



Evidential Nearest Neighbours in Active Learning

Daniel Zhu, Arnaud Martin, Yolande Le Gall, Jean-Christophe Dubois,
Vincent Lemaire

► To cite this version:

Daniel Zhu, Arnaud Martin, Yolande Le Gall, Jean-Christophe Dubois, Vincent Lemaire. Evidential Nearest Neighbours in Active Learning. Workop on Interactive Adaptive Learning (IAL) - ECML-PKDD, Sep 2021, Bilbao, Spain. hal-03327629

HAL Id: hal-03327629

<https://hal.science/hal-03327629>

Submitted on 27 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evidential Nearest Neighbours in Active Learning

Daniel Zhu¹, Arnaud Martin¹, Yolande Le Gall¹,
Jean-Christophe Dubois¹, and Vincent Lemaire²

¹ Univ Rennes, CNRS, IRISA, DRUID, rue E. Branly, Lannion
`firstname.lastname@irisa.fr`

² Orange Labs, France, `vincent.lemaire@orange.com`

Abstract. Active learning is a subfield of machine learning which allows to reduce the amount of data necessary to train a classifier. The training set is built in an iterative way such that only the most significant and informative data are used and labeled by an external person called oracle. It is furthermore possible to use active learning with the theory of belief functions in order to take erroneous labels due to the oracle's uncertainty and imprecision into account in order to limit their influence on the classifier's performance. In this article, we compare the classifier of the k nearest neighbours (kNN) to a variant based on belief functions from the theory of belief functions (EkNN), in a situation where some labels have been noised in order to model uncertain labels. We show that although the superiority of EkNN over kNN is not systematic, there are some interesting and modest results supporting the relevance of belief functions in active learning.

Keywords: Active Learning · Belief Functions · Theory of Dempster-Shafer · Nearest Neighbours

1 Introduction

In supervised machine learning, the size of the training set, *i.e.* the number of labeled examples, is often correlated to the performance of the learned model. Although having access to large database is no longer difficult nowadays, labeling the data remains an expensive task, especially when the application domain requires some expertise. Active learning offers a solution to this issue by reducing the number of labeled examples and ensuring that data to be labeled is selected by the model or a strategy [2,12]. Some classifiers can be combined with belief functions from the theory of Dempster-Shafer [5] in order to take the uncertainty and the imprecision in the labels into account, when the oracle – the person in charge of the labeling task – is not necessarily proficient in the domain.

In this paper, our contribution consists in the use of an evidential variant of the k nearest neighbours classifier which involves belief functions in active learning. More precisely, we compare this evidential classifier to the common k nearest neighbours in a context where the labels provided by the oracle are uncertain.

This article is organised as follow. First of all, we introduce active learning and belief functions in section 2. Our contribution is then described in section 3 and our experiments and results are presented in section 4. Finally the last section 5 concludes this paper.

2 State of the Art

In this section, we first introduce some notions about active learning (section 2.1) before dealing with the theory of belief functions (section 2.2).

2.1 Active Learning

Active learning (AL) is a subfield of machine learning which allows to limit the amount of training data to train a classifier. Its specificity lies in the construction step by step of a reduced training set, with a limited amount of information [10], by choosing only the most relevant and informative samples which provide an increasing in performances [2,12].

For a given classification problem, let us consider $\mathcal{X} \subset \mathbb{R}^d$, the set of samples described by $d \in \mathbb{N}^*$ features and \mathcal{Y} , the set of the different classes. The labeled and unlabeled samples are respectively gathered in \mathcal{L} and \mathcal{U} such that $\mathcal{X} = \mathcal{L} \cup \mathcal{U}$ and $\mathcal{L} \cap \mathcal{U} = \emptyset$. The aim is to label the minimum amount of data required by the model to reach a given performance or the best performance given a budget. The classifier first selects the sample $x \in \mathcal{U}$ whose contribution to the model is supposed to be the most significant before asking for its label to the oracle; this step is called a *query*. Once the label $y \in \mathcal{Y}$ of x is provided, the classifier learns it and updates its knowledge, moving x from \mathcal{U} to \mathcal{L} . Queries are formulated repeatedly in the same way until a certain stop criterion is satisfied. At the end of the complete learning process, it is highly probable that \mathcal{U} still contains a lot of samples but it is a matter of little importance as the classifier does not need to learn the whole dataset to perform efficiently.

This strategy is known as the *pool-based sampling* [8] and will be used thereafter.

There are several ways to select and to evaluate the relevance or the “informativeness” of a sample. The utility measures defined by the active learning strategies in the literature [12] differ in their positioning according to a dilemma between the exploitation of the current classifier and the exploration of the training data. Among these strategies, the *uncertainty sampling*, more dedicated to the exploitation part, is one of the most popular and can be divided into three main different types:

- the *least confident* prediction: the sample x_{LC}^* which minimises the probability of classification of its most probable class is queried:

$$x_{LC}^* = \arg \max_{x \in \mathcal{U}} 1 - P(y_x | x) \quad (1)$$

With y_x being the most probable class for x according to the classifier. However, this method depends only on the most probable class and do not take the other classes into consideration;

- the *margin sampling*: the sample x_M^* which minimises the difference between the probability of classification of the two most probable classes is considered as the most uncertain and will be sent to the oracle:

$$x_M^* = \arg \min_{x \in \mathcal{U}} P(y_x^{(1)}|x) - P(y_x^{(2)}|x) \quad (2)$$

With $y_x^{(1)}$ and $y_x^{(2)}$ being respectively the most and the second most probable classes for $x \in \mathcal{U}$ according to the classifier. The idea is quite natural: if the difference between the probabilities – the margin – is small, then the classification of x would be considered as ambiguous;

- the *entropy sampling*: the sample x_H^* which maximises the entropy of Shannon is considered as the most uncertain and will be sent to the oracle:

$$x_H^* = \arg \max_{x \in \mathcal{U}} - \sum_{y \in \mathcal{C}} P(y|x) \log(P(y|x)) \quad (3)$$

Unlike the previous methods, this approach considers every class and not only the most probable ones.

One can also notice that the uncertainty could be “divided” or viewed in two distinct parts [7]: *aleatoric* uncertainty which does not depends on the oracle’s knowledge but rather, is inherent to random phenomena – such as tossing a coin; and *epistemic* uncertainty which depends on the oracle’s knowledge and ignorance – for example, distinguishing two different species of bird. The second type of uncertainty is more appropriate in our context of active learning as the labeling task strongly relies on the oracle’s proficiency in the domain; in particular, it is possible to adapt the uncertainty sampling to epistemic uncertainty [9].

It is however important to notice that the uncertainty described in this section is related to the classifier and not to the oracle. The latter’s lack of knowledge and epistemic uncertainty might be modeled with the theory of belief functions introduced in the next section.

2.2 Theory of belief functions

The theory of belief functions also known as theory of Dempster-Shafer is an ideal tool for modeling uncertainty and imprecision [4,13]. Both of these imperfections might be common in labels when the oracle is not an expert in the domain. It is therefore necessary to model these phenomena.

Let us consider a set Θ called the *frame of discernment*, containing the elementary and exclusive hypotheses of a given problem. In the context of classification, it would be the set of classes, therefore: $\Theta = \mathcal{Y}$.

The theory is based on *belief functions* which are defined from 2^Θ to $[0, 1]$, with 2^Θ the power set of Θ . The *basic belief assignment* (BBA) $m : 2^\Theta \rightarrow [0, 1]$ is a belief function which satisfies the *normalisation condition*:

$$\sum_{X \in 2^\Theta} m(X) = 1 \quad (4)$$

A BBA allows to assign elementary belief on different combinations of hypothesis. For example, if we consider $X = \{\theta_1, \theta_2, \theta_3\} \in 2^\Theta$, then $m(X)$ is the confidence assigned to θ_1 , θ_2 and θ_3 altogether and cannot be subdivided among the different sub-hypothesis $\{\theta_1\}$, $\{\theta_2\}$ or $\{\theta_3\}$; thus, $m(X)$ supports the veracity of X as a whole.

In particular, if $m(\emptyset) = 0$, then the assumption of *closed world* is true, meaning that the frame of discernment is exhaustive. On the contrary, the assumption of *open world* prevails if $m(\emptyset) \neq 0$, which implies that unknown hypothesis exist and may not belong to Θ . In this article, the world will be supposed as closed.

The uncertainty about an hypothesis $\theta \in \Theta$ is modeled by the value of the BBA: the higher the BBA is (close to 1) and the more confidence there is in θ . However, if the value of BBA is low (close to 0), it means that few evidences support θ . The imprecision happens when the oracle hesitates between several hypotheses. This phenomenon is modeled by non empty and non singleton values of 2^Θ and can be extended to the situation of total ignorance, in which case, the entire belief is assigned to Θ in the following way: $m(\Theta) = 1$ and $m(X) = 0$ for all $X \in 2^\Theta - \{\Theta\}$. In a such context, every hypothesis is possible. For example, let us consider a classification problem of 10 classes so that $\Theta = \mathcal{Y} = \{\theta_i \mid i \in \llbracket 1, 10 \rrbracket\}$ and let us note $A = \{\theta_1, \theta_5, \theta_8\} \in 2^\Theta$. If, for a given sample $x \in \mathcal{X}$, the oracle believes that the class of x might be either θ_1, θ_5 or θ_8 without being able to determine which one is the most likely and without having evidences supporting other classes, then a BBA $m_x : 2^\Theta \rightarrow [0, 1]$ can be defined for x in the following way: $m_x(A) = s$ and $m_x(\Theta) = 1 - s$, with $s \in]0, 1[$. Assigning a certain amount of belief $s \neq 1$ in A represents the uncertainty. As A is the union of three different classes, that means that the oracle is furthermore imprecise; he does not particularly favour one answer among the three. Finally, the remaining $1 - s$ of the belief is assigned to Θ , which models the ignorance of the oracle.

The core feature of the Dempster-Shafer's theory is the conjunctive rule of combination of Dempster. It allows the combination of several BBA defined from the same frame of discernment. As a result, the hypotheses on which the BBA agree are enhanced. Let us consider $l \in \mathbb{N}^*$, the BBAs $(m_i)_{i \in \llbracket 1, l \rrbracket}$ defined on the same frame of discernment Θ will be combined into a single BBA m_\oplus :

$$\forall X \in 2^\Theta \quad m_\oplus(X) = \sum_{X_1 \cap \dots \cap X_l = X} \prod_{k=1}^l m_k(X_k) \quad (5)$$

Even though each BBA m_i respects the assumption of closed world, m_\oplus might not. It is possible to normalise m_\oplus in order to restore the closed world assumption.

tion:

$$\forall X \in 2^\Theta \quad m_{\text{Norm}}(X) = \begin{cases} \frac{m_\oplus(X)}{1-m_\oplus(\emptyset)} & \text{if } X \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

This rule of combination might be useful in decisions' rules. It is therefore possible to define classifiers based on belief functions as described in the next section.

3 An evidential classifier in active learning

Active learning is a paradigm of machine learning which reduces the amount of training data necessary to train the classifier. As uncertain oracle leads to errors in the labeling task, by coupling AL and the theory of belief functions [15], one can expect a certain robustness towards incorrect labels.

In this section, we present the evidential k nearest neighbours introduced for the first time in [5] (section 3.1) before explaining our approach and its interests in active learning (section 3.2).

3.1 Evidential nearest neighbours

The evidential k nearest neighbours classifier (EkNN) is a variant of the classical k nearest neighbours (kNN) based on belief functions [5].

Let us consider $x \in \mathcal{U}$ and $\tilde{x} \in \mathcal{L}$, one of the k nearest neighbours of x according to the euclidian distance. It is possible to define a BBA $m_{x,\tilde{x}}$ which supports the sole hypothesis that x and \tilde{x} belong to the same class $\theta \in \Theta$. A such BBA must also deal with the distance between x and \tilde{x} : the closer they are and the stronger the belief is. Therefore, the BBA might be defined as follow:

$$\forall X \in 2^\Theta \quad m_{x,\tilde{x}}(X) = \begin{cases} \alpha \exp(-\gamma_\theta d(x, \tilde{x})) & \text{if } X = \{\theta\}, \\ 1 - \alpha \exp(-\gamma_\theta d(x, \tilde{x})) & \text{if } X = \Theta, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

with α and γ_θ parameters and $d(x, \tilde{x})$ the Euclidean distance between x and \tilde{x} .

In the original paper [5], it is recommended to set α to 0.95 and γ_θ to the inverse of the mean distance between every training samples from the same class θ .

The combination rule of Dempster is then applied among the BBA which support the same classes. For each class, we get the BBA $(m_{x,(\theta)})_{\theta \in \Theta}$. Finally, the rule of combination is applied one more time among every BBA $(m_{x,(\theta)})_{\theta \in \Theta}$ and we get m_x which aggregates every original BBA supporting different classes.

The decision rule is then the following:

$$\forall x \in \mathcal{U} \quad y = \arg \max_{\theta \in \Theta} m_x(\{\theta\}) \quad (8)$$

In the context of active learning, it is necessary to take into account the time complexity of the training phase, as the classifier is updated after each query.

In the implementation of EkNN used in this article, the training phase consists to store the training data, then, to compute the γ_θ parameters which take the value of the inverse of the mean distance between every pair of samples from the same class $\theta \in \Theta$. If the euclidian distance is used, then the number of features d will necessarily influence the time complexity. Let us consider $n_\theta \in \mathbb{N}^*$, the number of samples whose class is θ . There are $N_\theta = \frac{n_\theta(n_\theta-1)}{2}$ unique couples of samples, which is also the number of distances to be computed. So the time complexity to compute the mean distance of the class θ is in $O((d+1)N_\theta)$. The global time complexity of the training phase (including every class of Θ) is then in $O((d+1)\sum_{\theta \in \Theta} N_\theta) = O(\frac{d+1}{2} \sum_{\theta \in \Theta} (n_\theta^2 - n_\theta)) = O(d \sum_{\theta \in \Theta} n_\theta^2)$. As the training set grows by adding samples, it is not necessary to recalculate the distances that have already been computed; in order to update the mean distance, adding the distances of the new training data and weighting them accordingly is sufficient.

In this context, the fact that the class of an unknown sample x might be one of its neighbour's would be a form of imprecision if there are several different classes. Moreover, the uncertainty could be deduced from the distance: the closer the neighbour is the more believable it is that x belongs to the same class as its neighbour. However, uncertainty and imprecision do not directly depend on the oracle in this situation but only on the training set used by EkNN.

3.2 Use of a belief functions-based classifier in active learning

The interest in using the theory of belief functions is to model uncertainty and imprecision in the data used in active learning. In particular, it becomes possible to model the ignorance, which would be difficult in the classical theory of probability [13]. In the context of crowdsourcing, [1] and [14] applied the theory of belief functions to model the contributors' uncertainty.

The approach described in this article consists to use EkNN in the context of active learning where some labels provided by the oracle are false. Parameters such as level of confidence or expertise estimation can not be taken into account in a such configuration as far as we know. Therefore, the oracle's uncertainty is not modeled in this article. The approach tends rather to limit the influence of erroneous labels given by the oracle. Once the learning phase is over, with some samples being mislabeled, the EkNN classifier uses the density of the distribution and its distances to modulate the influence of each neighbor of a sample to be labeled. It is finally the use of the combination rule (eq. (5)) that will contribute to enhance the hypotheses where the BBAs are in agreement, and therefore, to limit the indirect effect of the oracle's uncertainty.

Let us consider a dataset of two classes $\Theta = \{\theta_1, \theta_2\}$ and the classifier EkNN with $k = 5$. Let x be a sample to be labeled and its 5 nearest neighbours $(x_i)_{i \in [1,5]}$ represented in figure 1. Whether the oracle mislabeled the samples (case A) or not (case B), the goal after the learning phase for the classifier is to find the actual class of x . In case B, two neighbours have been mislabeled, therefore, class 1 is majoritary, but the two remaining class 2 (correctly labeled) are closer to x . The classifier will take the distance into account and attribute a

greater belief on θ_2 even though the actual class is minority in the neighbourhood (see table 2). The coordinates of the samples and their distances to x are described in table 1.

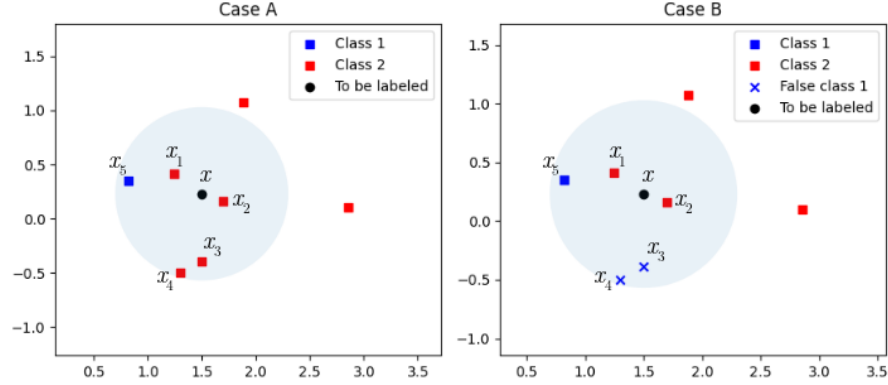


Fig. 1. A sample x to be labeled and its 5 nearest neighbours (within the circle). In case A, every sample has its actual label but in case B, two neighbours have been mislabeled by the oracle.

Table 1. Coordinates and distances to x of its 5 nearest neighbours of figure 1.

Sample	x	x_1	x_2	x_3	x_4	x_5
Coordinate	(1.5, 0.23)	(1.7, 0.15)	(1.3, 0.41)	(1.5, -0.39)	(1.3, -0.51)	(0.82, 0.35)
Distance to x	0	0.22	0.27	0.70	0.62	0.77

Table 2. Values of m_x of figure 1.

	$m_x(\emptyset)$	$m_x(\{\theta_1\})$	$m_x(\{\theta_2\})$	$m_x(\Theta)$
Case A	0	$5.97 \cdot 10^{-3}$	0.919	$7.49 \cdot 10^{-2}$
Case B	0	$5.05 \cdot 10^{-2}$	0.808	0.142

4 Experiments

In this section, we prove that EkNN is a viable classifier in active learning and that its robustness to uncertain labels is interesting. We first present the methodology and protocol used in our experiment (section 4.1) and then discuss and interpret the results (section 4.2).

4.1 Methodology

The EkNN classifier will be compared to kNN in order to highlight the performance of the former in active learning. The value of the nearest neighbours k is difficult to choose. A very small value will make the data very sensitive to noise while a larger value will lead to heavy computations. In this article, the value k will be arbitrarily set to 5 for each experiment and each dataset. These classifiers will also be used in association with uncertainty sampling (least confident prediction) and random sampling to select data for labeling. The secondary sampling method involves selecting the data to be labeled randomly so that its potential contribution or relevance to the model is ignored. Random sampling is often used to highlight a better efficiency of active learning’s sampling methods.

Several datasets are used to evaluate the classifiers’ behaviour. Synthetic random data have been generated from scikit-learn’s library [11] while some real datasets have been extracted from the UCI repository [6]. The different datasets are presented in tables 3 and 4.

Table 3. Synthetic datasets used in the experiments.

Name	#samples	#classes	Class distribution	#features
Synthetic A	1,000	2	90 %-10 %	10
Synthetic B	1,000	5	50 %-20 %-15 %-10 %-5 %	10
Synthetic C	1,000	5	75 %-10 %-5 %-5 %-5 %	10

Table 4. Real datasets used in the experiments. In the legend, **LRC** and **MRC** stand for “least represented class” and “most represented class”.

Name	#samples	#classes	LRC	MRC	#features
Speaker Accent Recognition	329	5	8.8 %	50 %	12
HCV	615	5	1.1 %	87 %	14
Letter Recognition (V vs. Y)	1,550	2	49 %	5.5 %	15
Wine Quality (Red Wine)	1,599	6	0.63 %	43 %	11
Pen-Based Recognition of Hand-written Digits (3 vs. 6 vs. 8)	3,166	3	33 %	33 %	16

To evaluate the performance of a classifier coupled to a given sampling method, the *accuracy* criterion is often used, but this metrics is not always relevant, especially when the class distribution is unbalanced [3]. An alternative is to use the *balanced accuracy*, denoted by a_b and defined in the following equation:

$$a_b = \frac{1}{2} \left(\frac{TP}{P} + \frac{TN}{N} \right) \quad (9)$$

With TP , TN , P and N being respectively the amount of true positive, true negative, of samples from the positive class and samples from the negative class.

The results for a combination of a classifier and a sampling method will be plot under the form of a learning curve of balanced accuracy according to the number of queries.

First, a simple comparison between EkNN and kNN is made according to the following protocol. The initial dataset \mathcal{X} is split into a test set \mathcal{T} consisting of 25 % of the samples while a “training” set contains the remaining samples. From the latter, 5 samples of each class is drawn, forming the *bootstrap* \mathcal{B} , and will be used to pre-train the classifier. The remaining of the “training” set becomes the *pool* \mathcal{P} so $\mathcal{X} = \mathcal{T} \cup \mathcal{B} \cup \mathcal{P}$. Then, the active learning phase begins: the classifier will forge queries from \mathcal{P} by selecting the sample according to a given sampling method (uncertainty or random sampling) until the number of 150 queries is reached. After each query, the oracle gives the label of the requested sample and the classifier updates its knowledge. Finally, the classifier computes the balanced accuracy on \mathcal{T} and adds it to the learning curve before making another query. The whole process is repeated 20 times such that an averaged learning curve is computed over the 20 learning curves for a given combination of classifier-sampling method.

Second, in order to add some uncertainty to the answers provided by the oracle, noise will be added to the labels, meaning that some of them are replaced by false values. Noise might not always be caused by uncertainty (the oracle can still provide false label while being confident and certain) but it will be sufficient in the current configuration as the uncertainty is studied through its consequences. The protocol used is the same as in the previous experiment except that a copy of the labels is generated with t % of them being noised. To answer queries, the oracle uses this noised copy instead of the original labels. Finally, the noised curves will be compared to the curves corresponding to the data that have not been noised.

4.2 Results

The comparison between EkNN and kNN in figure 2, whether with uncertainty or random sampling, suggests that the contribution of belief functions is most often interesting in AL. It is important to highlight that the EkNN is always superior or equivalent to kNN. This is the case for each dataset except for Speaker Accent Recognition where EkNN’s random sampling has significant low performance compared to the other curves. However, the difference is not always significant as the confidence intervals are often overlapping. Due to the definition of the balanced accuracy, the dataset with balanced class distribution presents higher score while the highly unbalanced one generate more learning difficulties. As a sidenote, the fact that uncertainty sampling is superior to random sampling is an expected behaviour, otherwise AL would not be an interesting learning paradigm.

When noise is added to label, it is worth mentioning that on dataset that are easy to classify, the confidence intervals of noised data is wider than non-noised data (Hard). This might be explained by the fact that noised data leads to bigger variance among results, and thus to a less precise balanced accuracy as it can

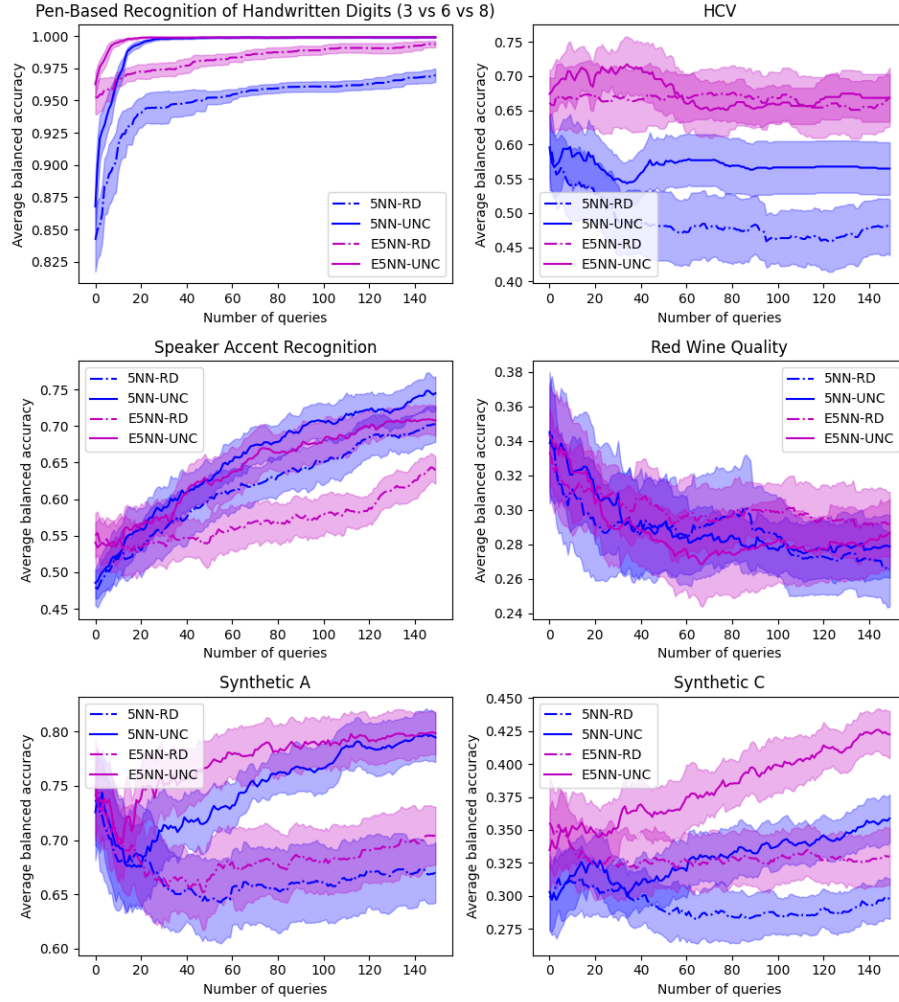


Fig. 2. Comparison of the balanced accuracy (with confidence intervals) between EkNN and kNN through uncertainty sampling (UNC) and random sampling (RD).

be seen in figure 4. Although, this is not always the case as shown in figures 3 and 5. Predictably in such cases, when the noise rate is high, the Hard and Noised curves of a same classifier move away from each other. It is more interesting to compare EkNN and kNN’s Noised curves. Again, the former’s curve is often above or at the same level as the latter’s curve. The gap between the confidence intervals, however, does not appear to be large as they often overlap each other; the gap is wider in figure 5.

Therefore, it appears that EkNN is slightly more robust to noise than kNN.

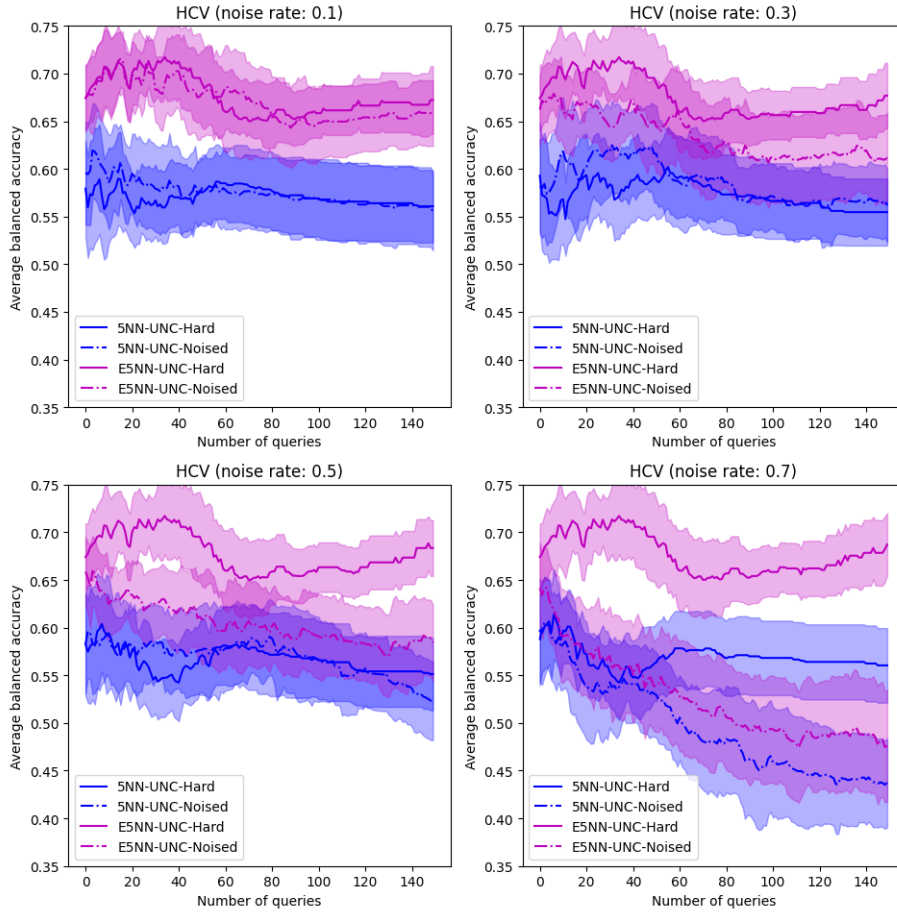


Fig. 3. Comparison of the balanced accuracy (with confidence intervals) between EkNN and kNN on HCV dataset through noised and non-noised (Hard) labels.

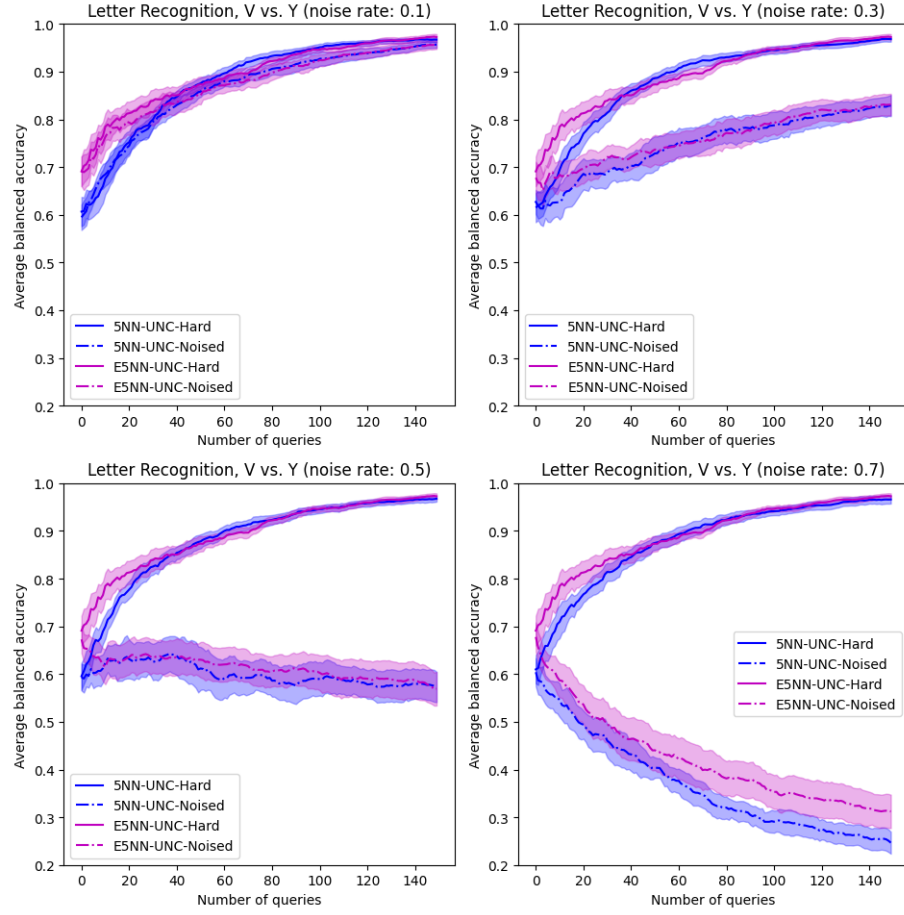


Fig. 4. Comparison of the balanced accuracy (with confidence intervals) between EkNN and kNN on Letter Recognition dataset through noisy and non-noisy (Hard) labels.

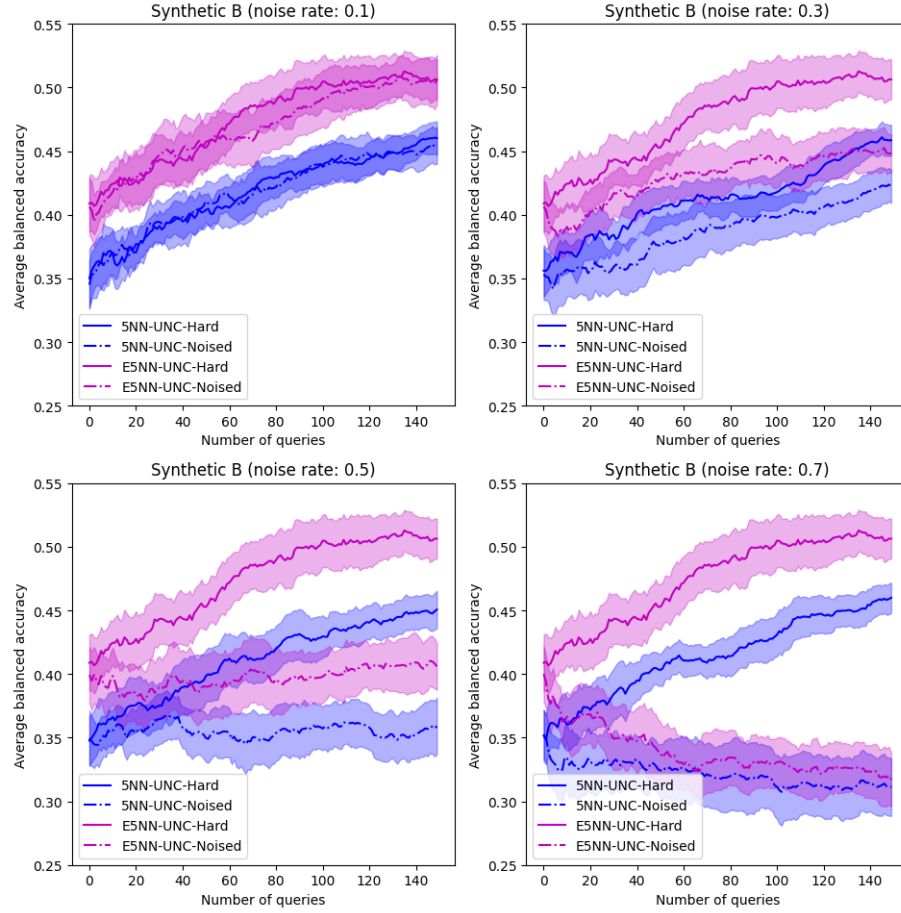


Fig. 5. Comparison of the balanced accuracy (with confidence intervals) between EkNN and kNN on synthetic dataset through noised and non-noised (Hard) labels.

5 Conclusion

Active Learning is a subfield of Machine Learning that aims to reduce the size of the training set and the amount of labels required. This paradigm might be coupled to the theory of belief functions in order to have a way to model uncertainty and imprecision amongst the data.

In this article, we attempt to show the efficiency of a classifier based on belief functions, EkNN, compared to a more common classifier, kNN. Our results suggest that in the majority of the experiments there is a real contribution of the EkNN classifier.

The imperfect label discussed in this article mostly covered the uncertainty aspect with the use of noised label. Further experiments on imprecision could be done in future work to complete this paper. This could be achieved by allowing the oracle to propose several classes instead of one after each query. However, a more flexible classifier would be required to treat imprecision as EkNN is not particularly adapted to deal with several classes per sample. Besides, the EkNN classifier requires heavy computation in AL as the distances between samples are re-computed after each query. Thus, it might be interesting to design a more efficient and adapted belief functions-based classifier for AL in order to treat both uncertainty and imprecision.

References

1. Abassi, L., Boukhris, I.: A worker clustering-based approach of label aggregation under the belief function theory. *Applied Intelligence* **49**(1), 53–62 (2019). <https://doi.org/10.1007/s10489-018-1209-z>
2. Bondu, A., Lemaire, V.: État de l’art sur les méthodes statistiques d’apprentissage actif. In: *Apprentissage Artificiel et Fouille de Données, AAFD* (2006)
3. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: *2010 20th International Conference on Pattern Recognition*. pp. 3121–3124 (2010). <https://doi.org/10.1109/ICPR.2010.764>
4. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38**(2), 325–339 (04 1967). <https://doi.org/10.1214/aoms/1177698950>, <https://doi.org/10.1214/aoms/1177698950>
5. Denoeux, T.: A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* **25**(5), 804–813 (1995). <https://doi.org/10.1109/21.376493>
6. Dua, D., Graff, C.: UCI machine learning repository (2017), <http://archive.ics.uci.edu/ml>
7. Hüllermeier, E., Waegeman, W.: Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning* **110**(3), 457–506 (2021). <https://doi.org/10.1007/s10994-021-05946-3>, <https://doi.org/10.1007/s10994-021-05946-3>
8. Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In: Croft, B.W., van Rijsbergen, C.J. (eds.) *SIGIR ’94*. pp. 3–12. Springer London, London (1994)

9. Nguyen, V.L., Destercke, S., Hüllermeier, E.: Epistemic uncertainty sampling. In: Kralj Novak, P., Šmuc, T., Džeroski, S. (eds.) *Discovery Science*. pp. 72–86. Springer International Publishing, Cham (2019)
10. Nodet, P., Lemaire, V., Bondu, A., Cornuéjols, A., Ouorou, A.: From Weakly Supervised Learning to Biquality Learning: an Introduction. In: *In Proceedings of the International Joint Conference on Neural Networks (IJCNN)* (2021)
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011)
12. Settles, B.: Active learning literature survey. *Computer Sciences Technical Report 1648*, University of Wisconsin-Madison (2009)
13. Shafer, G.: *A mathematical theory of evidence*, vol. 42. Princeton university press (1976)
14. Thierry, C., Casiez, G., Dubois, J.C., Le Gall, Y., Malacria, S., Martin, A., Pietrzak, T., Uro, P.: Interface de Recueil de Données Imparfaites pour le Crowd-Sourcing. EGC 2020 - Humains et IA, travailler en intelligence Atelier de la conférence (Jan 2020), <https://hal.inria.fr/hal-02465761>
15. Zhu, D., Martin, A., Dubois, J.C., Le Gall, Y., Lemaire, V.: Modèle crédibiliste pour l'échantillonnage en apprentissage actif. In: *Rencontres francophones sur la logique floue et ses applications, (LFA)* (2021)