*Electronic Letters on Computer Vision and Image Analysis 21(1):111-120, 2022*

# Feature selection based on discriminative power under uncertainty for computer vision applications

M. Chakroun$^{*+}$, S. A. Bouhamed$^{*+}$, I. K. Kallel$^{*+}$, B. Solaiman$^{+}$ and H. Derbel$^{*}$

$^{*}$ *Control and Energy Management Lab, ENIS, University of Sfax, Tunisia*
$^{+}$ *(iTi) Department, IMT Atlantique, Brest, France*

Received 15th of December, 2020; accepted 28th of May 2022

**Abstract**

Feature selection is a prolific research field, which has been widely studied in the last decades and has been successfully applied to numerous computer vision systems. It mainly aims to reduce the dimensionality and thus the system complexity. Features have not the same importance within the different classes. Some of them perform for class representation while others perform for class separation. In this paper, a new feature selection method based on discriminative power is proposed to select the relevant features under an uncertain framework, where the uncertainty is expressed through a possibility distribution. In an uncertain context, our method shows its ability to select features that can represent and discriminate between classes.

*Key Words*: Feature Selection, Computer Vision, Dimensionality, Uncertainty, Possibility Distribution.

## 1 Introduction

The emergence of Big Data analytic, mainly those containing a large number of features brings important requirements for strategies, which can effectively reduce the dimensionality of data. In real-world applications, the training data may be uncertain, imprecise or even missing due to unreliable information sources, noise, etc. These forms of imperfection are part of epistemic uncertainty. In this context, the possibilistic modeling is the appropriate framework not to remove these imperfections, but rather to consider them in the data modelling step [c1,c2].

Usually, in the modeling phase, data is characterized by using features. It is necessary to highlight that the use of all features, during the data modeling, increases the dimensionality and then the system complexity. In [3], the authors use 4 million features in their system. Surely, some of them are redundant or unnecessary. Therefore, they can be a source of different types of noise that can lead to a deterioration of the system performance.

The feature selection is becoming an increasingly important step in decision-making systems. Its main purpose is to reduce the dimensionality of the system, thereby reducing its complexity. Features do not have the

Correspondence to: <chakroun.marwa37@gmail.com>

Recommended for acceptance by <Prof Luiz M. Goncalves>
https://doi.org/10.5565/rev/elcvia.1361
ELCVIA ISSN: 1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

same importance towards classes. Some of them are good for class separation, while others are more reliable for class representation. The proposed method consists of selecting the relevant features under an uncertain framework, where the uncertainty is expressed through a possibility distribution.

Several feature selection methods are presented in the literature such as the Linear Forward Selection (LFS) [4], Principal Component Analysis (PCA) [5], Relief [6], Fisher score [7], and so on. Various overview papers of feature selection methods are presented in [8 -11]. Some of these methods have shown good results in some cases but have failed elsewhere. Therefore, we aimed to develop a feature selection strategy that can outperform conventional feature selection algorithms. Togaçar *et al.* [12] used a deep learning method based on convolutional neural networks (CNN) to classify plant species. In [13, 14], authors proposed feature selection methods based on CNN in order to classify and identify cancer types in microarray cancer data. It should be noted that if there is no good GPU, the CNN-based feature selection process will become very slow. Other recent interesting works have addressed feature selection under an uncertain framework [15, 16]. In this case, the relevant features are extracted from possibility distributions. In this paper, we are focused on the selection of features, which are modeled by possibility distribution. Indeed, the possibility theory could model both different kinds of information (categorical, numerical, symbolic, etc.) and deal with different types of imperfection (imprecision, ambiguity, etc.). It is, therefore, interesting to propose an efficient feature selection method in the possibility framework.

## 2 Possibility theory: An overview

Good modeling helps to improve the performance and reliability of computer vision systems since it considers the imperfections of information by using specific methods or theories to convert information into mathematical representations. In [1], the imperfect information modeling theory is reviewed. The choice of one of these theories strongly depends on the application and its complexity, the available knowledge and the type of imperfection [1]. For example, when there are a large number of samples and you only want to model random uncertainty information, probability theory is considered a very suitable theoretical framework. However, if the sample size is limited and other types of imperfections must be considered, the probability theory will not be the right choice.

For computer vision systems, the imperfection of data is part of epistemic uncertainty since it is due to the physical properties of the used visual sensor. In this context, the possibility theory is the most suitable framework for dealing with these imperfections. Therefore, special attention has been paid to this theory.

Possibility theory provides flexible mathematical tools and an effective model for dealing with uncertain information. It was first proposed by Zadeh [17] as an extension of the theory of fuzzy sets. In fact, while fuzzy set theory relaxes the framework of classical set theory to deal with imprecision, possibility theory is intended to deal with epistemic uncertainties. The possibility theory was then developed by many researchers such as Dubois and Prade [18-23]. In this section, we present an overview of the possibility theory.

### 2.1 Possibility distribution

Among the basic concepts used in the possibility framework, we have the possibility distribution, denoted by $\pi$ (.) defined as follows:

$$\begin{aligned} \pi : \Omega &\longrightarrow [0,1] \\ x_n &\longrightarrow \pi(x_n) \end{aligned} \tag{1}$$

which assigns to each singleton $x_n$ of $\Omega$, a value in the interval [0, 1]. The value $\pi(x_n)$ represents the possibility degree that the singleton $x_n$ is the only singleton that occurs. Considering this definition, two extreme knowledge situations can be defined:

- Complete knowledge: $\exists! x_n \in \Omega$, $\pi(x_n)$ =1 and $\pi(x_m)$ =0, $\forall x_m \in \Omega$, $x_m \neq x_n$
- Total ignorance: $\forall x_n \in \Omega$, $\pi(x_n)$ =1 (all singletons of $\Omega$ are quite possible)

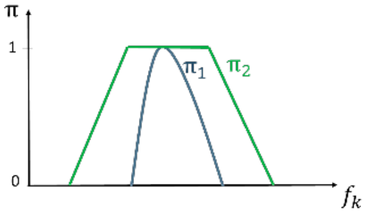

Figure 1: Two possibility distributions with different specificity

The estimation of the possibility distribution is one of the most discussed issues in possibility theory [1, 2]. In [1], the techniques for estimating the possibility distribution are divided into two categories: (i) techniques guided by expert knowledge, used to directly or indirectly express the general form of the possibility distribution, and (ii) techniques based on characteristic functions of other theories of uncertainty to infer the possibility distribution, namely: the probability distribution, the mass function, etc. In our work, we pay particular attention to the technique of estimating possibility distributions based on probability distributions transformation [18-23]. Note that this conversion represents conceptual consistency, as probability and possibility distributions apply to information that is susceptible to uncertainty.

### 2.2 Yager's specificity measure "$S_p$"

The Yager's specificity describes the change in the amount of information of a possibility distribution [24]. In this work, the "$S_p$" is used to evaluate the feature capacity for representing classes. In fact, the more specific the distribution, the better the class is represented. Let $\pi_1$ $\pi_2$ and two possibility distributions. As shown in Fig. 1, $\pi_1$ is more specific than $\pi_2$ according to a feature $f_k$ as met only when $\pi_1(\omega_i) \leq \pi_2(\omega_i)$, $\forall\, \omega_i \in \Omega$.

In the following, the proposed feature selection method, based on Yager's specificity measure of a possibility distribution to infer the discriminative power, is detailed.

## 3 Proposed feature selection strategy

### 3.1 Context and motivation

Let $\Omega$ = $\{C_1, C_2, \ldots, C_i, \ldots C_N\}$ the universe of discourse, $\{Obj_1, Obj_2, \ldots, Obj_i, \ldots Obj_N\}$ a set of objects belonging respectively to the classes $\{C_1, C_2, \ldots, C_i, \ldots C_N\}$ and acquired by a sensor. Let $\{f_1, f_2, \ldots, f_k, \ldots f_K\}$ be the features set extracted from the acquired data. These feature values measured on objects allow these objects to be attributed to one of the $\Omega$ classes. Sometimes we can measure the same value of a feature on objects belonging to different classes, as shown in Fig. 2. This situation can cause ambiguity when assigning unique class labels to objects, which can affect the efficiency of the classifier. Therefore, elucidating this ambiguity is important.

### 3.2 Proposed strategy

Our proposed strategy is based on both, feature capacity for class representation "$C_{p_{rep}}$" and feature capacity for classes separation "$C_{p_{sep}}$". The ability of a feature $f_k$, k ∈ {1,..., K}, in class representation is estimated based on the calculation of the average of all specificity measures $S_p(\pi^{f_k}_{C_i})$, i ∈ {1,..., N}. In our work, the capacity of a feature to separate classes is related to the maximum height of all intersection points "$M_h$" of N possibility distributions. In fact, the lower "$M_h$" according to a feature, the greater the capacity of the considered feature to separate the different classes. The discriminative power of a feature is determined as the average of its two capacities "$C_{p_{rep}}$" and "$C_{p_{sep}}$". In the following, the algorithm of the proposed feature selection method is given.

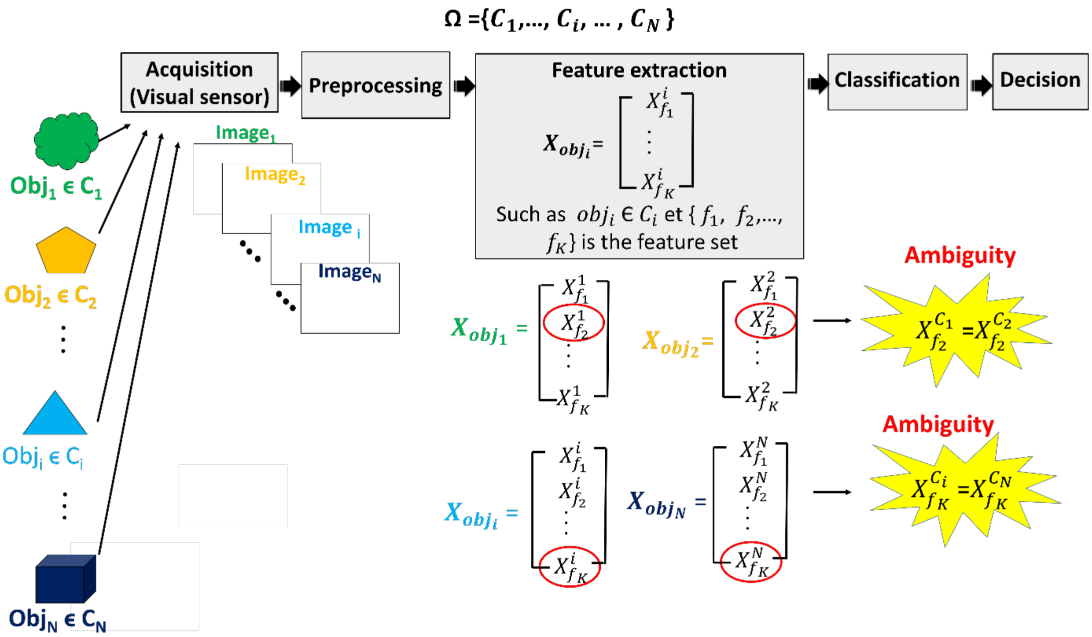
$\Omega = \{C_1,..., C_i, ... , C_N\}$
Acquisition (Visual sensor)
Preprocessing
Feature extraction
$X_{obj_i} = [X^i_{f_1} \; \vdots \; X^i_{f_K}]$
Such as $obj_i \in C_i$ et $\{f_1, f_2,..., f_K\}$ is the feature set
Classification
Decision
$Obj_1 \in C_1$
$Obj_2 \in C_2$
$Obj_i \in C_i$
$Obj_N \in C_N$
$Image_1$
$Image_2$
$Image_i$
$Image_N$
$X_{obj_1} = [X^1_{f_1}, X^1_{f_2}, \vdots, X^1_{f_K}]$
$X_{obj_2} = [X^2_{f_1}, X^2_{f_2}, \vdots, X^2_{f_K}]$
Ambiguity
$X^{C_1}_{f_2} = X^{C_2}_{f_2}$
$X_{obj_i} = [X^i_{f_1}, X^i_{f_2}, \vdots, X^i_{f_K}]$
$X_{obj_N} = [X^N_{f_1}, X^N_{f_2}, \vdots, X^N_{f_K}]$
Ambiguity
$X^{C_i}_{f_K} = X^{C_N}_{f_K}$

Figure 2: Uncertainty related to attributes in a classification process

**Algorithm: The proposed feature selection strategy**

```
Inputs: Ω={$\pi_{C_1}$, $\pi_{C_2}$, ..., $\pi_{C_N}$}, the set of features F= {$f_1$,$f_2$, ..., $f_K$}
Output: b: boolean value, b=1 if the feature is selected, b=0 if it is rejected
for k=1 to K do
    Step1: Calculate "$C^{f_k}_{prep}$"
    Calculate the Yager's specificity measure of each possibility distribution $\pi_{C_i}$
    $S_p(\pi^{f_k}_{C_i}) \longleftarrow$ Specificité Yager $(\pi^{f_k}_{C_i})$
    Calculate the average of all specificity measures of $S_p(\pi_{C_i})$ i=1,..., N
    $C^{f_k}_{prep} \longleftarrow \frac{\sum_{i=1}^{N} S_p(\pi^{f_k}_{C_i})}{N}$
    Step 2: Calculate "$C^{f_k}_{psep}$"
    "$H_k$": the set of heights of the intersection points of all possibility distributions
    $h^{f_k}_{i,j} \longleftarrow \pi^{f_k}_{C_i} = \pi^{f_k}_{C_j}$ i, j ∈ {1,..., N}; $i \neq j$
    $H_k \longleftarrow \{h^{f_k}_{i,j} \neq 0\}$
    $M^{f_k}_h \longleftarrow$ max($H_k$)
    $C^{f_k}_{psep} \longleftarrow$ 1-$M^{f_k}_h$
    Step 3: Calculate the discriminative power "$D_p$"
    $D^{f_k}_p \longleftarrow \frac{C^{f_k}_{prep}+C^{f_k}_{psep}}{2}$
    Step 4: Let $\delta$ an empirical threshold
    if $D^{f_k}_p > \delta$
        b ⟵ 1
    else
        b ⟵ 0
    end if end for
```

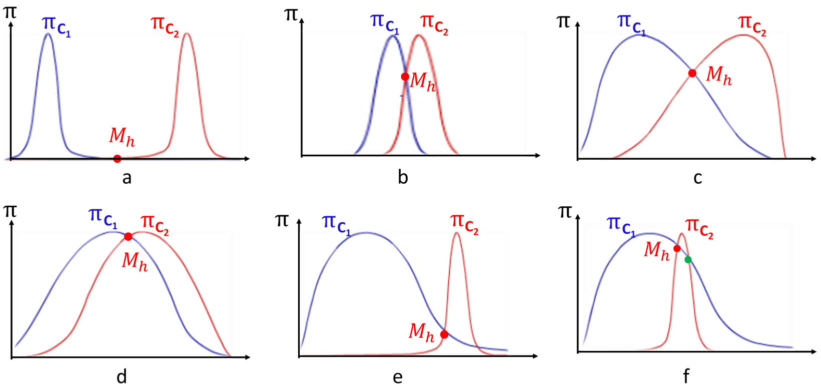

Figure 3: Different representation situations of two classes $C_1$ and $C_2$ according to the shapes and the disposition of their respective models

The more specific the model, the better it is informative and representative of its class. Fig. 3 shows that the separation between two classes is more marked as their respective models are more specific (narrower shapes). The two cases illustrated by Fig. 3(c) and (d) show two classes represented by non-specific models, which are not useful for separation.

At least one model must be relatively specific to be able to discuss the separability between classes. The heights of the intersection points between the models are informative about the separability between classes. The lower this value, the better the classes are separable. Fig. 3(a) and (e) illustrate the favorable separation situations between two classes, while Fig. 3(b) and (f) illustrate unfavorable situations for separation. The ideal case, illustrated by Fig. 3(a), corresponds to zero height.

The validation of the proposed feature selection method, as well as the obtained results, are presented in the following.

# 4 Experimental results

Two experiments are carried out to validate the proposed feature selection strategy in the possibilistic context. The first experiment is carried out on a synthetic data set, which is constructed to know in advance the ability of each feature regarding class discrimination and class representation. This experiment helps us to verify that the expected results are achieved. The second experiment consists in evaluating our strategy on different benchmark datasets and comparing it with other feature selection methods. The classification task is performed by an SVM classifier.

## 4.1 Synthetic database

The constructed synthetic data set, given in the form of possibility distributions conditional to each class, assumes a mixture of relevant and not relevant features. The universe of discourses of this data set is composed of three classes $\Omega = \{C_1, C_2, C_3\}$, described by nine features $F = \{f_1, f_2, .., f_9\}$.

From Table 1 and Fig. 4, we can conclude that:

- Ideal features: $\{f_1\}$ since its $C_{p_{sep}}$ is equal to 1 and it has the highest $C_{p_{rep}}$.
- Discriminant features: $\{f_2, f_5, f_6, f_9\}$ since their $C_{p_{sep}}$ and $C_{p_{rep}}$ are not too low.
- Features seem to be not discriminant: $\{f_3, f_4, f_7, f_8\}$ since their $C_{p_{sep}}$ are too low.

For the case where the threshold  is set empirically to 0.6, five features $\{f_1, f_2, f_5, f_6, f_9\}$ will be selected (Fig. 5)

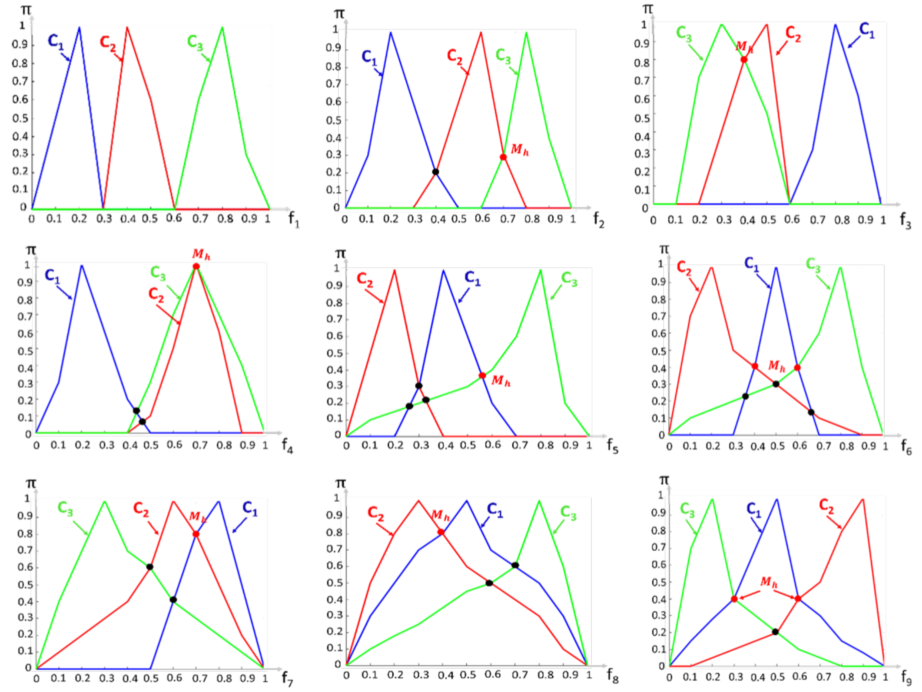

Figure 4: Possibility distribution of $C_1$, $C_2$ and $C_3$ according to each feature

Table 1: Discriminative capacities of each feature

| | Set of the height of intersection points $\{H_k\}$ | Maximum of the set of the height of intersection points $M_h$ | $Cp_{sep}$ | $Cp_{rep}$ |
|---|---|---|---|---|
| $f_1$ | $\{\varnothing\}$ | $\varnothing$ | 1 | 0.792 |
| $f_2$ | {0.2, 0.3} | 0.3 | 0.7 | 0.748 |
| $f_3$ | {0.8} | 0.8 | 0.2 | 0.758 |
| $f_4$ | {0.1, 0.05, 1} | 1 | 0 | 0.762 |
| $f_5$ | {0.15, 0.3,0.2, 0.35} | 0.35 | 0.65 | 0.707 |
| $f_6$ | {0.2, 0.4, 0.3, 0.4, 0.1} | 0.4 | 0.6 | 0.652 |
| $f_7$ | {0.6, 0.4, 0.8} | 0.8 | 0.2 | 0.53 |
| $f_8$ | {0.8, 0.5, 0.6} | 0.8 | 0.2 | 0.426 |
| $f_9$ | {0.4, 0.2, 0.4} | 0.4 | 0.6 | 0.604 |

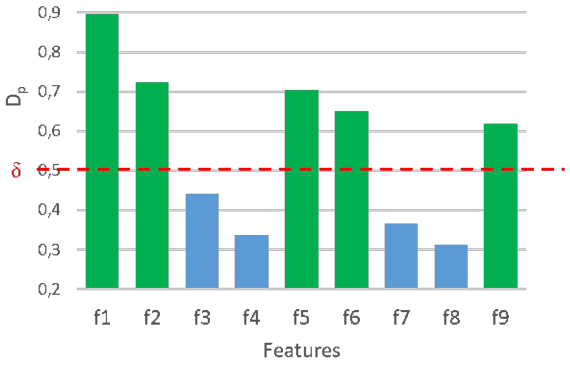

Figure 5: Discriminating power $D_p$ for each feature

## 4.2 Benchmark data set

To evaluate our strategy, we use four benchmark databases from UCI Machine Learning Repository [25], namely Glass database, iris flower database, Pima Indian diabetes database and vehicle silhouettes database. The obtained results are compared to those obtained from five literature feature selection approaches namely Principal Component Analysis (PCA), Linear Forward Selection (LFS), Fisher score, RELIEF and methods proposed in [15, 16] based on possibilistic models. The evaluation of the feature selection method is based on two concepts: the number of selected features (NSF) and the correct classification rate (CCR). Table 2 and Fig. 6 show the results obtained by each feature selection method based on the considered databases.

Table 2: Feature selection results obtained using the different methods

| Method | Iris | | Glass | | Diabetes | | Vehicle | |
|---|---|---|---|---|---|---|---|---|
| | NSF | CCR | NSF | CCR | NSF | CCR | NSF | CCR |
| All features | 4 | 96.2 | 9 | 95.3 | 8 | 78.1 | 18 | 82.8 |
| PCA | 2 | 81.3 | 6 | 93.9 | 4 | 75.6 | 7 | 74.7 |
| LFS | 2 | 95.3 | 7 | 93.9 | 4 | 77.8 | 10 | 70.6 |
| Fisher | 2 | 81.3 | 4 | 91.5 | 6 | 76.4 | 10 | 74.2 |
| Relief | 2 | 81.3 | 4 | 93.4 | 4 | 77.2 | 10 | 72.8 |
| Method proposed in [15] | 2 | 95.3 | 7 | 94.4 | 4 | 76.3 | 8 | 72.8 |
| Method proposed in [16] | 3 | 94.4 | 5 | 94.6 | 5 | 76.3 | 7 | 75.2 |
| Our strategy | 2 | 95.3 | 5 | 94.8 | 4 | 77.2 | 6 | 77.5 |

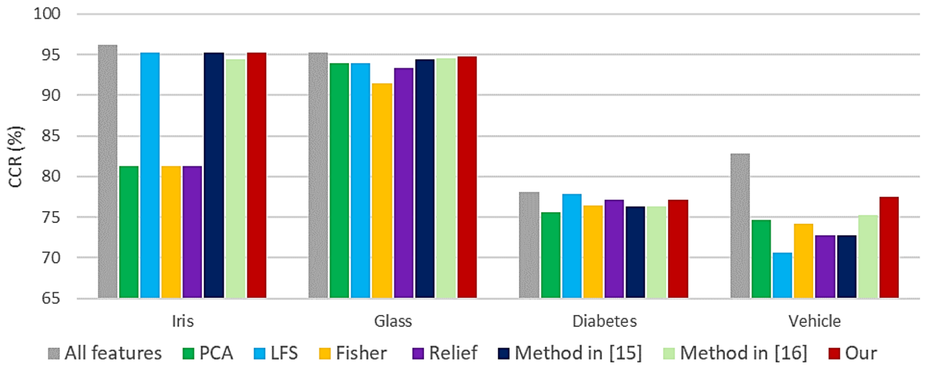

Figure 6: Correct classification rate calculated based on feature selection methods evaluated by 4 databases

As shown in Table2, the CCR is very dependent on the NSF, which makes difficult do highlight the contribution of our strategy. To avoid these dependencies, in [1], two rates are proposed to quantify the gain obtained by each selection method: i) the feature selection rate (FSR) defined by the equation (2), and ii) the classification effect rate (CER) defined by the equation (3). A CER value greater than 1 indicates that the correct classification rate has been improved before feature selection. A value of CER less than 1 means that the selection reduced the recognition. The higher the $|1 - CER|$, the greater the impact of the feature selection on the classification rate.

$$FSR = \frac{NSF}{NAF} \tag{2}$$

where:
NAF: number of all features

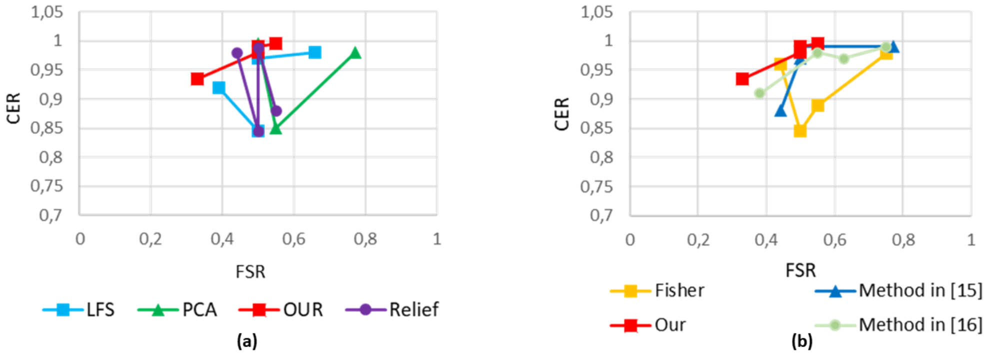

Figure 7: Variation of classification effect rate according to feature selection rate: (a) for LFS, PCA, Relief and our approach, (b) Fisher, method in [15], method in [16] and our approach

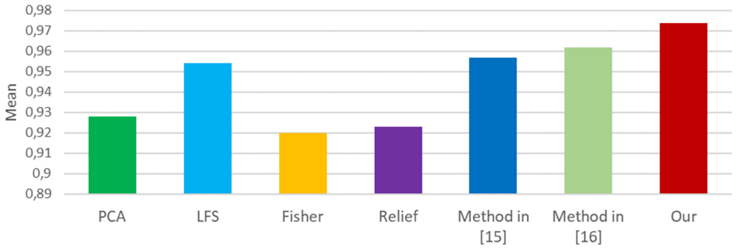

Figure 8: Mean of CER measures calculated based on feature selection methods evaluated by 4 databasess

$$CER = \frac{RSF}{RAF} \tag{3}$$

where:
RSF: recognition rate using selected features
RAF: recognition rate using all features

Each point in Fig. 7 corresponds to the performances (FSR, CER) achieved by a feature selection method for a dataset. It is quite visible that very few couples (method/dataset) have a CER greater than 1. This may be because features could have already been preselected. It should be noted that our proposed approach shows remarkable stability compared to the other approaches, with CER values very close to 1. This is visible by the linear, horizontal, tight, rectilinear shape, close to the unity, of its curve.

In order to have a more comprehensive understanding of the performance achieved by different methods, the mean is calculated as shown in Fig. 8. In fact, the mean value gives an overall estimate of the performance of the method.

Since we aim to assess the ability of feature selection methods to keep stability on the recognition rate, we propose to calculate the variance corresponding to each feature selection method. The method having the lowest variance value is the one who has the most recognition stability regarding the number of selected features. As shown in Fig. 9, our method attains the best stability score.

Our method allows us to obtain good classification results on any database while pushing the reduction of features as much as possible. Unlike our method and methods proposed in [15, 16], other methods in the literature have random performance. Therefore, we can conclude that the inference of the representation/separation capabilities of these methods leads to a good compromise between the feature reduction rate and the correct classification rate on any database. However, our method is not as complicated as methods proposed in [15, 16]. In fact, the method proposed in [15] mainly depends on the complexity of the Shapley index, which in-

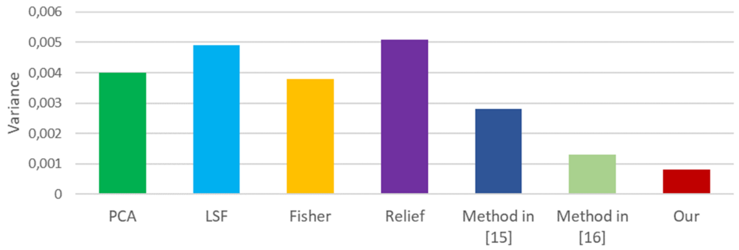

Figure 9: Variance of CER measures calculated based on feature selection methods evaluated by 4 databases

creases exponentially with the number of features, while the method proposed in [16] depends on the number of Cross-validation bases, which increases with the number of classes.

## 5 Conclusion

In this paper, we present a new method to select relevant features in an uncertain context where classes are presented through possibility distributions. The selection of a feature is realized according to its discriminative power that is defined as the average of the representation and separation capacities. The representation capacity is defined through the Yager' specificity measure [24]. In fact, the more specific the possibility distribution is the best the class is represented. The separation capacity of a feature depends on the height of the intersection points "$M_h$" of all possibility distributions linked to it. The lowest "$M_h$" height is the most the classes are separated.

Two experiments are carried out to evaluate the proposed feature selection method. Firstly, the proposed method is evaluated on synthetic data in order to verify that the expected results are achieved. Then, we evaluate our strategy based on benchmark datasets to compare its performance with literature feature selection methods (including 4 classic methods and a method that operates in the same context as ours). The possibilistic methods, namely the proposed method and those presented in [15, 16], have shown better performance than the classic methods. We noticed that our method has a slight advantage in terms of algorithm complexity. The evaluation was performed through the correct classification rate criterion using an SVM classifier.

## References

[1] I. K. Kallel, *Mécanismes de raisonnement possibiliste pour l'aide à la décision et l'interprétation de scènes*, Université de Sfax,⟨tel-02868499⟩, 2019.

[2] B. Solaiman and É. Bossé, Theory for the Design of Information Fusion Systems, Springer International Publishing, 2019.

[3] I.T. Phillips, A.K. Chhabra, "Empirical performance evaluation of graphic recognition systems", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(9):849-870, 1999.

[4] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," *Proc. of 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 213-220(2003).

[5] M. Gütlein, E. Frank, M. Hall, M. Karwath, "A large-scale attribute selection using wrappers", *IEEE Symposium on Computational Intelligence and Data Mining*, TN, USA, 332-339, 2009.

[6] L. I. Smith, A tutorial on principal components analysis, *Computer Science Technical Report No. OUCS-2002-12*, 2002, Retrieved from http://hdl.handle.net/10523/7534.

[7] I. Kononenko, “Estimating attributes: analysis and extension of RELIEF”, *Proc. of the 7th European Conference on Machine Learning*, Catania, Italy, 171–182, 1994.

[8] A. Bommert, X. Sun, B. Bischl, J. Rahnenfuhrer, M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data”, *Computational Statistics and Data Analysis*, 143, 2020.

[9] B. Venkatesh, J. Anuradha, “A Review of Feature Selection and Its Methods”, Cybernetics and information technologies, 19 (1): 3–26, 2019.

[10] J. Cai, J. Luo, S. Wang, S. Yang, “Feature selection in machine learning: A new perspective”, *Neurocomputing*, 300, 70–79, 2018.

[11] S. Darshan, C. Jaidhar, “Performance evaluation of filter-based feature selection techniques in classifying portable executable files”, *Procedia Comput. Sci.*, 125:346–356, 2018.

[12] M. Togaçar, B. Ergen, Z. Comert, “Classification of Flower Species by Using Features Extracted from the Intersection of Feature Selection Methods in Convolutional Neural Network Models”, *Measurement*, 158:1-12, 2020.

[13] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, “Gene selection and classification of microarray data using convolutional neural network,” *Proc. of the International Conference on Advanced Science and Engineering (ICOASE)*, Duhok, Iraq, 145-150, 2018.

[14] O. Ahmed, A. Brifcani, “Gene Expression Classification Based on Deep Learning,” *Proc. of the 4th Scientific International Conference Najaf (SICN)*, Al-Najef, Iraq, 145-149, 2019.

[15] S. A. Bouhamed, I. K. Kallel, D. S. Masmoudi et B. Solaiman, “Feature selection in possibilistic modeling”, *Pattern Recognition*, 48(11):3627-3640, 2015.

[16] M. Medhioub, S. A. Bouhamed, I. K. Kallel, N. Derbel and O. Kanoun, "Possibilistic Feature Selection Method based on Discriminant Power for Class Discrimination," *Proc. of the 17th International Multi-Conference on Systems, Signals Devices (SSD)*, 378-383, 2020.

[17] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility”, *Fuzzy Sets Syst.*, 1(11):3–28, 1978.

[18] D. Dubois et H. Prade, "Possibility Theory: An Approach to Computerized Processing of Uncertainty", *Plenum Press*, 1988.

[19] B. B. Devi, V. V. S. Sarma, “Estimation of fuzzy memberships from histograms”, *Inf. Sci*, 35(1):43-59, 1985.

[20] D. Dubois, H. Prade, S. Sandri “On possibility/probability transformations”, *Fuzzy Logic*, 12:103-112, 1993.

[21] G. Jumarie, “A General Approach to Approximate Reasoning via Probability-Possibility Conversion”, *International Conference on Systems, Man and Cybernetic*, 4: 656-661, 1993.

[22] G. J. Klir, J.F. Geer “Information-preserving probability-possibility transformations: recent developments”, *Fuzzy Logic*, 12:417-428,1993.

[23] M. Masson, T. Denoeux, “Infering a possibility distribution from empirical data”, *Fuzzy Sets and Syst.*, 157:319-340, 2006.

[24] R. R. Yager, “On the specificity of a possibility distribution”, *Fuzzy Sets and Syst.*, 50(3):279-292, 1992.

[25] UCI Machine Learning Repository [En ligne]. Available: http://archive.ics.uci.edu/ml/datasets.html.