



Hybrid one-class classifier ensemble based on fuzzy integral for open-lexicon handwritten Arabic word recognition

Bilal Hadjadji¹ · Youcef Chibani¹ · Hassiba Nemmour¹

Received: 31 January 2017 / Accepted: 19 July 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

One-class classifier (OCC) is involved for solving different kinds of problems due to its ability to represent a class distribution regardless the remaining classes. Its main advantage for multi-class classification is offering an open system and therefore allows easily extending new classes without retraining OCCs. So far, hidden Markov models, support vector machines and neural networks are the most used classifiers for Arabic word recognition, which provides a system with closed lexicon. In this paper, the OCCs are explored in order to perform an Arabic word recognition system with an open lexicon. Generally, pattern recognition systems designed by a single system suffer from limitations such as the lack of uniqueness and non-universality. Thus, combining multiple systems becomes an attractive research topic for performance and robustness enhancement. Fixed rules are commonly used as combiners for the hybrid OCC ensembles. The present paper aims to propose a combination scheme of OCCs based on the use of fuzzy integral (FI) operators. Furthermore, an alternative framework is proposed to design a parameter-independent and open-lexicon handwritten Arabic word recognition system as well as a new density measure function. Experimental results conducted on Arabic handwritten dataset using different types of OCCs with large number of classes highlight the superiority of FI for hybrid OCC ensembles.

Keywords One-class classifiers · Hybrid OCC ensemble · Fuzzy integral · Density measures · Open-lexicon Arabic word recognition

1 Introduction

Handwritten Arabic word recognition is an active research field due to its interesting use in different applications such as automatic sorting of postal mail, automatic bank check processing, bills processing, passport validation and, recently, for historical document reading via text to speech applications, helping blinds to read and recognizing handwritten historical documents [1–5]. Unlike Latin languages,

Arabic is written from right to left. Also, it has its own diacritical marking such as duma (˙), hamza (ء) and the madda (~). Regarding character shapes, Arabic script has two main properties. On the one hand, several letters share the same shape and differ only in the number and position of dots, such as “djim: ج,” “ha: ح” and “kha: خ.” On the other hand, some letters change their shape according to their position at the beginning, the medial or the end in the word. For instance, the letter “Aïn” can be written through four shapes that are: “ا, آ, ع, إ,” where the two last shapes are related to end positions which change if the word is fully connected or not.

So far, the Arabic word recognition is considered as one of the most challenging tasks of pattern recognition for its specific writing as well as its variability. In this context, the analytical and holistic approaches are the two possible ways for recognizing an Arabic word [6]. The first one consists of segmenting a word image into subwords or isolated characters, which are recognized through character recognition. Generally, this approach is employed when a very large vocabulary is available since it is impossible to construct a

✉ Bilal Hadjadji
bhadjadji@usthb.dz

Youcef Chibani
ychibani@usthb.dz

Hassiba Nemmour
hnemmour@usthb.dz

¹ Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants (LISIC), Faculty of Electronics and Computer Science, University of Science and Technology Houari Boumediene (USTHB), 32, El Alia, Bab Ezzouar, 16111 Algiers, Algeria

specific classifier for each word. Hence, the hidden Markov models (HMM) are probably the most probably used classifiers for solving very large vocabulary lexicon [7]. In contrast, the holistic approach is based on the global analysis where each word is considered as a single unit. This approach is appropriate for problems with large or medium vocabulary such as address postal recognition [8]. In this case, all kinds of classifiers can be used such as the binary support vector machines (SVM) [9, 10]. The main advantage of this approach is related to its efficient capture of the co-articulation and variability effects contained into word images handled by the same classifier [11].

Hence, the present work is focused on the use of the holistic approach to achieve handwritten Arabic word recognition. In this respect, HMM [12], SVM [13], neural networks [14] and hybrid of SVM and conventional neural networks [15] are extensively used. However, such classifiers provide an Arabic word recognition system with closed lexicon, because adding new word to the lexicon requires retraining all the system.

Nowadays, extended multi-class implementation to new classes is strongly required, for instance, in Arabic word recognition and handwritten writer identification. However, the existing classifiers need to retrain the system again on all classes such as the one-against-one (OAO) or one-against-all (OAA) implementations based on the SVM classifiers. Recently, one-class classifier (OCC) has been successfully used to achieve extensible multi-class implementations [16–21]. Indeed, extending the OCC to new classes does not require retraining the used classifiers for a second time. Furthermore, the OCC offers less computational cost in terms of training time and memory space [15, 16].

One-class classifiers are defined as a machine learning that models a restricted domain in a multi-dimensional pattern space using only a set of the target class [22]. Thus, the OCC is considered one among the nearest approach to the human learning for the classification task, due to its ability to learn the model of each class independently of the remaining classes. In this respect, the OCC has been successfully employed in many applications such as image retrieval [23], automated document retrieval and classification [24] as well as combining different biometric traits [25].

In this paper, the OCCs are explored in order to perform an open-lexicon Arabic word recognition system. Generally, pattern recognition systems designed by a single system suffer from limitations such as the lack of uniqueness and non-universality to the problem at hand [26]. Thus, combining multiple systems by taking advantage of each individual and avoiding their weakness may lead to the improvement of classification accuracy. Indeed, the benefits of multiple classifiers based on different features for the same problem have been suitable for various fields of pattern recognition,

including handwritten recognition [27], speech verification [28] and other applications [29].

Recently, it has been demonstrated that combining classifiers can also be effective for OCCs, and therefore, the existing classifier combination strategies can also be applied to OCCs [20]. Since information regarding only one class is available, combining OCCs becomes more difficult. Nevertheless, OCC ensemble has been explored to deal with variety of applications such as anomaly detection and distributed intrusion detection in mobile ad-hoc networks [30], image annotation [17] and other recognition applications. Hence, a distinction should be highlighted between OCC ensemble for solving one-class problems, multi-class implementation and ensemble of multi-class implementation, which represents the hybrid approach.

Usually, the combination step for the hybrid OCC ensemble is performed through the use of simple combination rules such as fixed rules [9, 17, 20, 31], error-correcting output code (ECOC) and decision template (DT) strategies [10, 32, 33]. However, fixed combiners fail in some difficult cases. Fixed rules are optimal for special cases in which the combined systems are similar in terms of performance and competence. Moreover, classifiers designed by various information sources are different from each other, because the members of the ensemble are built of diverse feature spaces [34]. Therefore, trained combiners are more suitable since a priori-knowledge about the ensemble is investigated in the test phase, favorites the more suitable classifier. Thus, the final decision is made by taking into account the competence of each member.

In this respect, a great effort has been done for proposing various combination methods and schemes including methods based on the Dezert–Smarandache theory (DSmT). For instance, Abbas et al. [35] proposed DSmT a new scheme based on the one-class support vector machine (OC-SVM) ensemble trained on different feature sets using the DSmT for handwritten digit recognition. The DSmT shows its superiority in term of performance versus the sum rule. However, the proposed scheme violates the best advantage of using OCCs as multi-class system since the extension to new classes achieves a closed system. Indeed, adding new classes requires updating all parameters and retrains the combination model.

Other alternative combination methods have been proposed based on fuzzy sets. In that case, fuzzy integral (FI) and the associated fuzzy measures initially introduced by Sugeno have been reported to give excellent results for classifier aggregation. Its main advantage is related to measuring the strength of both individual and subsets of classifiers. The ability of the fuzzy integral to enhance the results produced by multiple information sources has been proved in various application areas of pattern recognition [26–29].

In order to achieve an open and powerful hybrid OCC ensemble, Hadjadji et al. [36] proposed a combination scheme based on the use of fuzzy integral operators by studying their abilities against fixed rules. However, their parameters should be adjusted to each time when a new class is added to the system. Besides, the density measure representing the ensemble competence of each member is measured by its training accuracy. Consequently, all testing samples are represented by the same density measure values, which make this approach less efficient since each test sample has its suitable density measure. Therefore, more contribution should be given in respect to the more appropriate information source and each test sample.

To overcome these drawbacks, this paper proposes to investigate an alternative framework to design a parameter-independent open-lexicon Arabic word recognition system as well as a new density measure function. This framework allows providing a dynamic measure for each test sample without the need to measure its performance via the training or validation dataset. Since OCCs have not been evaluated yet for large number of classes, experiments are carried out on large dataset for Arabic handwritten word recognition and different types of OCC, leading to an extended view on the usefulness of proposed framework based on FI to the addressed problem.

Hence, the proposed paper tries to investigate the use of one-class classifiers and FI combination strategies for open-lexicon Handwritten Arabic word recognition. The main advantage of the proposed approach is to offer an open and efficient system, which is the first time that has been performed on handwritten Arabic word recognition. Indeed, the use of closed system to the application of address recognition, as an example, is considered as a shortcoming and therefore open-lexicon system is suitable for updating the list of addresses more efficiently.

The remaining of this paper is organized as follows. Section 2 reports a brief overview on hybrid OCC ensembles. Section 3 describes the mathematical formulation of the proposed hybrid OCC ensemble. In Sect. 4, experimental results are conducted on various types of OCC for handwritten Arabic word recognition with large number of classes in order to prove the effective use of the proposed combination scheme. Finally, the conclusion and future work are provided in the last section.

2 Brief overview on hybrid OCC ensembles

Hybrid OCC ensemble is defined as a multi-class implementation based on single OCC ensembles. It is designed for attempting to enhance the recognition performance and system robustness of the OCC multi-class implementation. As depicted in Fig. 1, the test pattern is assigned to one of

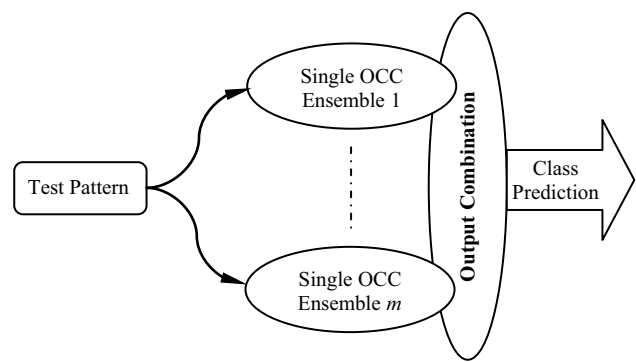


Fig. 1 Hybrid OCC ensemble topology

the predefined classes according to two steps. First, a single OCC ensemble per class is constructed from which outputs are combined through a predefined combination rule. Second, the outputs of multiple single OCC ensembles are combined through another combination rule to provide a multi-class decision and therefore to predict the class label.

Note that, in addition to performance improvement, the hybrid OCC ensemble preserves the property of an open system since it is possible to add new classes without retraining all single OCC ensembles. Hence, this scheme has firstly been explored by Juszczak and Duin [31] for classifying missing data in multi-class problems. In their proposed system, a single Parzen OCC ensemble is trained for each existing class in the training set. Each ensemble contains one classifier for each feature. During the classification step, only the available features are classified, which are combined using a fixed rule. Comparatively to their proposed method offers two advantages. First, it requires fewer classifiers to be trained. Second, it does not require retraining the system whenever missing feature values are occurred. Later, Goh et al. [17] pointed that the SVM is influenced by the problems of noisy and imbalanced training data. As solution, they proposed the use of OC-SVM to estimate support vectors of individual classes. Moreover, bagging scheme has been proposed to reduce class prediction variance. Finally, the overall class prediction of the multi-class implementation is the result of majority voting of the several bags. Furthermore, authors explored the mentioned advantages of OCC for constructing dynamic ensemble which can be used for a new class discovery. Experimental results showed the effectiveness of their proposed system. In a related work, Muñoz-Marí et al. [9] demonstrate that using a simple combination rule (e.g., average or product) on OCCs trained on different feature sets improves the image classification accuracy in remote sensing. They proved that the ensemble composed of the support vector data description achieves better results in comparison with mixture of Gaussian ensemble.

Yeh et al. [10] proposed a combination scheme by using conjointly Adaboost and OC-SVM with a well-designed

discriminant function for solving the multi-class classification problems. Classifiers returned by Adaboost are then aggregated through the sum combination rule. In another work, Krawczyk and Wozniak [32] follow the path of using the diversity measure that has been widely explored for conventional multi-class classifiers to achieve complimentary systems. In their proposed system, OC-SVM ensembles trained on different training sets are performed. Then, a diversity measure is applied on the constructed ensembles in order to select the suitable ones for each class. Finally, the selected classifiers are combined using ECOC [33] and DT [34] strategies to provide a robust system.

Krawczyk and Filipczuk [10] proposed an efficient medical decision support framework that allows distinguishing between benign, malignant and fibro adenoma cases, based on combining OCC trained on different features and the ECOC combination strategy. Experimental evaluation shows the superiority of their proposed system against some state-of-the-art systems. In similar concept, hybrid OCC ensemble has been also investigated for medical image classification [37]. The ensemble consists of one-class KPCA models trained on different features generated from each image class. Besides, a new product combination rule was proposed. The effectiveness of their proposed classification scheme was verified using a breast cancer biopsy image dataset and a 3D optical coherence tomography retinal image set. The proposed classification scheme provides competitive results on the two medical image sets compared to the state-of-the-art systems. Recently, in a related work Krawczyk et al. [38] presented a new method to design the hybrid OCC ensemble based on a single OCC ensemble. The main idea is to create single OCC ensemble for each class based on feature space partitioning. The combined classifiers are trained on the basis of clusters that lead to make use of individual classifier strengths. Experiments conducted on a wide range of benchmark datasets prove the validity and the flexibility of the proposed framework to work with different clustering algorithms. In order to improve their system, Cyganek and Krawczyk [33] proposed to split data using the nonnegative matrix factorization algorithm with sparse constraints. This framework relies on splitting the input data into compact and consistent clusters as well as automatic determination of the cluster's number. The proposed method shows high accuracy and fast classification.

3 Proposed hybrid OCC ensemble based on fuzzy integral

The hybrid system is composed of different single OCC ensembles, where each one is dedicated to represent one class from the set of classes. Consequently, each class is represented by a combination of different OCCs. Various combination

rules are possible for achieving an enhanced hybrid OCC ensemble. The present paper proposes to investigate the fuzzy integral (FI) operators for combining multiple single OCC ensembles.

In the following, the main concepts of the fuzzy integral and associated fuzzy measures are reviewed for a better clarity. Then, the hybrid system is described for combining multiple single OCC ensembles for multi-class classification by means of FI.

3.1 Background on fuzzy measure and fuzzy integral operators

Fuzzy integrals are nonlinear combiners defined with respect to fuzzy measures. Therefore, the main advantage of FI is its ability to combine the objective evidences in the form of expert decisions taking into account subjective evaluation of their competence expressed by the fuzzy measure. The present section reviews the main properties of fuzzy measures in addition to the used fuzzy integral operators.

3.1.1 Fuzzy measure and objective evidence

A measure is defined to express the importance of each information source and of each possible coalition of information sources. For a majority of applications when considering diverse information sources together they manifest some sort of positive/negative synergy. Therefore, the additive property of the measure may result too restrictive. In order to overcome this drawback, Sugeno introduced the concept of *fuzzy measure*.

Let $Z = \{z_1, \dots, z_L\}$ be a finite set of information sources, a set function $g(Z) : 2^Z \rightarrow [0, 1]$ is called fuzzy measure if it verifies the following properties [39, 40]:

1. $g(\phi) = 0, g(Z) = 1$
2. If $A, B \subset 2^Z$ and $A \subset B$, then, $g(A) \leq g(B)$
3. If $A_n \subset 2^Z$ for $1 \leq n \leq \infty$ and the sequence $\{A_n\}$ is monotone in the sense of inclusion then,

$$\lim_{n \rightarrow \infty} g(A_n) = g\left(\lim_{n \rightarrow \infty} A_n\right) \tag{1}$$

These properties show that the fuzzy measure is not necessary additive; therefore, if $A, B \subset Z$, and $A \cap B = \phi$, then:

$$g(A \cup B) \neq g(A) + g(B) \tag{2}$$

According to this inequality, Sugeno (1977) introduced the decomposable so-called λ -fuzzy measure satisfying the following additional property for $\lambda > -1$:

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \tag{3}$$

The λ -value can be defined as an interaction parameter between two sets A and B .

Let $Z = \{z_1, \dots, z_L\}$ be the set of information sources to be aggregated for attempting to achieve better performance. In this case, each information source z_i is then associated with a density measure $g^i = g(\{z_i\})$, which represents generally the performance of a single source. In this case, Sugeno defines a function $h : Z \rightarrow [0, 1]$ called an objective evidence with respect to a fuzzy measure g over Z . The set of objective evidences, denoted $h(z_i)$, is then rearranged from maximum to the minimum value as $h(z_1) > \dots > h(z_L)$ as well as their corresponding density measures g^i . Consequently, according to this order, the λ -fuzzy measure $g(A_i)$ of the new sequence $A_i = \{z_1, \dots, z_i\}$ can be computed recursively as follows:

$$g(A_1) = g(\{z_1\}) = g^1 \tag{4}$$

$$g(A_i) = g(A_{i-1}) + g^i + \lambda g(A_{i-1})g_i, \quad 2 \leq i \leq L \tag{5}$$

The value of λ is deduced by solving the equation $g(Z) = 1$, which corresponds to resolve the following equation:

$$\lambda + 1 = \prod_{i=1}^L (1 + \lambda g_i) \tag{6}$$

This value is obtained as the unique real root greater than -1 and not equal to zero.

3.1.2 Fuzzy integral operators

Sugeno integral I_s is the first FI operator defined for aggregating the objective evidence $h(z_i)$ with respect to a fuzzy measure $g(A_i)$ over Z . It is computed as:

$$I_s = \max_{i=1}^L [\min(h(z_i), g(A_i))] \tag{7}$$

This integral represents the simplest operator for FI combination. Hence, it has been extended through the definition of the Choquet integral. Its discrete formulation for a function $h : Z \rightarrow R^+$ with respect to fuzzy measure g is defined as:

$$I_C = \sum_{i=1}^L \{h(z_i) - h(z_{i-1})\} g(A_i) \tag{8}$$

where indices i have been permuted so that $0 \leq h(z_1) \leq \dots \leq h(z_L) \leq 1$ and $h(z_0) = 0$.

Attempting to enhance more efficiently the FI, other efficient operators have been proposed such as ordered weighted averaging (OWA). Two commonly versions are used, which are OWA [39] namely OWA-AND and OWA-OR,

respectively. The OWA-AND requires the calculation new evidences $\hat{h}(z_i)$ defined as:

$$\hat{h}(z_i) = \frac{(1 - \alpha)}{l} \sum_{i=1}^l h(z_i) + \alpha \min_{z_i \in Z} \{h(z_i)\} \tag{9}$$

These new evidences are then used into Sugeno integral, which is termed I_{S-AND} . In addition, this paper proposes to use these new evidences into the Choquet integral termed I_{C-AND} in order to evaluate its performance.

In contrast, the OWA-OR is performed using the same new evidences through the new decision function defined as follows:

$$I_{OR} = \frac{1 - \beta}{2^L} \sum_{i=1}^L \min(\hat{h}(z_i), g(A_i)) + \beta \max_{i=1}^L [\min(\hat{h}(z_i), g(A_i))] \tag{10}$$

Both operators need tuning parameters α and β in the unit interval in order to achieve better results than the Sugeno and Choquet operators [40]. In the following, the five different operators for FI are termed Sugeno (I_s), Choquet (I_C), S-AND (I_{S-AND}), C- AND (I_{C-AND}) and OR (I_{OR})

3.2 Hybrid OCC ensemble system

The hybrid OCC ensemble depicted in Fig. 2 is composed of m classes and L different information sources. Therefore, each class is represented by a single OCC ensemble which is composed of L OCCs trained on different information sources. Their normalized outputs are aggregated through a FI operator. Finally, the class label of the test pattern is assigned to the single OCC ensemble that achieves the maximum prediction.

Let $\{D_i, i = 1, \dots, m\}$ be the set of m single OCC ensembles and denote $D_i = \{d_{ij}, j = 1, \dots, L\}$ as the output vector composed of the output value d_{ij} provided by the OCC_{ij} trained on the i th information source of the j th class. The set of the output values can be represented in a matrix as follows:

$$D = \begin{pmatrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ D_m \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1L} \\ d_{21} & d_{22} & \dots & d_{2L} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ d_{m1} & d_{m2} & \dots & d_{mL} \end{pmatrix} \tag{11}$$

Several combination rules are possible to achieve the hybrid OCC ensemble, but all these rules need a unique interpretation of the outputs generated by the different classifiers for each test pattern x . Hence, the normalization of outputs for each classifier is required to perform a correct

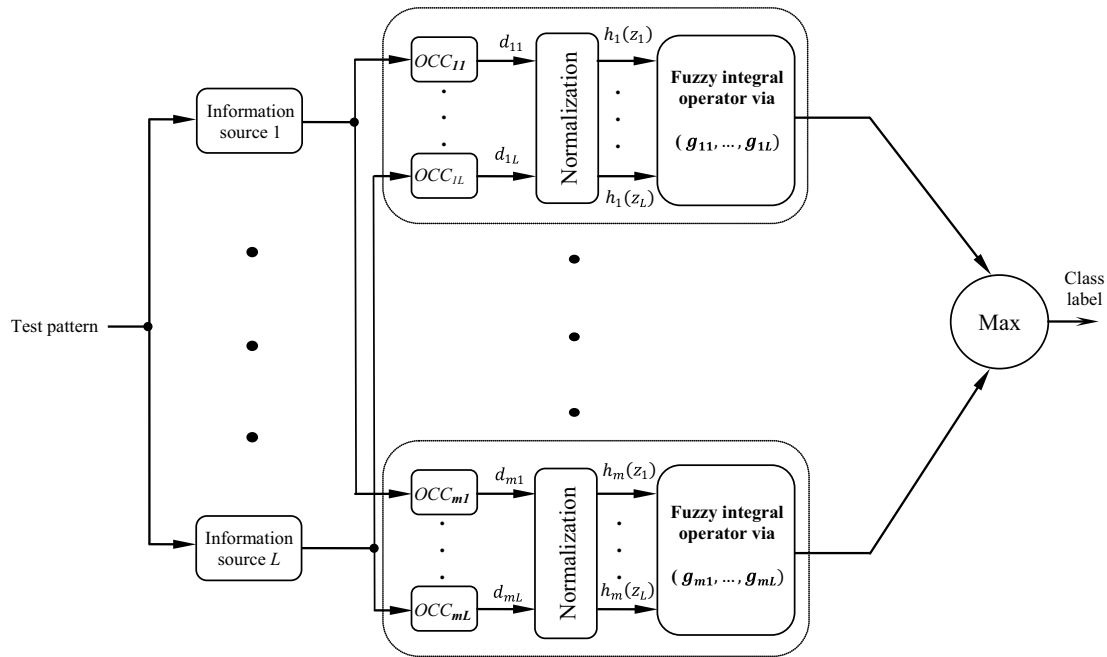


Fig. 2 Hybrid OCC ensemble scheme based on fuzzy integral

the combination. Presently, the simple exponential function is used for transforming the OCC output d_{ij} ranging from $]-\infty, 0]$ to $]0, 1]$ using the posteriori probability $P_i(c_j/x)$ as follows:

$$P_i(c_j/x) = \exp(d_{ij}(x)) \tag{12}$$

The evidence is then expressed as the posteriori probability taking the following form:

$$h_j(z_i) = P_i(c_j/x) \tag{13}$$

The successful key of FI depends on the appropriate formulation of the density measure associated to each information source. Consequently, if the density measures are well formulated then the fuzzy measures can be correctly defined, in order to make a well aggregation of the fuzzy integral. Generally, the density measures represent the competence of the ensemble member measured by its accurate performance. The use of the performance deduced from the training datasets is not well representative since it requires a new validation dataset to get a better evaluation of each information source. Moreover, all testing samples are represented by the same values, which make this approach less efficient. In fact, more contribution should be given in respect to the more appropriate information source and each test sample. To overcome this limitation, we propose an alternative approach in which

the density measure is considered as the similarity degree or correspondence between the class model and that of the test pattern, i.e., the closer the similarity is, the greater the value of the density measure is. Precisely, we introduce a new dynamic density measure, which is defined for each test sample as:

$$g_{ij}(x) = \exp(-\delta d_{ij}(x) - dm_{ij}^2) \tag{14}$$

Such that, $0 \leq g_{ij}(x) \leq 1$ and $0 < \delta \leq 1$.

The dm_{ij} is obtained by averaging values of the OCC_{ij} outputs using the training samples. δ is a positive value that has been introduced for calibrating more efficiently the density and therefore representing more accurately the contribution of each classifier and information source. Once accurate density measures obtained, fuzzy measures are deduced according to Eqs. 4–6.

Let $y(x)$ be the class label of a test word x . A specific FI is applied to combine the three classifier outputs for each class. Finally, the test word is recognized according to the maximum FI obtained score over all classes as follows:

$$y(x) = \arg \max_{j=1}^m (FIOP_j(h_j(z_i), g_j(A_i))), \quad 1 \leq i \leq L \tag{15}$$

FIOP is one of the defined FI operators.

In summary, the multi-class classification using the FI is performed according to the Algorithm 1

Algorithm 1:**Inputs:** Image word I , classifier models OCC_j^i , average training outputs dm_{ij} **Output:** Class label

```

1: for  $i \leftarrow 1$  to  $m$  do      /*  $i$  represents class variable */
2:   for  $j \leftarrow 1$  to  $L$  do  /*  $j$  represents the feature generation method variable */
3:     Calculate the word feature vector  $x$  of the image word  $I$ 
4:     Calculate the classifier output  $d_{ij}(x)$ 
5:     Compute the normalized output  $P_i(c_j/x)$  and the evidence value  $h_j(z_i)$  according to Eq. 12 and 13.
6:     Calculate the density measure according to Eq. 14
7:   end for
8:   Determine the fuzzy measures  $g_j(A_i)$ , for  $1 \leq i \leq L$  and  $1 \leq j \leq m$  using Eqs. 4, 5 and 6.
9:   Perform FI operator to determine the combination score via the evidence and fuzzy measure
10: end for
11: Assign the unknown word to the word class that provides the maximum FI combination score according to Eq. 15.

```

4 Experimental results

4.1 Dataset description

In order to evaluate the proposed approach on large number of classes, the well-known IFN/ENIT dataset [41] is selected containing more than 26,400 images of Tunisian town names written in Arabic script. Words are written by 411 writers using different writing tools. The IFN/ENIT is composed of four folds, A, B, C and D. Usually, results are carried out using three folds for training and one for testing. The present work is evaluated using 300 classes each one is trained by considering more than 10 samples per class. All results are reported in terms of the classification accuracy expressed in percentage (%).

4.2 Information source generation

Various techniques have been proposed for generating features from the word image [42]. Recently, the deep learning technique has been successfully used for representing handwritten words. Indeed, Bluche et al. [43] has that learning features with convolutional neural networks (CNN) is better than using hand-crafted features for handwritten word recognition. In other work, Poznanski and Wolf [44] employed the CNN to estimate its n-gram frequency profile, which is the set of n-grams contained in the word.

In this paper, the curvelet transform (CT) is used for its enhanced directional capacity to characterize edges and singularities along curves that compose handwritten Arabic word [44]. CT has been employed in various applications such as image denoising [45, 46], face and facial expression recognition [47, 48], compression [49], texture classification

[50], content-based image retrieval [51], character recognition [52, 53]. Recently, the CT has been successfully used for offline handwritten signature retrieval [54] and verification [55].

Presently, CT is first performed on the word image via the wrapping technique at different scales and different orientations to generate curvelet coefficients. The resulting ones are used for computing the energies, which allow characterizing the importance of the curvature contained into the word image. In order to capture more efficiently the local information, CT is performed on different sections of the word image grid. Finally, the feature vector is achieved by concatenating all computed wedge energies for the defined image sections.

For the combination stage, different information sources should be generated from the word image [35, 41]. Also, three different ways are used for partitioning an image into variety of sections (before applying the curvelet transform), namely equispaced, equimass and equimass adaptive grids, respectively. A uniform or equispaced grid [56] creates rectangular regions for sampling, where each one has the same size and shape. It is performed via placing the grid lines at equally spaced locations along the x-axis of the word image creating the vertical regions. Similarly, the horizontal regions are produced by placing grid lines at equally spaced locations down the y-axis. Conversely, an equimass grid creates different rectangular regions having the same number of black pixels, also known as the mass, of the word image [56]. Consequently, each region is found by partitioning horizontally and vertically the word image using its mass histogram. Hence, the total mass between all adjacent points on either the x-axis or the y-axis are as close to equal as possible. Additionally, the equimass adaptive is used, which is

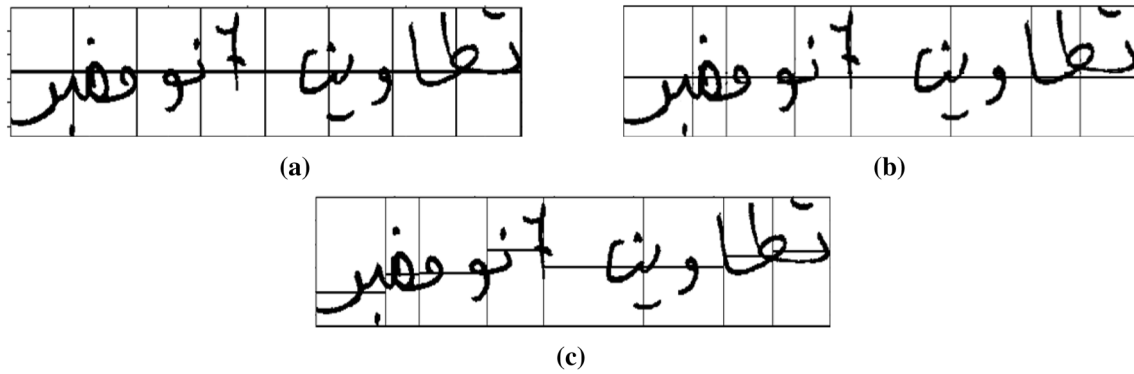


Fig. 3 Arabic handwritten word decomposition having the same grid size 2×8 : **a** equispaced grid, **b** equimass grid and **c** equimass adaptive grid

a small modification of the equimass grid. It is based on the computation along the x-axis for each horizontal region and not for the entire image as in equimass grid. Figure 3 shows an example of partitioning the handwritten word image using three methods.

4.3 Experimental design

To evaluate the hybrid OCC ensemble parameter-independent and open-lexicon Arabic word recognition system, experiments are conducted into two steps: design step and evaluation step. For the design step, a number of classes are randomly selected from the whole dataset in order to deduce the optimal parameters for feature generation, training parameters for OCCs and the combination model. In the evaluation step, the remaining classes are used for evaluating the robustness of the proposed system taking the same parameters found during the design step. In other words, when a new class is added to the hybrid OCC ensemble for evaluation, the same parameters are used, as they have been tuned during the design step.

The hybrid OCC ensemble is composed of three OCCs for each class; each one receives its own feature vector according to the selected feature generation method. Hence, each OCC is separately trained on its own information source. In order to have a large view on the usefulness of the proposed architecture, experiments carried out on different types of OCCs which are principal component analysis (PCA), K-means and nearest neighbor (NN) [22].

Therefore, different systems are built according to the used type of OCC. These classifiers are selected for their success in many applications and for the reduced numbers of parameters to be tuned during their training.

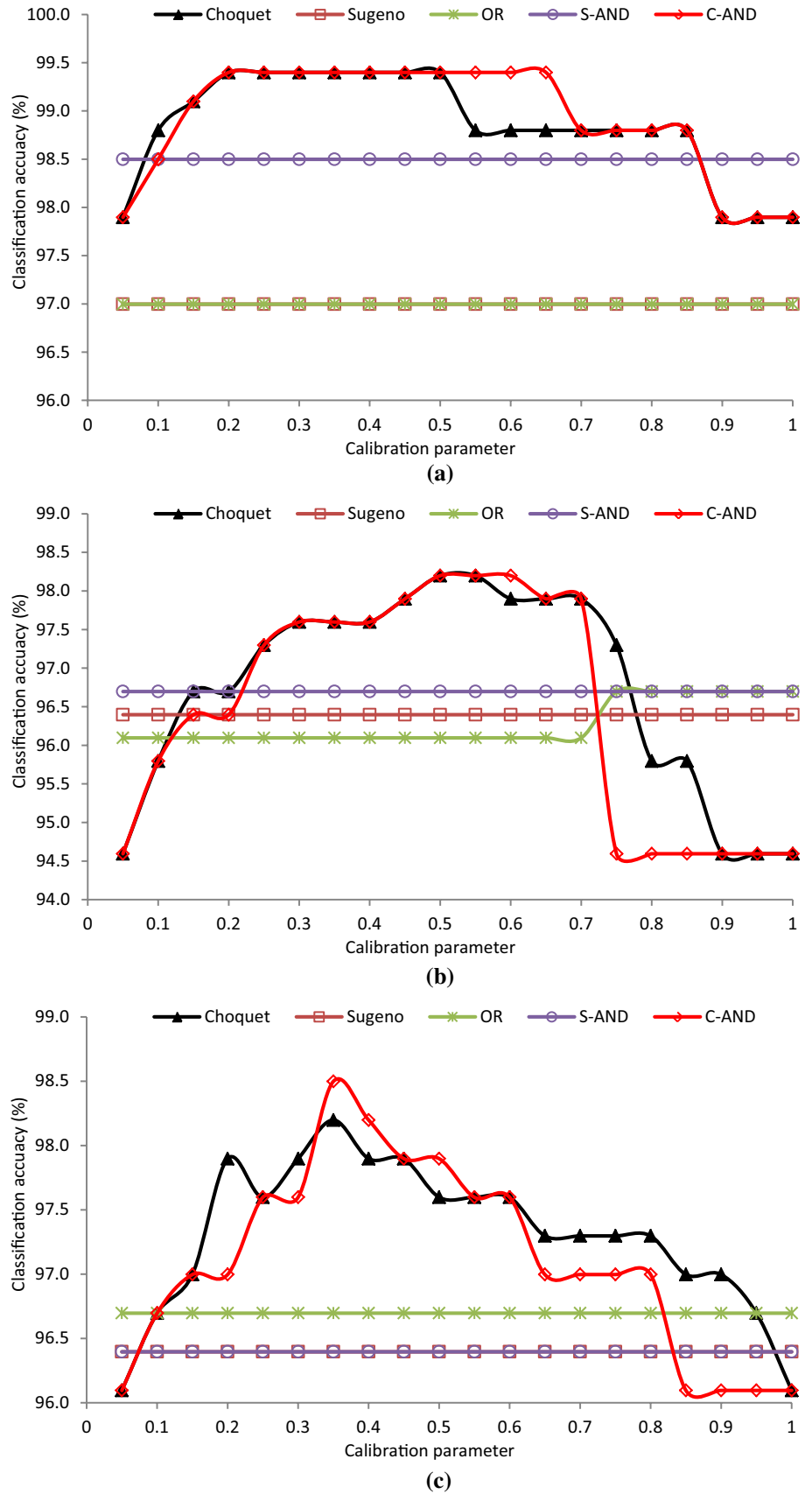
During the design step, the proposed open classification system is highly affected by the grid size and the classifier parameters. For the NN classifier, no parameters are required to be tuned unlike to other classifiers. For the K-Means, an only parameter should be carefully tuned corresponding to the number of cluster. The PCA also needs an only parameter corresponding to the number of eigenvectors. For finding the optimal parameters of the grid size and classifiers, 10 classes are selected randomly having more than 10 samples per class. To build the OCC models, the set of samples for each class is divided into two subsets namely Subset 1 and Subset 2, respectively. Subset 1 is used for training the OCCs, while Subset 2 is used for finding the optimal parameters of the grid size and classifiers.

Results expressed in terms of classification accuracy for three information sources with different grid sizes obtained with the best classifier parameters are reported in Table 1. As can be seen, the grid size affects significantly the classification accuracy whatever the used classifier. Indeed, when the grid size increases, the classification accuracy is also enhanced. Therefore, the suitable grid parameters are selected to achieve the best accuracy. In this case, the grid size 2×8 offers the best accuracy and optimal feature vector size for all classifiers and for three grid types. Therefore, this grid size is selected for all next experiments.

Table 1 Classification accuracy (%) of individual OCCs with various equispaced (ES), equimass (EM) and equimass adaptive (EA) grid sizes

Grid size	# Features	PCA			K-means			NN		
		ES	EM	EA	ES	EM	EA	ES	EM	EA
2×2	192	80.78	84.98	86.78	81.98	81.38	82.23	74.17	74.17	74.40
2×4	384	91.89	93.39	93.99	89.18	91.29	91.89	81.38	84.38	85.70
2×6	576	94.29	95.19	95.79	92.49	92.49	95.19	84.08	86.78	89.18
2×8	768	95.79	96.09	96.69	93.09	93.99	96.69	93.09	93.99	96.39

Fig. 4 Effect of calibration parameter δ on the different operators: **a** PCA, **b** K-means and **c** NN classifiers



Once designing the individual classifier, the combination is performed by means of the FI operators. As already mentioned, the use of FI requires a careful tuning of parameters, which are the calibration parameter (δ) related to the fuzzy densities according to Eq. 14, and the couple (α, β) related to OR, C-AND and S- AND operators, respectively. Similarly to individual systems, suitable parameter (δ, α, β) values are tuned using also the 10 classes. Results for the used operators with different values of the calibration parameter (δ) ranging from 0.1 to 1 with a step 0.1 are depicted in Fig. 4. Moreover, for OR, C-AND and S-AND operators, the presented results are obtained with best parameters of the couple (α, β) varied in the range [0, 1]. From the presented results, it is worth noting that the Choquet and C-AND are the best and the most suitable FI operators. However, they are highly affected by the calibration parameter δ . Therefore, a careful tuning should be carried in order to achieve the best performance. For instance, for the NN classifier, C-AND operator achieves a classification accuracy from 96.09 to 98.49% when varying the calibration parameter in the range [0.1, 1], which justifies its introduction to the proposed density measure function. Table 2 reports the optimal parameters (δ, α, β) selected during the design step to achieve the best classification accuracy.

4.4 Combination results

Results for the individual classifiers and the hybrid OCC ensembles with different combination rules according to the selected OCC are reported in Table 3. When comparing the individual classifiers, we can note that PCA is the most suitable for this application. Moreover, for a fairly comparison and analysis of the achieved combination schemes, various FI operators are evaluated against fixed combiners including

Table 2 Optimal parameters (δ, α, β) selected for each operator and OCC during the design step

Classifier	Operator		
	S-AND (δ, α)	C-AND (δ, α)	OR (δ, α, β)
PCA	(0.05, 0.3)	(0.2, 0.3)	(0.05, 0.2, 0.6)
K-means	(0.05, 0.5)	(0.5, 0.5)	(0.75, 0.3, 0.5)
NN	(0.05, 0.8)	(0.35, 0.8)	(0.05, 0.6, 0.3)

average (Avg), product (Prod), minimum (Min) and maximum (Max) [33,55, 57]. Obtained result reveals that combining information sources allows improving considerably the classification accuracy than the single source for all classifiers. For instance, when using the hybrid PCA for the test set A, the classification accuracy is improved by more than 7% against the best single information source, which confirm the effectiveness of using multiple systems than using a single one. Secondly, for the combination strategy, we can clearly observe that the average and product are the best combiners form the fixed group and on the other side, we find that Choquet is the best form the FI group. Besides, when comparing the fixed group against the FI one, we can note that the Choquet and its extension C-AND combiners offer an improved recognition rate than the best fixed aggregators. Therefore, FI combiners are more suitable than fixed ones for hybrid OCC ensemble, Choquet integral seems the most suitable for achieving the hybrid OCC ensemble, since it yields better results with C-AND without any parameter.

In order to have extended view about the performance of the proposed system, results are provided in Table 4 with

Table 4 Classification accuracy (%) of individual classifiers and hybrid one-class ensembles with different combination rules for 300 classes according to Top-N

	OCC ensemble	Sources			Fuzzy integral operator
		ES	EM	EA	Choquet
Top-1	PCA	71.14	73.32	76.26	83.73
	K-means	60.43	62.65	67.35	77.64
	NN	60.22	62.69	68.15	77.55
Top-2	PCA	79.87	80.16	82.88	87.66
	K-means	68.67	68.96	71.22	78.00
	NN	68.15	69.12	71.08	77.62
Top-3	PCA	87.98	90.19	91.97	92.60
	K-means	71.75	72.47	74.32	80.29
	NN	71.22	72.32	74.85	80.16
Top-4	PCA	90.35	91.60	93.25	95.70
	K-means	74.15	74.90	76.74	82.84
	NN	73.85	75.00	76.57	82.78
Top-5	PCA	91.95	92.66	94.88	96.17
	K-means	75.57	76.59	78.38	83.96
	NN	75.48	76.80	78.67	83.73

The bold result defines the best classifier

Table 3 Classification accuracy (%) of individual classifiers and hybrid one-class ensembles with different combination rules for 300 classes

OCC ensemble	Sources			Fixed rules				Fuzzy integral operators				
	ES	EM	EA	Avg	Prod	Max	Min	Sugeno	C-AND	OR	Choquet	S-AND
PCA	71.14	73.32	76.26	82.02	82.02	71.11	77.27	79.91	81.86	81.84	83.73	78.18
K-means	60.43	62.65	67.35	76.43	76.32	62.61	69.07	70.51	75.12	71.43	77.64	71.88
NN	60.22	62.69	68.15	76.26	76.21	62.64	69.75	72.45	75.67	73.23	77.55	73.22

different types for each sort of used OCC to perform the open hybrid OCC ensembles. From the obtained results, we can obviously notice the effect of the selected top for the word recognition. Furthermore, the OC-PCA-based open hybrid OCC ensemble appears to perform the best word recognition accuracy, and provides 83.73 and 96.17% for Top-1 and Top-5, respectively.

In order to show the effective use of the Choquet operator against the remaining combination rules, we use the McNemar’s test [58], which allows comparing statistically two systems. More precisely, a contingency table is constructed in order to calculate the p value.

The McNemar’s test has the ability to provide whether one system is significantly better than another according to the p value. A small p value indicates a significant difference of the classification accuracy between two systems to be compared. In contrast, when the p value exceeds 0.05 then the null hypothesis is considered. In this case, both systems perform closely and the difference is too small to decide the superiority of one system than the other.

Table 5 reports the p value of the Choquet operator against the remaining combination rules for PCA, K-means and NN classifiers according to Top-1. The obtained results show that the Choquet operator is significantly different from the other combination rules for the different type of classifiers except for C-AND operator, where small difference can be observed in some cases for which p values exceed 0.05.

McNemar’s test evinces that Choquet and C-AND are more robust than the fixed rules whatever the used OCC and datasets. Indeed, the obtained p values are very small which shows the high difference between the different aggregators.

These findings demonstrate the success of Choquet operator for improving the results of the hybrid OCC ensemble, which is due to its ability to capture interactions among the OCCs and attribute the right importance for each information source. According to the definition of Choquet FI, the appropriate weight values are dynamically deduced from the fuzzy measure dynamically for each test pattern which represents its successful key. Moreover, the proposed method for dynamic density measure, which is associated with fuzzy measure, offers a better adaptation allowing to give more importance to the relevant information source relatively to the others.

4.5 Stability assessment of combination rules

The proposed hybrid OCC ensembles using FI operators are evaluated to show the behavior of the parameter-independent open classification when new classes are progressively added to the system. Also, we use the stability criterion, which defines the ability of an OCC or hybrid OCC ensemble to maintain the same classification accuracy when the number of classes is progressively increased in the system.

Hence, the classification accuracy is computed for the hybrid OCC ensemble by adding progressively new classes each time from 10 to 300 using parameters found in the design step. Figure 5 depicts the classification accuracy versus the number of classes achieved by the Choquet, C-AND and Sugeno operators against best fixed rules average and product to perform the hybrid OCC ensembles. Roughly speaking, we clearly can notice that the most combiners achieve similar classification accuracy for few classes. However, when adding progressively new classes, the used combination methods behave differently for more complicated problem. When comparing the set of aggregators, the Choquet is the most stable and widely better than the fixed combiners since they keep the same classification accuracy while adding new classes. Indeed, the gap between the best and the worst aggregators increases progressively when extending the hybrid OCC ensemble for more complex problem via adding new classes. Consequently, the Choquet operator provides the best performance and are the least affected when adding new classes. Conversely, surprising results are achieved by Sugeno FI, which is highly affected by the number of classes and shows its inappropriate use for combining hybrid OCC ensembles.

4.6 Discussions

Nowadays, extended multi-class implementation to new classes is strongly required, for instance, in Arabic word recognition and handwritten writer identification. Indeed, the use of OCCs for solving the multi-class classification problem has been discussed by many researchers due to its offered properties. This type of classifier attempts to model each class separately from the others that allows designing an open multi-class system. This property is highly desirable for the actual systems since it is possible to add new classes without retraining the classification system again

Table 5 The p values of McNemar’s test for Choquet FI versus the other combination rules for hybrid OCC ensembles according to Top-1

OCC ensemble	Fixed rules				Fuzzy integral operators			
	Avg	Prod	Max	Min	C-AND	OR	Sugeno	S-AND
PCA	8.2810^{-13}	2.2810^{-13}	1.1710^{-8}	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
NN	$1.50 10^{-7}$	$2.16 10^{-8}$	$< 10^{-16}$	$< 10^{-16}$	$3.50 10^{-5}$	$9.74 10^{-14}$	$< 10^{-16}$	$< 10^{-16}$
K-means	1.77×10^{-6}	2.01×10^{-7}	$< 10^{-16}$	$< 10^{-16}$	$4.28 10^{-5}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$

Fig. 5 Stability study of combination rules for hybrid one-class ensembles performed on three OCCs: **a** PCA, **b** K-means, **c** INN

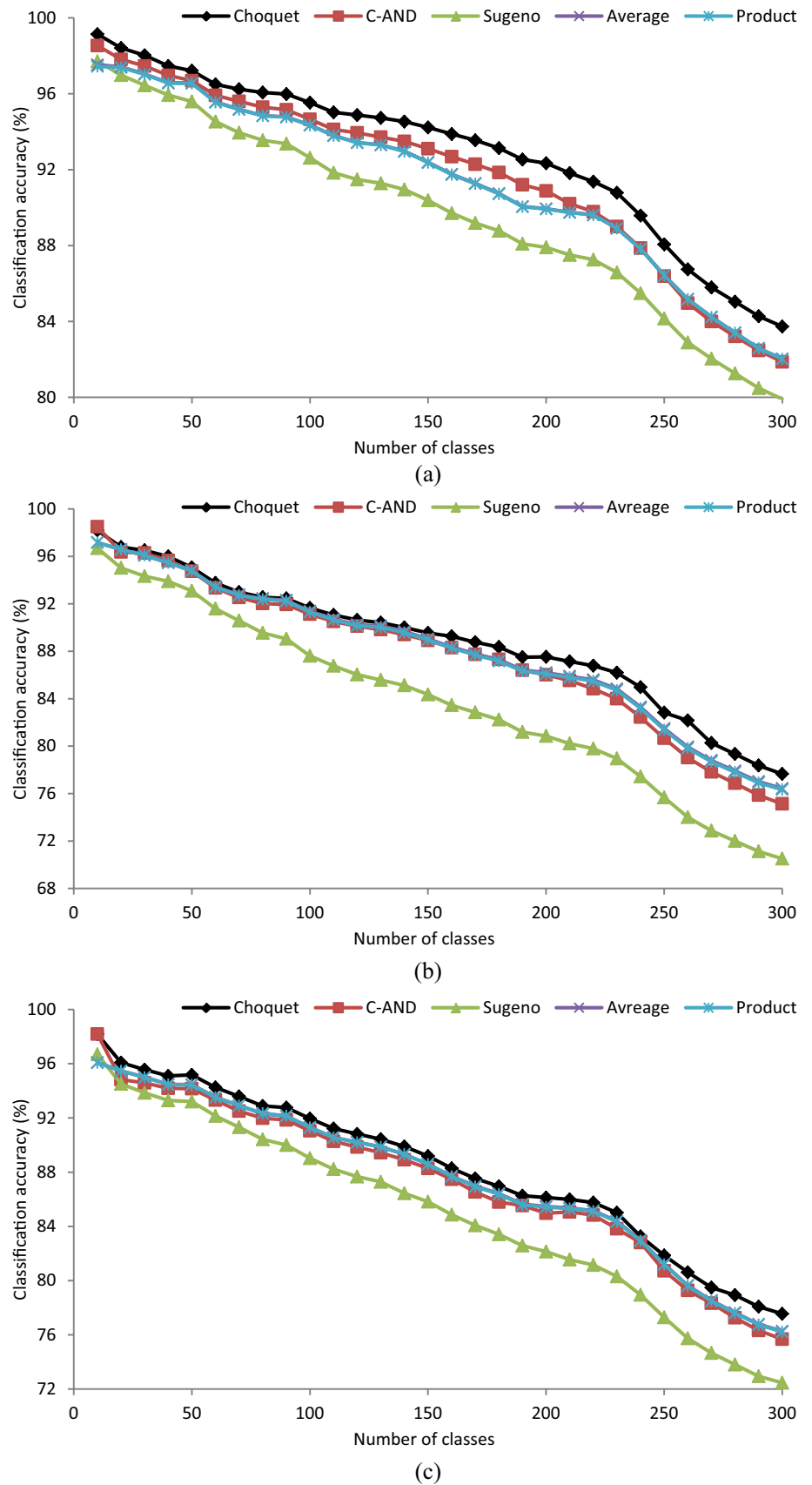


Table 6 Classification accuracy (%) achieved by the proposed open-lexicon handwritten word recognition based on OCC comparatively to other state-of-the-art systems use SVM classifier

Reference	#Classes	Features	Classifier	Accuracy (%)
Nemmour and Chibani [59]	24	Ridgelet transform	SVM	84.00
Khalifa and Bingru [60]	56	SIFT	SVM	91.70
Alalshekmubarak et al. [61]	24	Uniform grid	SVM	95.27
	56			92.37
Proposed system	24	Curvelet	OC-PCA	98.21
	56			96.70
	300			83.73

on the whole classes. However, using individual OCC does not allow designing a robust classification system. Therefore, the design of hybrid OCC ensemble is required in order to achieve the best possible classification performance and robustness as well as keeping the properties of OCC. Thus, this paper discusses the appropriate use of the FI for best designing of the open-lexicon handwritten Arabic word recognition system. Various works reported that FI provides excellent results for classifier aggregating. However, when adapting FI for the hybrid OCC ensemble, a problem is faced which is the generation of the density measure. Indeed, it is usually estimated using the training datasets. However, this estimation is considered not representative and leads to require new validation dataset to get a better evaluation of each information source. Furthermore, all testing samples are represented by the same density values, which make this approach less efficient since each test sample has its suitable density measure. Therefore, more contribution should be provided in respect to the more appropriate information source and each test sample. Thereby, a dynamic density measure is proposed having the ability to attribute the appropriate values to each information source as well as its adaptation for each test pattern.

In order to design a parameter-independent system, a new framework is proposed that relies on using a separated datasets for tuning parameters and the selection of optimal parameter values. This is an efficient property since, once the parameters are found, they are considered to be the same for all existing classes and also for the new ones.

In this step, we have seen the effect of all parameters including descriptor, classifier, density measure and OWA operators. From the set of parameters, we notably mention the impact of the calibration parameter used in the density measure to enhance the combination performance. It can be also noticed that the calibration parameter is more suitable for the Choquet and C-AND operators. From all experimental results we clearly notice the achieved improvement by the use of the FI combination scheme based on Choquet operators against the other operators and fixed rules. More precisely, the Choquet FI shows its stability to maintain the classification accuracy roughly

stable when adding new words to the lexicon against the other combination rules.

In order to situate the proposed open-lexicon handwritten word recognition system based on OCC, a comparative study has been performed versus some recent studies those have explored the SVM classifier for the IFN/ENIT dataset.

Table 6 reports the classification accuracy achieved by the open-lexicon system versus other systems using SVM classifier. The results obtained reveal that the open-lexicon system based on the OC-PCA offer better results against the other systems use SVM classifier in terms of classification accuracy, as well as the advantage of keeping the system open for new class words. Moreover, we should also highlight the impact of FI for improving the results and its good adaptation with the OCC to keep the concept of open lexicon. This would suggest that using OC-PCA may be more representative for Arabic words than the SVM classifier [59–61].

5 Conclusion

This paper has investigated the usefulness of combining OCCs for open-lexicon handwritten Arabic word recognition. The main advantage of the proposed approach is related to an offered open system, which is the first time that has been applied for handwritten Arabic word recognition. Indeed, for instance, the use of closed system for address recognition can be considered as an inconvenient and therefore an open-lexicon system is suitable for updating the list of addresses more efficiently. Furthermore, due to the fact that fixed rules are the standard used combiners for the hybrid OCC ensemble, the proposed work attempted to study the potential of FI operators via proposing a combination scheme for combining ensemble of OCCs designed by different feature generation methods.

Experimental analysis is conducted on different types of OCC and Arabic handwritten word datasets having high number of classes. The results prove the superiority of FI against fixed combiners those represent our base of comparison, whatever the selected type of OCC. Furthermore, the Choquet operator seems to be the most suitable and powerful among FI aggregators. Thus, this study suggests keeping

fuzzy integral operators as a way for achieving robust hybrid OCC ensembles for performing open-lexicon handwritten Arabic word recognition.

For future work, we plan to propose a new architecture of the hybrid OCC ensemble, which relies on the dissimilarity learning approach in order to generate a generic model to perform the open-lexicon classification system.

References

1. Lawgali A (2015) A survey on Arabic character recognition. *Int J Signal Process Image Process Pattern Recognit* 8(2):401–426
2. Shatnawi M (2015) Off-line handwritten Arabic character recognition: a survey. In: *Proceedings of the international conference on image processing, computer vision, and pattern recognition (IPCVC)*
3. Alginahi YM (2013) A survey on Arabic character segmentation. *Int J Doc Anal Recognit (IJ DAR)* 16(2):105–126
4. Likforman-Sulem L, Sigelle M (2008) Recognition of degraded characters using dynamic Bayesian networks. *Pattern Recognit* 41(10):3092–3103
5. Ye Q, Doermann D (2014) Text detection and recognition in imagery: a survey. *IEEE Trans Pattern Anal Mach Intell* 37(7):1480–1500
6. Farah N, Souici L, Sellami M (2005) Classifiers combination and syntax analysis for Arabic literal amount recognition. *Eng Appl Artif Intell* 19(1):29–39
7. Knerr S, Augustin E, Baret O, Price D (1998) Hidden Markov model based word recognition and its application to legal amount reading on French checks. *J Comput Vis Image Underst* 70(3):404–419
8. Plötz T, Fink GA (2009) Markov models for offline handwriting recognition: a survey. *Int J Doc Anal Recognit* 12(4):269–298
9. Muñoz-Marí J, Camps-Valls G, Gómez-Chova L, Calpe-Maravilla J (2007) Combination of one class remote sensing image classifiers. In: *International geoscience and remote sensing symposium*, pp 1509–1512
10. Krawczyk B, Filipczuk P (2014) Cytological image analysis with firefly nuclei detection and hybrid one-class classification decomposition. *Eng Appl Artif Intell* 31:126–135
11. Vinciarelli A (2002) A survey on off-line cursive word recognition. *J Pattern Recognit Soc* 35(7):1433–1446
12. Jayech K, Mahjoub MA, Amara NEB (2016) Synchronous multi-stream hidden markov model for offline Arabic handwriting recognition without explicit segmentation. *Neurocomputing* 214:958–971
13. Hmeidi I, Hawashin B, El-Qawasmeh E (2008) Performance of KNN and SVM classifiers on full word Arabic articles. *Adv Eng Inform* 22:106–111
14. AlKhateeb JH, Pauplin O, Ren J, Jiang J (2011) Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition. *Knowl Based Syst* 24:680–688
15. Elleuch M, Maalej R, Kherallah M (2016) A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition. In: *International conference on computational science*, vol 80, ICCS 2016, pp 1712–1723
16. Sun BY, Huang DS (2003) Support vector clustering for multi-class classification problems. In: *The congress on evolutionary computation*, Canberra, Australia, pp 1480–1485
17. Goh KS, Chang EY, Li B (2005) Using one-class and two-class SVMs for multiclass image annotation. *IEEE Trans Knowl Data Eng* 17(10):1333–1346
18. Ban T, Abe S (2006) Implementing multi-class classifiers by one-class classification methods. In: *International joint conference on neural networks*, Vancouver, Canada, pp 327–332
19. Rabaoui A, Davy M, Rossignol S, Ellouze N (2008) Using one-class SVMs and wavelets for audio surveillance. *IEEE Trans Inf Forensic Secur* 3(4):763–775
20. Yeh CY, Lee ZY, Lee SJ (2009) Boosting one-class support vector machines for multi-class classification. *Appl Artif Intell* 23(4):297–315
21. Boehm O, Hardoon DR, Manevitz LM (2011) Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms. *Int J Mach Learn Cybernet* 2:125–134
22. Tax DMJ (2001) One-class classification, PhD thesis, Delft University of Technology, Netherlands
23. Kwang-Kyu S (2007) An application of one-class support vector machines in content-based image retrieval. *Expert Syst Appl* 33(2):491–498
24. Manevitz L, Yousef M (2007) One-class document classification via neural networks. *Neurocomputing* 70:1466–1481
25. Bergani C, Oliveira LS, Koreich AL, Sabourin R (2009) Combining different biometric traits with one-class classification. *Signal Process* 89:2117–2127
26. Kwak K-C, Pedrycz W (2005) Face recognition: a study in information fusion using fuzzy integral. *Pattern Recognit Lett* 26(26):719–733
27. Chiang JH, Gaber PD (1997) Hybrid fuzzy-neural systems in handwritten word recognition. *IEEE Trans Fuzzy Syst* 5:497–510
28. Pham T, Wagner M (2000) Similarity normalization for speaker verification by fuzzy fusion. *Pattern Recognit* 33:309–315
29. Chiang JH (1999) Choquet fuzzy integral-based hierarchical networks for decision analysis. *IEEE Trans Fuzzy Syst* 7:63–71
30. Cabrera JBD, Gutiérrez C, Mehra RK (2008) Ensemble methods for anomaly detection and distributed intrusion detection in mobile ad-hoc networks. *Inf Fusion* 9:96–119
31. Juszczak P, Duin RPW (2004) Combining one-class classifiers to classify missing data. In: *5th international workshop, multiple classifier systems*. Cagliari, Italy, pp 92–101
32. Krawczyk B, Wozniak M (2014) Diversity measures for one-class classifier ensembles. *Neurocomputing* 126:36–44
33. Cyganek B, Krawczyk B (2015) Data classification with ensembles of one-class support vector machines and sparse nonnegative matrix factorization. In: *7th Asian conference on intelligent information and database systems*, Bali, Indonesia, pp 526–535
34. Kuncheva L (2004) *Combining pattern classifiers: methods and algorithms*. Wiley-Interscience Publication, Hoboken
35. Abbas N, Chibani Y, Belhadi Z, Hedir M (2013) A DSMT based combination scheme for multiclass classification. In: *16th international conference on information fusion*, Istanbul, Turkey, pp 1950–1957
36. Hadjadj B, Chibani Y, Nemmour H (2014) Fuzzy integral combination of one-class classifiers designed for multi-class classification. In: *11th international conference on image analysis and recognition*, Vilamoura, Portugal, pp 320–328
37. Zhang Y, Zhang B, Coenen F, Xiao J, Lu W (2014) One-class kernel subspace ensemble for medical image classification. *EURASIP J Adv Signal Process* 17:1–13
38. Krawczyk B, Woźniak M, Cyganek B (2014) Clustering-based ensembles for one-class classification. *Inf Sci* 264:182–195
39. Cho S-B, Kim JH (1995) Combining multiple neural networks by fuzzy integrals for robust classification. *IEEE Trans Syst Man Cybern* 25(2):380–384

40. Cho S-B (1995) Fuzzy, aggregation of modular neural networks with ordered weighted averaging operators. *Int J Approx Reason* 13(4):359–375
41. <http://www.ifnenit.com/download.htm>
42. Rath TM, Manmatha R (2003) Features for word spotting in historical manuscripts. In: 7th international conference on document analysis and recognition, vol 1, Edinburgh, Scotland, August 3–6 2003, pp 218–222
43. Bluche T, Ney H, Kermorvant C (2013) Feature extraction with convolutional neural networks for handwritten word recognition. In: 12th international conference on document analysis and recognition (ICDAR), pp 285–289
44. Poznanski A, Wolf L (2016) Cnn-n-gram for handwriting word recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2305–2314
45. Candès E, Demanet L, Donoho DL, Ying L (2006) Fast discrete curvelet transforms. *Multiscale Model Simul* 5:861–899
46. Starck J, Candes EJ, Donoho DL (2002) The curvelet transform for image denoising. *IEEE Trans Image Process* 11(6):670–684
47. Mandal T, Wu QMJ (2008) Face recognition using curvelet based PCA. In: International conference on pattern recognition, Florida, USA, pp 1–4
48. Saha A, Wu QJ (2010) Facial expression recognition using curvelet based local binary patterns. In: IEEE international conference on acoustics speech and signal processing, Texas, USA, pp 2470–2473
49. Majumdar A (2009) Image compression by sparse PCA coding in curvelet domain. *SIVIP* 3(1):27–34
50. Arivazhagan S, Ganesan L, Kumar TGS (2006) Texture classification using curvelet statistical and co-occurrence features. In: International conference on pattern recognition, Hong Kong, pp 938–941
51. Sumana IJ, Islam MM, Zhang D, Lu G (2008) Content based image retrieval using curvelet transform. In: The 10th workshop on multimedia signal processing, Australia, pp 11–16
52. Majumdar A (2006) Bangla basic character recognition using digital curvelet transform. *J Pattern Recognit Res* 2(1):17–26
53. Kazemi FM, Izadian J, Moravejjan R, Kazemi EM (2008) Numeral recognition using curvelet transform. In: IEEE/ACS international conference on computer systems and applications, Doha, Qatar, pp 606–612
54. Shirdhonkar MS, Kokare M (2011) Off-line handwritten signature retrieval using Curvelet transforms. *Int J Comput Eng* 3(4):1658–1665
55. Guerbai Y, Chibani Y, Hadjadji B (2015) The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. *Pattern Recognit* 48(1):103–113
56. Favata J, Srikantan G (1996) A multiple feature/resolution approach to handprinted digit and character recognition. *Int J Imaging Syst Technol* 7(4):304–311
57. Duin RPW (2002) The combining classifier: to train or not to train? In: 16th international conference on pattern recognition, Canada, pp 765–770
58. Dietterich TG (1998) Approximate statistical tests for combining supervised classification learning algorithms. *Neural Comput* 10(7):1895–1923
59. Nemmour H, Chibani Y (2011) Handwritten Arabic word recognition based on ridgelet transform and support vector machines. In: International conference on high performance computing and simulation, HPCS, pp 357–361
60. Khalifa M, BingRu Y (2011) A novel word based Arabic handwritten recognition system using SVM classifier. In: Shen G, Huang X (eds) ECWAC 2011, part I. CCIS, vol 143. Springer, Heidelberg, pp 163–171
61. Alalshekmubarak A, Hussain A, Wan QF (2012) Off-line handwritten Arabic word recognition using SVMs with normalized poly kernel. *Neural Inf Process Lect Notes Comput Sci* 7664:85–91