# Multi-Attribute Decision Fusion for Pattern Classification

Gabriel Awogbami
Department of Electrical and Computer Engineering
North Carolina A&T State University
Greensboro, NC, USA
giawogba@aggies.ncat.edu

Norbert Agana
Department of Electrical and Computer Engineering
North Carolina A&T State University
Greensboro, NC, USA
naagana@aggies.ncat.edu

Abdollah Homaifar
Department of Electrical and Computer Engineering
North Carolina A&T State University
Greensboro, NC, USA
homaifar@ncat.edu

***Abstract*— Classification is an important problem that is found in several domains. It has a wide range of application in medical diagnosis, decision making, and target classification. Several methods have been used for classification with varying degrees of success. Although multi-criteria decision analysis (MCDA) has been applied to classification problem, its application to pattern classification is relatively new. In this work, we propose a multi attribute decision fusion based on the classical multi-criteria decision algorithm (MADF-MCDA) for pattern classification problems. Multi-criteria decision making is a technique for selecting an alternative from a group of known alternatives based on certain criteria. The MCDA concept has been applied to many problems in sciences, business and engineering disciplines. In the proposed method, every attribute is considered as a criterion upon which a final decision is taken. For every attribute, an attribute model is built using a set of Gaussian membership functions for different classes. We assign a weight to each attribute model. This is because each attribute used in the modelling of an attribute model may have a different contribution to the final decision. It is expected that a model used for classification should have high distance among classes. The assigned weights are determined by the areas of intersection among the classes. This study also compares the performance of the proposed method with standard classification techniques such as K-nearest neighbor (K-NN), and support vector machine.**

***Keywords—Multi attribute decision fusion, multicriteria decision analysis, pattern classification***

## I. INTRODUCTION

Decision fusion is a powerful technique used for integrating multiple decisions from different sources to produce a better decision as compared to individual decision. According to [1], data fusion is the synergistic combination of information made available by knowledge sources such as sensors. This is done in order to gain better understanding about a process. The problem of classification is one of fundamental issues in data mining and knowledge discovery [2]. Classification is the process of assigning a new instance into a known set of classes based on training dataset containing instances whose classes are known. The task of classification is a decision process. Classification has offered many practical applications in pattern recognition [3], [4], decision making [5], medical diagnosis [6], and fault diagnosis [7], [8], etc. Several techniques have been proposed in literature for classification tasks, including, support vector machine, Bayesian network, decision tree, naive Bayes, and K nearest neighbors [9]. Recently, researchers have explored the concept of multicriteria decision making analysis (MCDA) and its applications in solving classification tasks. In [10], a multicriteria method was proposed for classifying 60,000 world tourist locations into four categories. Similarly, in [11], a multicriteria decision classifier was applied to multi-class classification with relatively good accuracy. Though MCDA has been applied to classification problem, its application in pattern classification is relatively new. In majority of multicriteria methods for classification, the importance of each criterion is based on subjective judgement of the expert in the domain. Expert knowledge is sometimes limited. Therefore, it becomes expedient to seek a method such that the evaluation of each criterion is more objective. In this work, we propose a multi attribute decision fusion based on the classical multicriteria decision algorithm (MADF-MCDA) for pattern classification. In this approach, the evaluation of relative importance of each attribute criterion is completely data driven. Hence, MADF-MCDA is an objective technique. The rest of this paper is organized as follows: Section II gives a brief description of MCDA problem. In Section III, we present the proposed method. Experimental results and discussion are the focus of Section IV. Finally, in Section V, conclusion and future work are presented.

## II. CLASSICAL MCDA PROBLEM

Multi-criteria decision analysis fundamentally consists of four steps [11].

1) Define alternatives and criteria
2) Determine the weight of each criterion
3) Evaluate each alternative with respect to each criterion
4) Aggregate the score of each alternative from all criteria
5) Choose the best alternative based on certain decision rule.

### *MCDA Problem Formulation*

Consider an MCDA problem with a known set of alternatives $C = [C_1, C_2, \dots, C_P]$ and a given set of criteria $A = [A_1, A_2, \dots, A_q]$. Where $p$ and $q > 2$. Each alternative $C_i$ is a possible decision to make and every criterion is characterized by a weighting factor $w_j \in [0, 1]$, $j = 1,2,\dots, q$. The weighting factor are normalized such that $\sum_j w_j = 1$. The normalized weighting factor is defined as $w = [w_1, w_2, \dots, w_q]^T$. The membership scores for each of the alternatives $C_i$ corresponding to each criterion $A_j$ is denoted by $C_{ij}$ . Therefore, the $p \times q$ score matrix, S can be defined as [12]

$$S = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1q} \\ C_{21} & C_{22} & \dots & C_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ C_{p1} & C_{p2} & \dots & C_{pq} \end{bmatrix} \quad (1)$$

## III. PROPOSED METHOD

The proposed method as shown in Figures 1 and 2 essentially consists two phases : (1) the training phase, and (2) the testing phase. The training phase involves attribute modelling and determination of attribute weight. In the testing phase, every attribute of a query instance is mapped to its corresponding attribute model, to obtain scores for different classes which represents its degree of belonging to various classes. Then, a weighted combination of all the scores from all attributes is computed for each class. The class with highest value is considered as the class of the query instance..

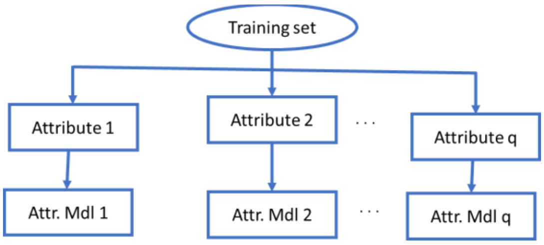


Fig. 1. Training Phase

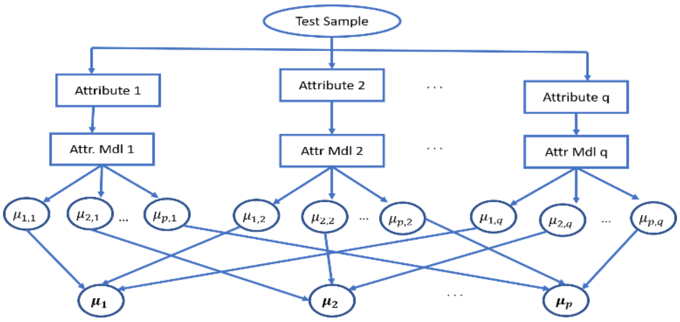


Fig. 2. Testing Phase

### A. Attribute Model

With the assumption that every attribute is independent of each other, we model each attribute using a set of Gaussian membership functions. The choice of Gaussian membership function was attributed to the fact that most physical phenomena obey Gaussian distribution [13]. A Gaussian membership function is fully characterized by its mean and standard deviation. Each class $C_i$ $[i = 1, 2, \dots, p]$ and attribute $A_j$ $[j = 1, 2, \dots, q]$ . We compute the mean $\bar{x}_{ij}$ and standard deviation $\sigma_{ij}$ as

$$\mu_{ij}(x) = \exp\left(-\frac{(x - \bar{x}_{ij})^2}{2\sigma_{ij}^2}\right) \quad (2)$$

where

$$\bar{x}_{ij} = \frac{1}{T}\sum_{k=1}^{T} x_{ijk}$$

and

$$\sigma_{ij} = \sqrt{\frac{1}{T-1}\sum_{k=1}^{T}(x_{ijk} - x_{ij})^2}$$

$x_{ijk}$ is the $k^{th}$ sample value given attribute $j$ and class $i$. And $\mu_{ij}(x)$ is the associated Gaussian membership function. $T$ is the sample size for each class.

### B. Determination of Weight of Attribute/ Criterion

In classification problems, it stands to reason that every attribute may not have equal contribution to the classification process. To this end, we need to devise a procedure to evaluate relative weight of each of the attribute models. The discriminative power of the attribute model can be used to represent its weight. This implies that models with disjointed membership functions are assigned higher weight than the one with overlapped regions. The weight can be defined as the ratio of non-overlapping area to the total area. If $\alpha_j$ is the weight of attribute j, then the weight can be expressed as:

$$\alpha_j = \frac{Non - overlapping\ Area}{Total\ Area} \quad (3)$$

or

$$\alpha_j = \frac{Total\ area - Overlapped\ Area}{Total\ Area} \quad (4)$$

$$\alpha_j = 1 - \frac{Overlapped\ Area}{Total\ Area} \quad (5)$$

Let us define

$$e_j = \frac{Overlapped\ Area}{Total\ Area}$$

Therefore, (5) becomes

$$\alpha_j = 1 - e_j \quad (6)$$

Consider a 3-class problem (Class A, B, C), we can illustrate the overlapped area by using the Venn diagram for the classical set theory as shown in Figure 3. We denote the overlapped region as OVR

$$OVR = d + e + f + g \quad (7)$$

Where

$$g = A \cap B \cap C$$
$$d = (A \cap B)only = (A \cap B) - g$$
$$e = (A \cap C)only = (A \cap C) - g$$
$$f = (B \cap C)only = (B \cap C) - g$$

Thus, OVR becomes

$$OVR = (A \cap B) + (A \cap C) + (B \cap C) - 2g \quad (8)$$

Suppose for a given attribute, the membership functions for Classes A, B, and C are $\mu_a(x), \mu_b(x)$, and $\mu_c(x)$ respectively. And their corresponding means and standard deviations are $(x_a, \sigma_a)$, $(x_b, \sigma_b)$, and $(x_c, \sigma_c)$. Then the area of overlapped region can be defined as

$$OVR = \int_u^v \min[\mu_a(x), \mu_b(x)]dx + \int_u^v \min[\mu_a(x), \mu_c(x)]dx + \int_u^v \min[\mu_b(x), \mu_c(x)]dx - 2\int_u^v \min[\mu_a(x), \mu_b(x), \mu_c(x)]dx \quad (9)$$

Where

$$u = \min[(x_a - 3\sigma_a), (x_b - 3\sigma_b), (x_c - 3\sigma_c)]$$

and

$$u = \max[(x_a + 3\sigma_a), (x_b + 3\sigma_b), (x_c + 3\sigma_c)]$$

And the total Area can be defined as

$$Total\ Area = \int_u^v max[\mu_a(x), \mu_b(x), \mu_c(x)]dx \quad (10)$$

We assume that the limit of integration is between $-3\sigma$ from the mean of the leftmost Gaussian function and $+3\sigma$ to the mean of the rightmost membership function. Thus, $e_j$ becomes

$$e_j = \frac{OVR}{\int_u^v \max[\mu_a(x), \mu_b(x), \mu_c(x)]dx} \quad (11)$$

Invariably, $\alpha_j$ turns out to be

$$\alpha_j = 1 - \frac{OVR}{\int_u^v max[\mu_a(x), \mu_b(x), \mu_c(x)]dx} \quad (12)$$

However, for a two-class problem (Class A and B), $e_j$ collapses to Jaccard similarity index [14]. Jaccard similarity $S_J(A, B)$ can be defined as

$$S_J(A, B) = \frac{\int_u^v min[\mu_a(x), \mu_b(x)]dx}{\int_u^v \max[\mu_a(x), \mu_b(x)]dx} \quad (13)$$

The normalized weight $w_j$ becomes

$$w_j = \frac{\alpha_j}{\sum_l \alpha_l}, (l = 1,2, \ldots, q) \quad (14)$$

The area under a function was implemented in MATLAB using numerical integration toolbox *cumtrapz.*

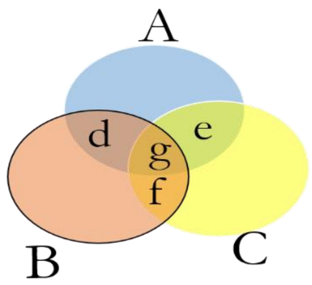


Fig. 3. Venn Diagram.

### C. Similarity between test Sample and Attribute Model

We obtain similarity between the test sample and different attribute models by mapping its attribute values to its corresponding attribute models. Each attribute model can be viewed as a classifier. In [15], it was reported that aggregation of output of classifiers depends on the information obtained from individual classifiers. There are three types of output levels: (1) the abstract level, (2) the rank level, and (3) the measurement level. In this problem, the output of each classifier belongs to measurement levels. This means that each attribute model returns a p-dimensional vector, $[\mu_{1j}, \mu_{2j}, \ldots, \mu_{pj}]^T$. Where $\mu_{1j} \in [0,1]$ is the support that a sample $x = [x_1, x_2, \ldots, x_q]^T$ is from class $i$ given attribute $j$.

### D. Consttruction of the Decision Matrix

We present the outputs of the all the q attribute models for a given query instance in form of a decision matrix D as shown in (15)

$$D = \begin{bmatrix} \mu_{11}(x_1) & \cdots & \mu_{1j}(x_j) & \cdots & \mu_{1q}(x_q) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{i1}(x_1) & \cdots & \mu_{ij}(x_j) & \cdots & \mu_{iq}(x_q) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{p1}(x_1) & \cdots & \mu_{pj}(x_j) & ,,, & \mu_{pq}(x_q) \end{bmatrix} \quad (15)$$

### E. Aggregation Function

The aggregation function is a weighted combination of scores. This is done by multiplying matrix D with the normalized weight vector w. The weighted score received by each class becomes

$$\mu_i = \sum_{j=1}^{q} w_j \mu_{ij}(x_j) \quad (16)$$

The decision rule: Assign $x$ to class $m$ if

$$\sum_{j=1}^{q} w_j \mu_{mj}(x_j) = \max_i \sum_{j=1}^{q} w_j \mu_{ij}(x_j) \quad (17)$$

## IV. EXPERIMENTAL RESULTS AND DISCUSION

### A. Experiment

The IRIS dataset collected by the Fisher was used to test the performance of the proposed method. The IRIS data set consists of three types of flowers: Setosa (S), Versicolor (C), and

Virginica(V). Each class has 50 samples resulting in a total of 150 samples. There are four features to describe each sample: Sepal Length(SL), Sepal Width(SW), Petal Length(PL), Petal Width(PW). Each feature is a criterion used to build an attribute model. We have a total of four attribute models. The experimental procedure is described in the following steps:

(1) For each attribute, construct Gaussian membership functions for different classes. The four attribute models are shown in Fig. 4, 5, 6, and 7. After one experiment, the following parameters of the Gaussian membership functions are obtained from the training set as shown in Table I. A total of 99 samples were selected as training set while the remaining 51 samples were chosen as testing set.

(2) Determine the weight of each attribute model using (9) - (14). The results of the weight calculation for the different attribute models are shown in Table II. The normalized weight vector, $w = [0.2202, 0.1443, 0.3129,\ 0.3226]^T$.

(3) For every test sample, map its attribute values with the associated attribute models. For example, if we choose a test sample from the virginica class whose four attribute values are: SL=6.4cm, SW=3.2cm, PL=5.3cm, and PW=2.3cm.

TABLE I. SIMULATION PARAMETERS

| **Attribute** | Parameters | S | C | V |
|---|---|---|---|---|
| Sepal Length | Mean | 5.1143 | 5.9792 | 6.6560 |
| | sd | 0.3665 | 0.5022 | 0.7292 |
| Sepal Width | Mean | 3.4857 | 2.7750 | 2.9600 |
| | sd | 0.3381 | 0.3193 | 0.3786 |
| Petal Length | Mean | 1.4667 | 4.3167 | 5.600 |
| | sd | 0.1683 | 0.4743 | 0.6434 |
| Petal Width | Mean | 0.2619 | 1.3625 | 1.9880 |
| | sd | 0.1024 | 0.2261 | 0.2438 |

(4) Map every attribute value of the query sample to its associated attribute model to obtain its degrees of belonging to different classes. This is done by finding its similarity with the corresponding attribute model. The degrees of belonging are the scores received by different classes given the attribute model. For notational simplicity, we will use the following subscripts to represent the different classes: Setosa='1', Versicolor='2', and Virginica='3'. As shown in Figure 8, the corresponding membership values for the attribute SL=6.4cm for all the classes are, $\mu_1(6.4) = 0.0021, \mu_2(6.4) = 0.7033, and\ \mu_3(6.4) = 0.9402$. This is repeated for the remaining three attributes of SW, PL, and PW respectively.

(5) Construct a decision matrix $D$ such that every column represents the scores received by different classes by a given attribute model and every row represents different scores received by a given class from different attribute models. The results for the selected instance are presented in (18). We rounded up membership value to 4 decimal places.

$$D = \begin{bmatrix} 0.0021 & 0.6998 & 0.0000 & 0.0000 \\ 0.7033 & 0.4124 & 0.1166 & 0.0020 \\ 0.9402 & 0.8180 & 0.8491 & 0.4409 \end{bmatrix} \quad (18)$$

TABLE II. ATTRIBUTE WEIGHTS

| Attribute | Total Area | Overlapped Area | Weight | Normalized Weight |
|---|---|---|---|---|
| SL | 2.7572 | 1.0288 | 0.6269 | 0.2202 |
| SW | 1.4890 | 0.8774 | 0.4107 | 0.1443 |
| PL | 2.9058 | 0.3175 | 0.8907 | 0.3129 |
| PW | 1.3250 | 0.1082 | 0.9183 | 0.3226 |

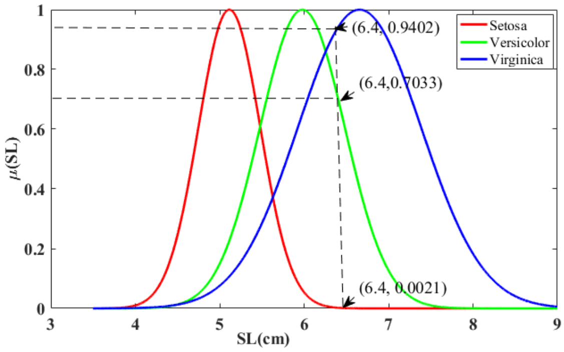


Fig. 4. Sepal Length Attribute Model

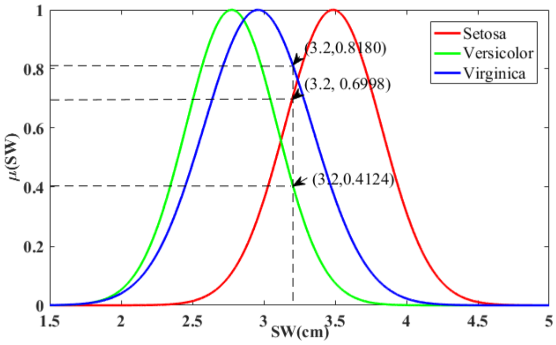


Fig. 5. Sepal Width Attribute Model.

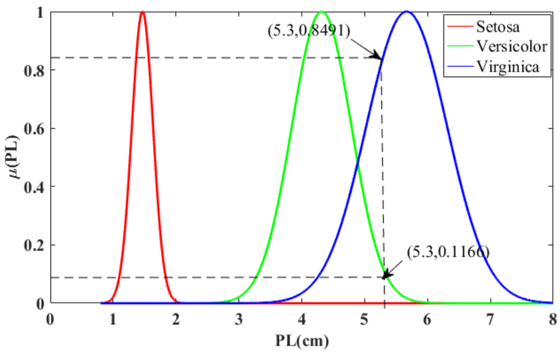


Fig. 6. Petal Length Attribute Model.

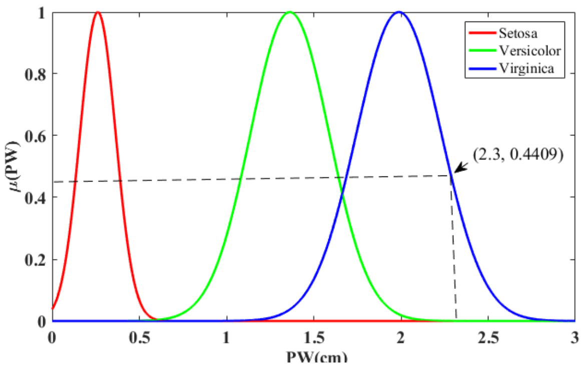


Fig. 7. Petal Width Attribute Model.

(6) Compute the weighted combination of scores received by each class. This is accomplished by multiplying the decision matrix with the weight vector. The following results were obtained $\mu_1(x) = 0.1014, \mu_2(x) = 0.2515, \mu_3(x) = 0.7330$. Since class 3 has the maximum value, we assign the test instance to Class 3 which is Virginica. The proposed method (MADF-MCDA) was compared with 3-nearest neighbors (3-NN), 5-nearest neighbors (5-NN), and support vector machines (SVM). The experiment was repeated with 200 random splits for training and test sets, the average accuracy of each of the methods is shown in Table III and Fig. 8.

Table III. Results

| METHOD | 3-NN | 5-NN | SVM | MADF-MCDA |
|---|---|---|---|---|
| ACCURACY | 0.948 | 0.952 | 0.961 | 0.935 |

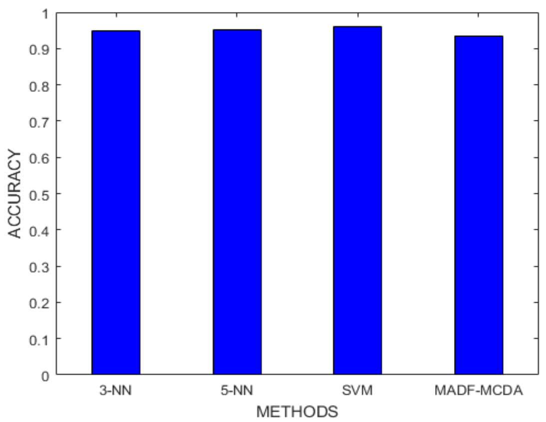


Fig. 8. Accuracy of Different Methods.

### *B. Discussion*

As shown in Table III and Figure 8, it can be observed that, though the performance of the proposed method is not the best of the four methods, its performance is only comparable. The whole idea behind this work is to extend the concept of the classical MCDA to pattern classification. By analogy, every attribute is seen as a criterion and every class is seen as an alternative.

## V. CONCLUSION

We have successfully extended the concept of the classical multi criteria decision analysis (MCDA) to multi attribute decision fusion (MADF) for pattern classification. The proposed method (MADF-MCDA) is an objective technique in the sense that it is driven by data. A spin-off of this method is the evaluation of the relative contribution of each attribute to the multi-class classification. This is especially useful in applications where feature selection is required. The proposed method was tested on the famous IRIS dataset. The results obtained are comparable with that of some standard classification techniques such as the K-nearest neighbors (k-NN) and support vector machines(SVM). As part of future work, we will explore other MCDA algorithms such as Analytical Hierarchy Process (AHP), Techniques for Order of Preference by Similarity to Ideal Solutions (TOPSIS) to improve the accuracy of the proposed method. We hope to also generalize the method of finding the overlapped area to consider cases where there are more than 3 classes.

## ACKNOWLEDGMENT

This work is supported by Air Force Research Laboratory and OSD under agreement number FA8750-15-2-0116. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory and OSD or the U.S. Government.

## REFERENCES

[1] M. A. Abidi and R. C. Gonzalez, Data fusion in robotics and machine intelligence. Academic Press Professional, Inc., 1992.J. ford: Clarendon, 1892, pp.68–73.

[2] F. Azuaje, “Review of” data mining: Practical machine learning tools and techniques” by witten and frank,”

[3] N. M. Nasrabadi, “Pattern recognition and machine learning,Journal of electronic imaging, vol. 16, no. 4, p. 049901, 2007

[4] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification. John Wiley & Sons, 2012.

[5] S.-J. Chen and C.-L. Hwang, “Fuzzy multiple attribute decision making methods,” in Fuzzy multiple attribute decision making. Springer, 1992, pp. 289–486.I.

[6] Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine learning, vol. 46, no. 1, pp. 389–422, 2002.

[7] S. Dash, R. Rengaswamy, and V. Venkatasubramanian, “Fuzzy-logic based trend classification for fault diagnosis of chemical processes,” Computers & Chemical Engineering, vol. 27, no. 3, pp. 347–362, 2003.

[8] R. Isermann, Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer Science & Business Media, 2006.

[9] X. Xu, J. Zheng, J.-b. Yang, D.-l. Xu, and Y.-w. Chen, “Data classification using evidence reasoning rule,” Knowledge-Based Systems, vol. 116, pp. 144–151, 2017.

[10] D. Mailly, I. Abi-Zeid, and S. Pepin, “A multi-criteria classification approach for identifying favourable climates for tourism,” Journal of Multi-Criteria Decision Analysis, vol. 21, no. 1-2, pp. 65–75, 2014.

[11] C.-T. Chen, P.-F. Pai, and W.-Z. Hung, “Applying multicriteria decision classifier in multi-class classification,” in Fuzzy Systems and Knowledge Discovery (FSKD), 2011,Eighth International Conference on, vol. 1. IEEE, 2011, pp. 607–610.

[12] J. Dezert, D. Han, and H. Yin, “A new belief function based approach for multi-criteria decision-making support,” in Information Fusion (FUSION), 2016 19th International Conference on. IEEE, 2016, pp. 782–789.

[13] Z. Zhang, Z. Hao, S. Zeadally, J. Zhang, B. Han, and H.- C. Chao, “Multiple attributes decision fusion for wireless sensor networks based on intuitionistic fuzzy set,” IEEE Access, vol. 5, pp. 12 798–12 809, 2017.

[14] P. Jaccard, “ ´ Etude comparative de la distribution florale dans une portion des alpes et des jura,” Bull Soc Vaudoise Sci Nat, vol. 37, pp. 547–579, 1901.

[15] L. Xu, A. Krzyzak, and C. Y. Suen, “Methods of combining multiple classifiers and their applications to handwriting recognition,” IEEE transactions on systems,man, and cybernetics, vol. 22, no. 3, pp. 418–435, 1992.