Weighted Fuzzy Dempster-Shafer Framework for Multi-Modal Information Integration

Yu-Ting Liu, Yang-Yin Lin, *Member, IEEE*, Shang-Lin Wu, Amar Marathe, Nikhil R. Pal, *Fellow, IEEE*, Chun-Hsiang Chuang, Jung-Tai King and Chin-Teng Lin*, *Fellow, IEEE*

***Abstract*— This study proposes an architecture based on a weighted fuzzy Dempster-Shafer framework (WFDSF), which can adjust weights associated with inconsistent evidences obtained from different classification approaches, to realize a fusion system for integrating multi-modal information. The Dempster–Shafer theory (D–S theory) of evidence enables us to integrate heterogeneous information coming from multiple sources to reap collaborative inferences for a given problem. To conquer various uncertainties associated with the collected information, it assigns beliefs and plausibilities to possible hypotheses of each decision-maker and uses a combination rule to fuse multi-modal information. For information fusion, an important step in D–S aggregation is to find an appropriate basic probability assignment (BPA) scheme for allocating support to each possible hypothesis/class, which still remains an arduous unsolved problem. Here we propose a mathematical structure to aggregate weighted evidences extracted from two different types of approaches: fuzzy Naïve Bayes and nearest mean classification rule.**

Yu-Ting Liu and Shang-Lin Wu are with the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: ytliu.ece00g@g2.nctu.edu.tw; slwu19870511@gmail.com).

Yang-Yin Lin is with the Brain Research Center, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: oliver.yylin@gmail.com).

Amar Marathe is with the Army Research Laboratory-HRED, Aberdeen Proving Ground, MD, USA (e-mail: amar.marathe@case.edu).

Nikhil R. Pal is with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700108, India (e-mail: nikhil@isical.ac.in).

Chun-Hsiang Chuang and Jung-Tai King are with the Brain Research Center, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: chchuang@ieee.org; jtchin2@gmail.com).

*Chin-Teng Lin is with the Brain Research Center and the Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: ctlin@mail.nctu.edu.tw).

Further, an intuitionistic belief assignment is employed to deal with uncertainties between hypotheses/classes. Finally, twelve benchmark problems from UCI machine learning repository for classification are employed to validate the proposed WFDSF based scheme. In addition, an application of WFDSF to practical brain-computer interfaces (BCIs) problem involving multi-modal data fusion is demonstrated in this study. Experimental results show that the WFDSF is superior to several existing methods.

***Index Terms*— Weighted Fuzzy Dempster-Shafer Framework (WFDSF), Data Fusion, Evidence Theory, Basic Probability Assignment (BPA), Fuzzy Evidential Reasoning**

I. Introduction

Recently in systematic design, a considerable amount of multi-modal information (e.g., sound, images, videos) is often simultaneously recorded and used. The variety of information helps systems to boost performance in several practical applications, such as robotic control [1], pattern recognition [2]–[4], and medical imaging [5], [6]. In each of these applications, distinct information sources provide different useful information and assist the system to make a unified decision/action. For any real life system different sources of information may not necessarily provide evidence in favor of only one particular decision. This may happen because of different types of uncertainty. For example, in a BCI application, some EEG channels may suggest a drowsy state of the driver, while some other channels, due to noise/uncertainty, may suggest an alert state. In other words, in this case, the two sources apparently provide inconsistent evidences for the same task. Developing an effective approach for integrating the multi-modal information turns out to be an important issue in designing decision making systems.

Data fusion is a technique that exploits heterogeneous data from multiple sources to reap collaborative inferences, i.e., inferences achieved via collaboration of multiple sources of information or collaboration of multiple decision makers such as multiple classifiers. Usually inferences obtained via collaboration are better than that by individual decision maker/source. The aim of data fusion is to ameliorate the quality of final decision by aggregating pieces of evidences from multiple information sources while decreasing the uncertainty in decision. More importantly, such a technique should efficiently exploit redundancy and complementariness between sources taking a global view to facilitate achieving optimal system performance.

Khaleghi *et al.*[7] made a generic and comprehensive review on the state of the art techniques of data fusion and explored concepts, benefits and challenges of each existing methodology. One of the distinguished approaches to model uncertainty /evidence is the Dempster–Shafer theory (D-S theory) [8], [9]. This is an effective method for dealing with uncertainty and integration of information from multiple sources. The sources could be different experts' opinions, multi-modal source of data, or even individual features characterizing an object. The D-S theory is a generalized variant of Bayesian probability theory, which introduces the notion of assigning beliefs and plausibilities to possible hypotheses of each decision-maker and provides a combination rule to fuse multi-modal information. Unlike the traditional mechanism in Bayesian theory, the D–S theory allows each source to incorporate information in different levels of detail. This property brings a significant benefit that it assigns a possibility mass to sets or intervals; hence, the fusion system can efficiently deal with both probabilistic (or objective) uncertainty and epistemic (or subjective) uncertainty. The D-S theory provides a mechanism for representing and processing uncertain, imprecise and incomplete information from different sources. Finally, the optimal decision can be made by means of Dempster's rule [8] of combination to integrate

information/evidences from different sources. In this context, it is worth mentioning the concept of ignorance function that is defined to measure the degree of uncertainty of an expert when assigning numerical values as membership degrees of an object to a given class [10].

Although D-S theory has been successfully applied in many practical fusion systems [11]–[18], an effective approach to assign evidences for each information source is still a crucial issue. As already mentioned, an important step in applying the D–S theory for data fusion is to find an appropriate basic probability assignment (BPA) for allocating evidences to each possible state/hypothesis/class. The BPA should be robust in the sense that a small perturbation in the data and/or parameters defining the BPA should not alter the assigned mass significantly. Intuitively for each of the possible states, the BPA represents a kind of "degree" of belief which supports the claim that the actual system belongs to that state.

BPA functions are usually designed heuristically according to the characteristic of data collected from multiple sources. Depending on the problem and available information, however, more objective and statically sound methods have also been developed for construction of belief, for example, the predictive belief function of Denœux [19]. Different authors have addressed this issue using different approaches in past studies. Two major approaches to the BPA determination are: 1) generative approaches [3], [20]–[22] based on density estimation, in which posterior probability estimates are computed from conditional densities and prior probabilities from distributions of different categories using Bayes' theorem [23], and 2) discriminative approaches [5], [24]–[26] based on distance function, in which directly posterior class probabilities are estimated.

For generative approaches, a mass/BPA is usually referred to as a probability, which is often associated with a probability distribution/fuzzy membership /possibility function. In [21], Zhu *et*

al. developed a method of automatically determining the mass function in D-S structure for image segmentation, in which basic probabilities assigned to a pixel are derived from membership degrees of gray levels. To effectively resolve the image segmentation problem, Boudraa *et al*. [27] estimated BPAs using fuzzy membership degrees derived from gray-level image histograms. Additionally, an intuitionistic view of the Dempster–Shafer model by utilizing the concept of intuitionistic fuzzy sets is described by Yager [20]. In fuzzy sets theory, imprecise membership can be represented in two different ways, using interval-valued fuzzy sets and intuitionistic fuzzy sets. Unlike ordinary fuzzy sets, in an intuitionistic fuzzy set, uncertainty is represented by a membership and a non-membership, where the sum of membership and non-membership is less than or equal to one. Like an interval valued fuzzy set, for the D-S framework, the imprecise probability associated with a set can be represented by an interval defined by the belief and plausibility measures. Inspired by the analogy between imprecise interval-valued membership of a fuzzy sets and the imprecise probability of a belief structure, Yager obtained an intuitionistic representation of the imprecise probability associated with a set. Xu *et al*. [3] adopted a normal distribution model for each attribute of data to construct a nested BPA structure, which can avoid high conflict between evidences. In [22], Masson and Denoeux incorporated D-S theory with the concept of the credal partition to address the problem of computing a credal partition from object data.

For discriminative approaches, BPAs are built based on the notion of similarity of a pattern to be classified with training patterns, which is often referred to as distance, specificity, and consistency. Denoeux [26] proposed a novel classification approach by combining k-nearest neighbors (k-NN) rule and D–S theory of evidence to establish the k-NN DST rule. Further, Pal and Ghosh [25] combined the underlying philosophy of rank nearest neighbor (RNN) rule with D-S theory for

determining the mass function of BPA. In addition, D-S theory based neural networks are presented in [5], [24], and the determination of BPA for each input pattern is implemented by using relationship with reference patterns as the evidence to each possible hypothesis.

These approaches have their own pros and cons; therefore, several authors attempted to combine both types to inherit benefits from each of them. Smets and Kennes [28] established the transferable belief model (TBM) to quantify uncertainty using belief function which is not related to any underlying probability model. Using this framework, in [29] Denœux and Smets proposed two approaches to pattern classification. One is TBM model-based classifier that depends on the general Bayesian theorem, and the other one is TBM case-based classifier that utilizes the concept of similarity of a test pattern with the training patterns. A novel approach to determine BPA is proposed in [30], which employs core samples extracted from the training dataset to calculate a relevance ratio based on incoming datum and the core samples. Then BPAs are assigned based on the relevance between the test data and the selected core samples.

In this study, we present a novel architecture based on a weighted fuzzy Dempster-Shafer framework (WFDSF) for the classification task, which takes into consideration both of the aforementioned approaches. The proposed method utilizes a mathematical structure to aggregate the weighted evidences extracted from both the generative-type approach and the discriminative-type approaches. The belief function used in the WFDSF is established based on simple models, which can universally aggregate multi-modal evidences from various sources for the recognition task. The tunable weighted mechanism in this structure enables us to use a training procedure to find the optimal parameters that should be used for different types of approaches, and it assists the D-S framework to adaptively deal with different problems. In the proposed method, the basic probability of an incoming datum to each class is assigned by calculating the fuzzy

membership degree to each cluster and the distance between cluster centroids representing each category. Further, an intuitionistic belief assignment is employed to deal with uncertainties between hypotheses/classes. In order to reap a comprehensive decision from the datum for the recognition task, the evidences for an incoming pattern are generated from each attribute/feature and then combined using Dempster's rule of combination [8] to get the final evidence. We evaluated the proposed scheme extensively using several real-world datasets obtained from the UCI machine learning repository [31], and the experimental results show that the performance of our system is superior to that of its rivals in this study.

The novel aspect of this work is that the fuzzy Naïve Bayes method [32], [33] and the nearest mean classification rule [25], [26], [34] are employed simultaneously for calculating the basic probability of each hypothesis as an example of integrating both of the previously mentioned types of approaches. The proposed model utilizes concepts from previous approaches [3], [5], [20], [21], [24]–[26], [35], and the basic probabilities are assigned by a "compound-type" (generative and discriminative) approach. In contrast to typical approaches in which basic possibilities are determined inflexibly, the tunable weights in the proposed model for BPA determination provides a compatible framework to aggregate evidences on either of the two types or a compound type. These weights can efficiently introduce robust and sufficient information for modeling the evidences. Additionally, our proposed method exploits an intuitionistic belief assignment to represent the level of uncertainty between different classes, and consequently the uncertainties that adversely affect the system performance can be well-controlled. Finally, the proposed architecture can efficiently process multi-modal information from different sources to achieve a compatible fuzzy DST-based fusion architecture.

The rest of this paper is organized as follows. Section II briefly introduces the Dempster–Shafer

theory of evidence as well as its related concepts. Section III introduces how the proposed weighted fuzzy D-S structure determines BPAs. Section IV presents the experimental results on a recognition task that are obtained from the proposed system. Section V demonstrates how the proposed system can effectively integrate multi-modal physiological signals for a practical brain-computer interface (BCI) experiment. Finally, Section VI offers concluding remarks.

II. Background on Dempster–Shafer Theory of Evidence

The Dempster–Shafer theory of evidence, initiated by Dempster [9] and then mathematically formalized by Shafer [8], is a general framework for modeling and reasoning with epistemic uncertainty, which allows us to combine evidences from multiple sources to arrive at combined degree of belief on different hypotheses. Compared with the Bayesian model, D-S theory is a comprehensive approach to deal with uncertainty and imprecision with a theoretically attractive evidential reasoning framework. The D-S theory consists of three main notions: 1) assigning appropriate beliefs and plausibilities to possible hypotheses, 2) employing the D-S rule of combination for fusing independent items of evidence, and 3) making the final decision on the choice of optimal hypothesis in a flexible and rational manner.

A. Frame of Discernment

The Dempster–Shafer theory of evidence is a mathematical framework for reasoning under uncertainty based on modeling of evidence [8]. A set of hypotheses θ is defined on a finite domain Θ, which is named the frame of discernment in D-S framework.

$$\Theta = \{\theta_1, \theta_2, \ldots \theta_N\} \tag{1}$$

The set Θ is constituted of N exhaustive and exclusive hypotheses. Information sources can distribute mass values on the power set of the frame of discernment, denoted by 2^{Θ}.

$$2^{\Theta} = \{\varnothing, \{\theta_1\}, \{\theta_2\}, \ldots \{\theta_N\}, \{\theta_1, \theta_2\} \ldots \{\theta_1, \theta_2, \ldots \theta_N\}\} \quad (2)$$

where ∅ denotes the empty set. The N subsets, each containing only one element, are called singletons.

B. Mass Functions, Focal Elements and Basic Probability Assignment

When the frame of discernment is determined, uncertain evidences for each possible proposition in 2^{Θ} are represented by a (normalized) mass function m to the interval [0,1]. Function m has two properties as follows:

$$m(\varnothing) = 0 \quad (3)$$

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

The mass function is also called the basic probability assignment (BPA) function. For a particular element A in 2^{Θ}, $m(A)$ characterizes the degree of relevant evidence that supports the proposition that an unknown datum belongs to the set A alone but not to any subset of A in particular. Any subset A in 2^{Θ} with $m(A | A \in 2^{\Theta}) > 0$ is called a focal element. Focal elements are members of 2^{Θ} on which the available evidence focuses.

C. Belief and Plausibility Measure Functions

Since $m(A)$ measures the belief that one commits to the set A exactly and not to any proper subset of A, to obtain the total belief committed to A, one must add to $m(A)$ the quantities $m(B)$ for all proper subsets B of A. Given a mass function m, a belief measure and a plausibility measure can be uniquely determined to reap a total BPA committed to a particular proposition. The belief function Bel is defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{5}$$

where all $B \subseteq A; A \in 2^{\Theta}$. $Bel(A)$ indicates the total evidence or belief that the element belongs to the set A ($x \in A$) as well as to the various subsets of A. On the other hand, the plausibility function Pl is defined as:

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{6}$$

$Pl(A)$ represents the total evidence or belief that the element belongs to set A ($x \in A$) as well as to the various special subsets of A plus the additional evidence or belief associated with sets that overlap with A. We can interpret the relationships between $Bel(A)$ and $Pl(A)$ as follows:

$$Pl(A) = 1 - Bel(\overline{A}) \tag{7}$$

$$Pl(\varnothing) = 0 \tag{8}$$

The belief function $Bel(A)$ signifies the total amount of probability distributed among the elements of A, which establishes a lower limit on the probability of A. In contrast, the plausibility function $Pl(A)$ describes the total belief degree related to A, which establishes an upper limit on the probability of A. D-S theory presents an explicit measure of belief about a particular proposition A and its complement $\bar{A}$ as a length of the interval $[Bel(A), Pl(A)]$ called belief interval. The basic probability can also be viewed as determining a set of probability distributions P over 2^{Θ} satisfying:

$$Bel(A) \leq P(A) \leq Pl(A) \tag{9}$$

This fundamental imprecision in the determination of the probabilities reflects the incompleteness of the available information. The above inequalities reduce to equalities in the case of a Bayesian belief function.

D. Dempster's Rule of Belief Combination

To deal with uncertain, imprecise and incomplete data from multiple sources, the D-S theory allows each source to contribute information in different levels of detail, which provides a mathematical architecture to aggregate relevant information from disparate sources. That is, we can reap a joint decision based on the evidence or belief determined by several experts independently.

Let m_1 and m_2 be two mass functions induced by two independent sources of evidence. Dempster's rule of combination [8], [9], which is denoted as $m = m_1 \oplus m_2$ within the framework of evidence theory, integrates the two BPAs, m_1 and m_2, to yield a combined BPA as:

$$m_1 \oplus m_2(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{10}$$

$$\kappa = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{11}$$

where all A, B and $C \in 2^{\Theta}$, $A \neq \emptyset$, $m_1 \oplus m_2(\emptyset) = 0$. κ is the degree of conflict, called the conflict coefficient, between m_1 and m_2. It is calculated by the sum of products $m_1(B) m_2(C)$ for all focal elements B in m_1 and C in m_2 where $B \cap C$ is null. The larger the value of κ is, the more conflicting are the two sources. When $\kappa = 1$, it implies that these two evidences are in complete logical contradiction. κ is also a normalization constant to let the joint BPA observe the property $\sum_{A \in 2^{\Theta}} m(A) = 1$ in D-S theory. Dempster's rule of combination is associative and commutative; hence, the total belief mass function resulting from the combination of all information sources is defined as:

$$m_{total} = m_1 \oplus m_2 \oplus \ldots m_j \tag{12}$$

where j is the total number of information sources. m_{total} aggregates information from all of the individual sources and represents aggregated mass function after the D-S fusion process.

E. Pignistic Probability Transformation

After combination, a decision can be uniquely determined by a belief measure Bel, a plausibility measure Pl, or pignistic probability transform P_{pig}. Since Bel and Pl represent the belief interval $Bel(A) \leq P(A) \leq Pl(A)$ for a particular proposition in the hypotheses, P_{pig} offers an appropriate and fair inference with a probability measure for a decision. The pignistic probability transform of the final mass values was defined in transferable belief model (TBM) proposed by Smets and Kennes [28], [36]. It is worth noting that the term "*pignistic*" originates from the word *pignus*, meaning 'bet' in Latin. The Pignistic probability employs principle of insufficient reason to transform the mass function to a probability measure for decision making. The estimate of the Pignistic probability is a crisp value in a belief interval and can be defined as:

$$P_{pig}(A) = \sum_{\substack{B \in 2^{\Theta} \\ A \cap B \neq \varnothing}} \frac{|A \cap B| \times m(B)}{|B| \times (1 - m(\varnothing))} \quad \forall A \in 2^{\Theta} \tag{13}$$

where $m(\emptyset) = 0$, the equation above can be simplified as:

$$P_{pig}(A) = \sum_{\substack{B \in 2^{\Theta} \\ A \cap B \neq \varnothing}} \frac{|A \cap B| \times m(B)}{|B|} \quad \forall A \in 2^{\Theta} \tag{14}$$

where $|B|$ is the number of singleton elements in set B. Based on (14) the pignistic probabilities for singletons can be written as:

$$P_{pig}(\{a\}) = \sum_{a \in B} \frac{m(B)}{|B|} \quad \forall a \in \Theta \tag{15}$$

III. Weighted Fuzzy Dempster-Shafer Framework

As mentioned earlier, to determine BPA functions, there are two major types of approaches: 1) generative approaches based on density estimation and 2) discriminative approaches based on distance function. Generative types of methods usually employ estimation of posteriori probabilities to represent the corresponding belief mass function for each category. On the other hand, discriminative type approaches capture relative information, e.g., distance, specificity and similarity. In this section, we discuss the fuzzy Naïve Bayes method [32] and the nearest mean classification rule [25], [26], [34] and employ them simultaneously for obtaining the basic probability assignment thereby integrating both of the aforementioned types of approaches. Furthermore, we present a weighted regulatory architecture based on a fuzzy Dempster-Shafer mechanism for pattern classification, in which compound approaches are taken into consideration. The weighted regulatory mechanism for BPA determination provides a compatible framework to gather evidences either on the individual types or on a compound type.

A. Generative-type Approach - Fuzzy Naïve Bayes

The generative type approaches are often associated with probability distribution, fuzzy membership and possibility function to deal with recognition problems. Traditional non-fuzzy approaches, e.g., Naïve Bayes [37], typically employ probabilistic methods to handle uncertainty for system identification problems. Naïve Bayes is an inductive technique that is frequently adopted for recognition tasks. The inductive learning mechanism obtains a probability distribution on each potential category from the training data and estimates a posterior probability for each incoming/unknown data. Let $X_i, i = 1 \dots p$ be p independent variables (features) and each object, X, is represented by a p-dimensional feature vector. Let Y be the class label associated with X; $\in C = \{C_1, C_2, \dots, C_N\}$. Then the posterior probability $P(Y|X_i), i = 1 \dots p, Y \in \{C_1, C_2, \dots, C_N\}$ can be

defined by the Bayesian rule as follows:

$$P(Y \mid X_i \ldots X_P) = \frac{P(X_i \ldots X_p \mid Y) \cdot P(Y)}{P(X_i \ldots X_p)} = \frac{\prod_i P(X_i \mid Y) \cdot P(Y)}{P(X_i \ldots X_p)} \quad (16)$$

In contrast to the canonical Bayesian probability theory, fuzzy set theory (FST) [38] provides a framework for handling uncertainty, imprecision, and complexity in real world applications. It provides a flexible framework of modeling uncertainty so that each unknown sample may not belong to a sharply defined class or cluster. This characteristic of comprehensiveness helps deal with real-world datasets that often possess ambiguities caused by various factors, e.g., measurement bias and noise contamination, which cannot be adequately modeled by probability theory. This property of FST suggests that the combination of DST and FST would be an effective approach to establish a comprehensive structure for processing uncertain information. Therefore, we employ the fuzzy Naïve Bayes [33] approach to determine BPAs in this study. In the Fuzzy Naïve Bayes approach, each variable takes linguistic values defined by a fuzzy partition of the domain of each variable. Consequently, the Naïve Bayes is a special case of the fuzzy Naïve Bayes that uses crisp rather than fuzzy variables. In the proposed method, the basic probabilities of an incoming datum to each class are assigned by using fuzzy membership degree to each category.

To utilize the fuzzy Naïve Bayes approach for determining BPAs, we first partition the domain of each variable into fuzzy sets (linguistic values). Consequently, the training of the fuzzy Naïve Bayes approach proceeds in two phases as follows: 1) applying exploratory data analysis to obtain an appropriate fuzzy partition of domain of each variable, and 2) estimating a conditional probability for each sample under consideration according to the fuzzy membership function. In the exploratory data analysis step, actually the fuzzy c-means clustering algorithm is used to find a set of c clusters. Then using the cluster centroids, c, linguistic values are determined. Finally, the conditional probability to class C_i decided by the fuzzy Naïve Bayes approach is assigned as the

basic probability used in the DST architecture.

$$m(C_i) = \mu_{C_i}(x) \quad . \tag{17}$$

We emphasize here that, although we are calling the assignment at (17) a BPA, this may not result in a valid BPA as the sum of assignments may not be equal to one. Later we shall use a normalization scheme to define a valid BPA.

B. Discriminative-type Approach – Nearest Mean Classification Rule

The discriminative approaches are built on the notion of similarity of a pattern to be classified with training patterns, which are often referred to as distances, specificity, and consistency. The nearest mean classification (NMC) [34] rule is a distinguished discriminative approach, which is both simple and robust. In this study, we utilize the distance of a test data point from the nearest class mean to obtain the posterior probability for defining the BPA. The objective of the NMC rule is to assign an incoming sample to a class that has the nearest centroid with respect to the sample. Let $X = \{x_1, x_2, \ldots, x_M\}$ be a collection on M training samples that have corresponding labels $\{y_1, y, \ldots, y_M\}$, where x_i is a feature vector in a p-dimensional metric space. Each x_i is assumed to possess a class label $y_i \in C$, where $C = \{C_1, C_2, \ldots, C_N\}$ is a set of N classes, indicating that x_i belongs to one of the classes in C. Further, the centroid vector of each class can be denoted by $\{v_1, v_2, \ldots, v_N\}$. The incoming sample is then classified according to the minimum Euclidean distance to the class centroid. The square Euclidean distance is computed by

$$d(x^s, v_i) = \left(x^s - v_i\right)^T \left(x^s - v_i\right) \tag{18}$$

Consequently, we assigned the predicted label $f(x)$ for an unknown sample x^s as follows:

$$f(x) = \arg\min_{i \in C} d(x^s, v_i) \tag{19}$$

Let $f(x)$ be the predicted label by the NMC. It is decided according to the minimum $d(x^s, v_i)$. Although $f(x)$ is the system output of NMC for a classification task, we shall not use it in the proposed system. Rather what we need is only the distance function $d(x^s, v_i)$. We shall use the distance of the test point from the class means to define BPAs.

Finally, in accordance with the D-S formalism, the mass function $m(x)$ can be defined using an exponential function over the Euclidean distance as.

$$m(C_i) \propto e^{-d(x^s, v_i)} \tag{20}$$

where $C = \{C_1, C_2, \ldots, C_N\}$ is the set of N hypotheses (classes) used in the DST.

C. Procedure of the WFDSF

A system flowchart of the proposed WFDSF is shown in Fig. 1. The decision is obtained through a four-step process, and the complete procedure is detailed as follows.

1) Step 1 - Attribute Division

For a given multivariate dataset, each individual attribute/feature is considered an independent source of information. An attribute can also be multivariate, if these multiple features are obtained from the same source. In order to reap a comprehensive decision from the datum for the recognition task, the WFDSF gathers BPAs from each attribute separately. After the process of attribute decomposition, the test dataset with P attributes are divided and transformed into P independent modules in this paper.

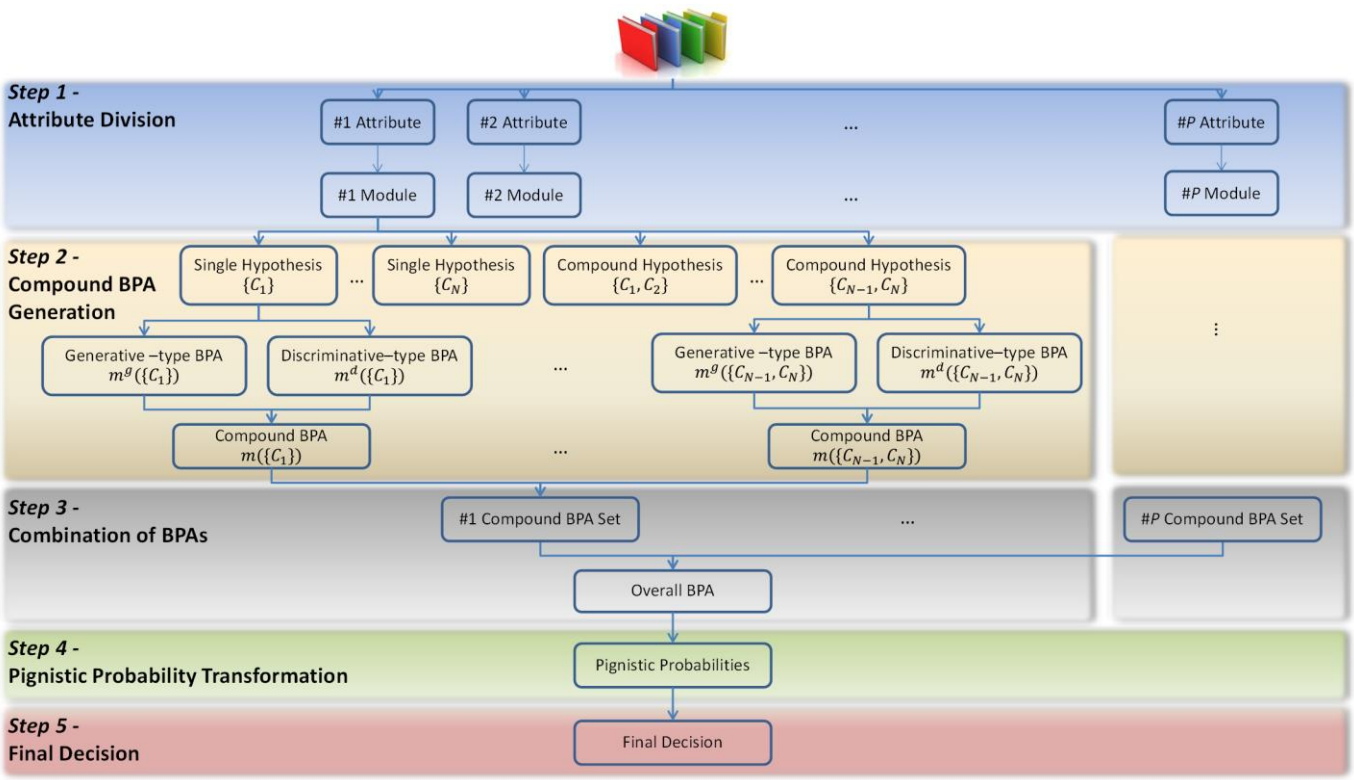

Fig. 1. Procedure of the WFDSF.

2) Step 2 - Compound BPA Generation

To present the procedure to determine BPAs, we take C as the frame of discernment:

$$\Theta = C = \{C_1, C_2, \ldots, C_N\} \tag{21}$$

Further, the focal elements of the power set of the frame of discernment, 2^{Θ}, are denoted by

$$\Omega = \{\{C_1\}, \{C_2\}, \ldots, \{C_N\}, \{C_1, C_2\}, \ldots, \{C_i, C_j\}, \ldots \{C_{N-1}, C_N\}\} \tag{22}$$

where a compound element $\{C_i, C_j\}$, $i \neq j$, is an uncertain hypothesis in D-S formalism, and in this study we do not consider focal elements with cardinality more than two.

Unlike previous studies [24]–[26] using the complementary notion to determine BPAs for compound hypotheses, this study employs an intuitionistic belief assignment to model the region of uncertainty between hypotheses/classes. Figure 2 represents the density of class i (green) and

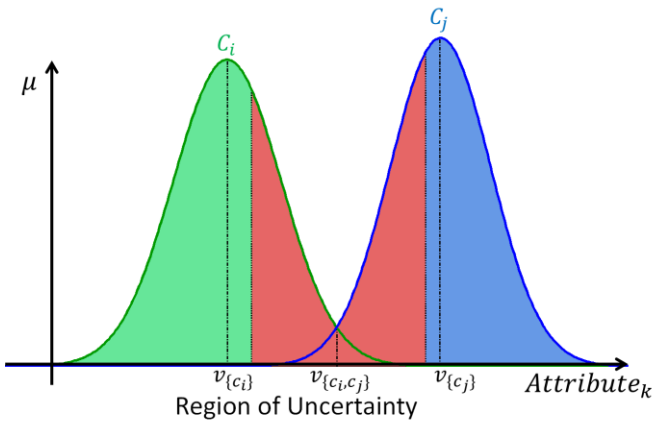

Fig. 2. Gaussian distributions of different classes, where the green curve and the blue curve represent the Gaussian distribution corresponding to c_i and c_j, respectively. The red region is defined as the Region of Uncertainty (ROU) in this example.

class j (blue) based on the k^{th} attribute, where each class is modeled by a Gaussian distribution. The overlapped region is depicted using red color, we call this region as Region of Uncertainty (ROU). It is clear that samples falling in the ROU will be hard to discriminate because they simultaneously possess properties of two different classes. Specifically, samples falling in the ROU are likely to result in classification error in the recognition task. Thus, it is natural to represent the ROU by the compound hypothesis $\{C_i, C_j\}$. For each attribute, N Gaussian distribution functions and C_2^N ROU functions can be obtained as models of different single and compound hypotheses, respectively.

We use the fuzzy Naïve Bayes method [32], [33] and the nearest mean classification rule [25], [26], [34] to calculate the basic probability of each focal element. For each attribute, fuzzy membership functions $\{\mu_{\{c_1\}}, \mu_{\{c_2\}}, \ldots \mu_{\{c_N\}}\}$ are obtained to represent the level of membership of each object to different classes when represented by independent individual attributes. Given an incoming sample, for feature x, the calculated membership value is

$$\mu_{\{C_i\}}(x) = exp(-\frac{[x-m_i]^2}{2\sigma_i^2}) \quad (23)$$

where m_i and σ_i^2 are the mean and variance of the Gaussian membership function, respectively, of

the C_i class.

Since for a compound hypothesis, $\{C_i, C_j\}$. , the object can belong to either of class C_i or class C_j, a fuzzy AND operator is used to assign the mass associated with $\{C_i, C_j\}$. Thus, the generative basic probability of each hypothesis calculated by the fuzzy Naïve Bayes approach is defined as

$$\begin{aligned} m_x^g(\{C_i\}) &= \mu_{\{C_i\}}(x) \\ m_x^g(\{C_i, C_j\}) &= \mu_{\{C_i, C_j\}}(x) = \mu_{\{C_i\}}(x) \wedge \mu_{\{C_j\}}(x) \end{aligned} \tag{24}$$

We note here again, that (24) may not lead to a valid BPA without proper normalization. In (24), for $\wedge$ we can use any t-norms [39]. In this study we use minimum as the t-norm.

Now the Euclidean distance between the incoming sample and the centroid vector $\{v_{\{c_1\}}, v_{\{c_2\}}, \ldots v_{\{c_N\}}\}$ of each class is utilized to determine the discriminative basic probability according to the NMC rule. Consequently, we use the ROU as the support for the compound hypothesis, and define the crossover point $v_{\{c_i, c_j\}}$ as the centroid to compute the mass for the compound hypothesis. In particular, the crossover point $v_{\{c_i, c_j\}}$ is defined as the point with the maximum AND value calculated by the distributions of two different hypotheses/classes.

$$v_{\{c_i, c_j\}} = \arg\max \mu_{\{C_i\}}(x) \wedge \mu_{\{C_j\}}(x) \tag{25}$$

where $v_{\{c_i, c_j\}}$ lies in the interval $[v_{\{c_i\}}, v_{\{c_j\}}]$, as depicted in Fig. 2.

One possible way to define the discriminative BPAs is to use an exponential function of distances from the NMC as follows:

$$\begin{aligned} m_x^d(\{C_i\}) &= e^{-d(x, v_{\{C_i\}})} \\ m_x^d(\{C_i, C_j\}) &= e^{-d(x, v_{\{C_i, C_j\}})} \end{aligned} \tag{26}$$

Furthermore, we propose a weighted regulatory architecture to gather different types of evidences and integrate them. The generative type BPAs, $m_x^g(\{\cdot\})$, generated by the fuzzy Naïve

Bayes approach and the discriminative type BPAs, $m_x^d(\{\cdot\})$, calculated by the NMC rule are integrated by

$$\phi_x(\{\cdot\}) = m_x^g(\{\cdot\})^\alpha \bullet m_x^d(\{\cdot\})^\beta \tag{27}$$

where $0 \le \alpha, \beta \le 1$ are regulatory parameters that adaptively determine the importance of the two types of evidences. The weighted regulatory mechanism enables us to use the training data to search for appropriate weights for different sources of evidences. The optimal values of the regulatory parameters are found using grid search by minimizing the training error. We note here that m_x^g and m_x^d are not BPAs per se because the sum of basic assignments may not be equal to one. However, in the combined BPA mentioned below we enforce this condition.

The global BPA for feature x, $m_x(\{\cdot\})$ is defined as:

$$\begin{aligned} m_x(\{C_i\}) &= \frac{\phi_x(\{C_i\})}{L} = \frac{(\mu_{\{C_i\}}(x))^\alpha \bullet (e^{-d(x, v_{\{C_i\}})})^\beta}{L} \\ m_x(\{C_i, C_j\}) &= \frac{\phi_x(\{C_i, C_j\})}{L} = \frac{(\mu_{\{C_i, C_j\}}(x))^\alpha \bullet (e^{-d(x, v_{\{C_i, C_j\}})})^\beta}{L} \end{aligned} \tag{28}$$

where L is a normalizing factor to make (28) a valid BPA.

$$L = \sum_{i=1}^{N} \phi_x(\{C_i\}) + \sum_{i=1}^{M} \sum_{j=1}^{M} \phi_x(\{C_i, C_j\}) \tag{29}$$

3) Step 3 - Combination of BPAs

The BPAs generated from different attributes/features are then combined to get the overall BPA using Dempster's rule of combination [8]. According to the formula (10) -(12), we can reap an overall BPA from each independent source of information.

4) Step 4 - Pignistic Probability Transformation

After combination of all BPAs, based on (15), the overall BPA is transformed into a pignistic

probability focusing on singletons for decision making.

5) Step 5 – Final Decision

Finally, decision is made using the pignistic transformation. The hypothesis (the class) with the maximum pignistic probability is chosen as the predicted class of the sample in the test data.

The proposed method quantifies the evidence from each information source and assigns basic probabilities on single hypotheses as well as on compound hypotheses. In our approach, we use the ROU to define a compound hypothesis, which can help to avoid overestimation of uncertainty compared to the use of the complementary approach. In addition, to exploit the characteristics of different sources, we employ a weighted regulatory architecture to assign mass values on single and compound hypotheses. We use a training mechanism to find the appropriate weights, i.e., α and β, that should be used for different types of evidences. To further strengthen the WFDSF, we can use different classification approaches to determine BPAs. However, the appropriate selection of classification approaches is not the main focus of this study.

IV. Experiments – Benchmark Data

The proposed WFDSF in this paper has been successfully validated on twelve real-world datasets from the UCI machine learning repository [31]. Before presenting the experimental results, we will step through the BPA assignment portion of the WFDSF method. To do so, we provide a simple numeral example on the calculation of BPAs using the Iris dataset [40]. After completing the numerical example, the proposed method is compared with eight state-of-the-art classifiers and three fusion approaches based on D-S theory to demonstrate the improvement realized by the proposed method.

A. Computation of Basic Probabilities

In this section, we utilize the Iris dataset to demonstrate the WFDSF procedure for computation of BPA. The Iris dataset involves three different classes of Iris flowers: 1) *Setosa*, 2) *Versicolour*, and 3) *Virginica*. It comprises 50 samples with four attributes, sepal length (SL) in cm, sepal width (SW) in cm, petal length (PL) in cm, and petal width (PW) in cm, for each of the three classes. The four attributes are considered four independent sources to provide different information in the experiment. In this example, 25 samples from each class are randomly selected as the training data, the remaining half is used as the test data.

The WFDSF procedure is divided into five steps as shown in Fig. 1. The mechanics of applying this procedure to the Iris dataset is outlined in the five steps shown in Fig. 3: 1) Centroids of

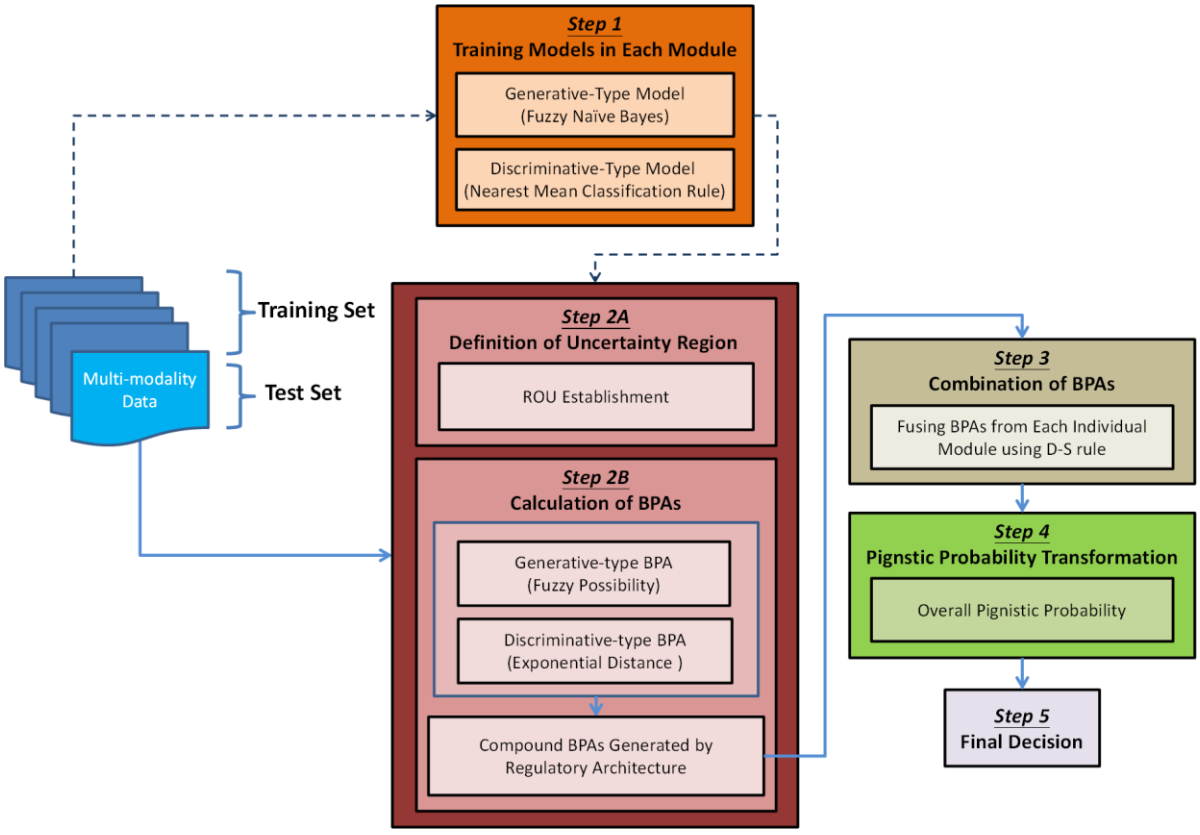

Fig. 3. Generalized procedure of the fusion system based on the WFDSF.

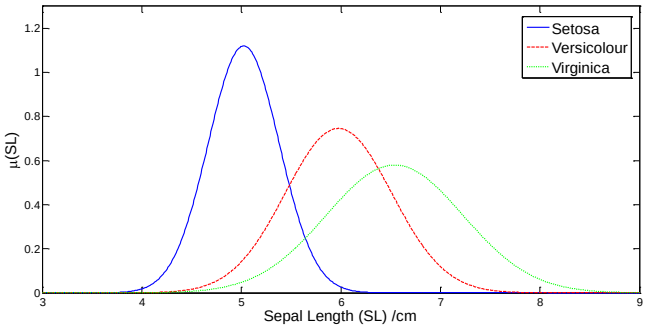

Fig. 4. There are three normal distribution models for the SL attribute. The blue, red and green curves represent distributions of *Setosa*, *Versicolour*, and the *Virginica*, respectively.

different classes used in the fuzzy Naïve Bayes and the NMC classification rule are generated and selected for each attribute in the training set; 2) The corresponding ROU between two classes is defined as explained above and used for mass assignment to compound hypothesis. Samples in the test set are then assigned BPAs by the same procedure. To optimize the WFDSF structure, the weights of the regulatory architecture are optimized with a grid search during the training phase; 3) The BPAs generated from each individual attribute are combined; 4) The overall final BPA is then transformed into a pignistic probability; 5) The hypothesis with the maximum pignistic probability is chosen as the predicted class for decision making. These five steps are described in more detail for the Iris example below:

Step 1 - Training Models in Each Module

In the training phase, four attributes in the training set are divided into four independent modules. The densities of the SL attribute assuming normal distribution in the three classes are illustrated in Fig. 4, and the complete centroid information used in the fuzzy Naïve Bayes and the NMC classification rule is shown in Table I.

Step 2A - Definition of Uncertainty Region

We utilize formula (25) to compute the centroids of ROUs for the associated compound

TABLE I
CENTROIDS OF HYPOTHESES IN THE ATTRIBUTE MODULE

Hypothesis	*Attribute Module*			
	SL	SW	PL	PW
{*Setosa*}	5.0240	3.4400	1.4440	0.2480
{*Versicolour*}	5.9760	2.7880	4.2760	1.3560
{*Virginica*}	6.5400	2.9920	5.5320	2.0080
{*Setosa, Versicolour*}	5.4046	3.1029	2.1356	0.5944
{*Setosa, Virginica*}	5.5409	3.2235	2.2860	0.6458
{*Versicolour, Virginica*}	6.2226	2.8831	4.8353	1.6109

hypothesis in each module (attribute). In this example, for the SL attribute, the centroid of ROU between *Setosa* and *Versicolour* is 5.4046, the centroid of ROU between *Setosa* and *Virginica* is 5.5409, and the centroid of ROU between *Versicolour* and *Virginica* is 6.2226. The detailed centroid information of single and compound hypotheses in each attribute is reported in Table I. Note that the weights of the regulatory architecture have already been updated with a grid search using the training data. During the training, we choose that pair of parameters which minimizes the error on the training data. If sufficient training data are available, one can use a part of that as the validation data and make better choices of the regulatory parameters.

Step 2B - Calculation of BPAs

For each sample in test data, four sets of BPAs are calculated from each individual module. An incoming sample from the *Versicolour* class, whose attributes are: SL = 6.4 cm, SW = 3.2, PL = 4.5 cm and PW = 1.5 cm, is selected as test example. Table II and Table III show the BPAs of each attribute module extracted using the fuzzy Naïve Bayes and the NMC classification rule, respectively. After assigning appropriate regulatory parameters, the compound BPAs obtained are as listed in Table IV.

TABLE II
BPAS OF EACH ATTRIBUTE MODULE EXTRACTED USING THE FUZZY NAÏVE BAYES APPROACH

Hypothesis	*Attribute Module*			
	SL	SW	PL	PW
$m^{\theta}_{x}(\{Setosa\})$	0.0000	0.2761	0.0000	0.0000
$m^{\theta}_{x}(\{Versicolour\})$	0.2661	0.0847	0.8682	0.8678
$m^{\theta}_{x}(\{Virginica\})$	0.4786	0.3275	0.0723	0.0843
$m^{\theta}_{x}(\{Setosa, Versicolour\})$	0.0000	0.0515	0.0000	0.0000
$m^{\theta}_{x}(\{Setosa, Virginica\})$	0.0000	0.1992	0.0000	0.0000
$m^{\theta}_{x}(\{Versicolour, Virginica\})$	0.2553	0.0611	0.0595	0.0479

TABLE III
BPAS OF EACH ATTRIBUTE MODULE EXTRACTED USING THE NMC RULE

Hypothesis	*Attribute Module*			
	SL	SW	PL	PW
$m^{d}_{x}(\{Setosa\})$	0.0741	0.1614	0.0222	0.0822
$m^{d}_{x}(\{Versicolour\})$	0.1921	0.1359	0.3768	0.2489
$m^{d}_{x}(\{Virginica\})$	0.2552	0.1666	0.1680	0.1730
$m^{d}_{x}(\{Setosa, Versicolour\})$	0.1085	0.1862	0.0443	0.1162
$m^{d}_{x}(\{Setosa, Virginica\})$	0.1243	0.2004	0.0515	0.1224
$m^{d}_{x}(\{Versicolour, Virginica\})$	0.2458	0.1495	0.3371	0.2573

Step 3 - Combination of BPAs

The BPAs generated from each individual module are integrated to get the overall BPA using Dempster's rule of combination [8], and we obtain the overall BPA as shown in columns one and two of Table V.

TABLE IV

BPAS OF EACH ATTRIBUTE MODULE EXTRACTED USING COMPOUND APPROACH

Hypothesis	*Attribute Module*			
	SL	SW	PL	PW
$m_x(\{Setosa\})$	0.0159	0.1771	0.0000	0.0000
$m_x(\{Versicolour\})$	0.2547	0.1288	0.5083	0.4486
$m_x(\{Virginica\})$	0.3611	0.1867	0.1694	0.2336
$m_x(\{Setosa, Versicolour\})$	0.0212	0.1633	0.0000	0.0000
$m_x(\{Setosa, Virginica\})$	0.0260	0.2089	0.0000	0.0000
$m_x(\{Versicolour, Virginica\})$	0.3210	0.1352	0.3223	0.3178

The regulatory parameters, α and β, are set to 0.13 and 0.96, respectively, on the Iris data.

TABLE V

THE OVERALL BPA AND THE PIGNISTIC PROBABILITIES

Hypothesis	Overall BPA	*Class*	Pignistic Probabilities
$m_x(\{Setosa\})$	0.0000	$P_{pig}(\{Setosa\})$	0.0000
$m_x(\{Versicolour$	0.6077	$P_{pig}(\{Versicolo$	0.6163
$m_x(\{Virginica\})$	0.3752	$P_{pig}(\{Virginica$	0.3837
$m_x(\{Setosa, Versicolour$	0.0000		
$m_x(\{Setosa, Virginica\})$	0.0000		
$m_x(\{Versicolour$ $Virginica\}$	0.0171		

Step 4 - Pignstic Probability Transformation

The overall BPA is then transformed into a pignistic probability distribution for decision making. In this case, we reap the pignistic probabilities as shown in columns three and four of Table V.

Step 5 - Final Decision

From Table V we find that the hypothesis/class *Versicolour* possesses the maximum pignistic

probability (0.6163); hence, we assign the *Versicolour* class to the incoming data as its predicted class for decision-making after the fusion process, which in this case, coincides with its actual class.

According to the result shown in Table IV, only the PL and the PW provide significant evidences that the incoming datum belongs to the correct class. In contrast, the evidences collected from the remaining attributes mainly contribute to other single or uncertain hypotheses. After combining all the evidences collected from the four sources using the WFDSF, we obtain a collaborative decision that suggests the correct hypothesis. However, this final decision would be different if we utilize the traditional voting strategy to process evidences from individual sources. Specifically, the WFDSF structure enhances the fusion system in two phases: 1) explicitly handling the uncertainty between different classes and 2) efficiently integrating evidences from different sources. Therefore, the WFDSF -based fusion system can provide a reliable and robust decision for the task of classification.

B. Evaluation Methods

The performance of the proposed WFDSF-based fusion system is compared with eight state-of-the-art classifiers and three D-S theory based fusion methods, including Naïve Bayes (NB) [37], nearest mean classifier (NMC) [34], k nearest neighbor (k-NN) [41], Decision Tree (REPTree) [42], support vector machine (SVM) [43], SVM with radial basis function (SVM-RBF) [43], multilayer perceptron (MLP) [44], RBF network (RBFN) [45], k nearest neighbor D-S theory (kNN-DST) [26], normal distribution-based classifier (NDBC) [3], and evidential calibration (Evi-Calib) [46].

TABLE VI
BASIC INFORMATION OF THE UCI DATASETS

Dataset	#Samples	#Class	#Attribute	Missing value
Iris	150	3	4	NO
Heart	270	2	13	NO
Wine	178	3	13	NO
Australian	690	2	14	YES
Climate	540	2	18	NO
Hepatitis	155	2	19	YES
Waveform	5000	3	21	NO
Parkinsons	197	2	22	NO
Forest	523	4	27	NO
Ionosphere	351	2	34	NO
Spambase	4601	2	57	YES
Sonar	208	2	60	NO

The proposed WFDSF model has been evaluated extensively using twelve real-world datasets from the UCI machine learning repository [31], including Iris, Heart, Wine, Australian, Climate, Hepatitis, Waveform, Parkinsons, Forest, Ionosphere, Spambase and Sonar. Three datasets (Australian, Hepatitis and Spambase) contain missing values for some attributes. One of the major advantages of the WFDSF is that these missing values can be addressed as ignorance for the actual state of the corresponding variables in the framework of the evidence theory. Specifically, if x_i is a missing value for the attribute i of the total P attributes, we only collect the evidences (BPAs) from the remaining $P - 1$ attributes, which means that we regard this attribute as ignorance, $m(C) = 1$. Thus, there is no requirement for any external management of the missing values. The general information of these UCI datasets is described in Table VI.

TABLE VII

ACCURACY WITH PERCENTAGE (%) OF COMPARATIVE MODELS ON TWELVE UCI DATASETS

Dataset	Classifiers								D-S Theory Based Fusion Methods			
	NB	NMC	k-NN	REPTree	SVM	SVM-RBF	MLP	RBFN	kNN-DST	NDBC	Evi-Calib	WFDSF (Proposed)
Iris	94.67	90.67	95.33	92.00	94.67	94.67	93.33	92.67	95.33	94.00	94.67	**96.00**
Heart	82.59	60.37	57.78	70.74	83.70	82.96	75.19	81.85	76.30	82.59	83.70	**85.56**
Wine	95.51	70.44	70.19	84.92	96.62	96.63	94.93	95.49	93.84	96.63	97.17	**98.32**
Australian	79.56	64.21	67.40	80.59	80.29	79.86	82.32	82.61	78.41	80.01	80.60	**85.20**
Climate	89.63	82.79	87.78	89.45	93.71	93.89	91.48	92.59	87.22	85.93	94.82	**95.55**
Hepatitis	76.76	52.14	65.71	71.64	79.96	76.76	74.93	81.32	80.57	79.40	79.88	**83.85**
Waveform	80.80	79.98	77.98	75.06	80.86	85.56	72.62	83.10	77.94	79.74	80.78	**86.71**
Parkinsons	68.75	70.77	83.02	80.94	70.13	81.03	74.39	82.05	78..01	70.26	81.64	**88.10**
Forest	85.47	82.58	88.34	81.05	86.24	81.80	81.26	82.80	83.56	85.67	86.62	**89.05**
Ionosphere	80.91	78.07	83.78	84.57	82.04	80.33	83.48	76.57	86.03	81.49	82.33	**86.89**
Spambase	81.61	67.64	80.29	83.59	82.00	82.92	66.46	70.66	79.50	80.00	81.74	**85.89**
Sonar	66.28	60.12	79.78	70.70	66.32	65.90	66.86	66.84	**79.81**	72.57	68.26	77.02
Average	81.88	71.65	78.12	80.44	83.05	83.53	79.77	82.38	83.50	82.36	84.35	**88.18**
STD	8.53	10.92	10.42	6.82	8.73	8.15	9.32	8.15	6.13	7.26	7.69	5.68

For each comparative model, a five-fold cross validation is applied, i.e., 80% of the data (four folds) are randomly selected to build the training data set, while the remaining one fold serves as test data. This process is repeated 5 times, and the average of the five runs is then used for comparison.

C. Evaluation Results

The average classification accuracy (%) on the twelve UCI datasets for the proposed scheme is

TABLE VIII

TOTAL DIFFERENCE WITH PERCENTAGE (%) FROM THE MAX PERFORMANCE ACROSS TWELVE UCI DATASETS

Dataset	Classifiers								D-S Theory Based Fusion Methods			
	NB	NMC	k-NN	REP Tree	SVM	SVM-RBF	MLP	RBFN	kNN-DST	NDBC	Evi-Calib	WFDSF (Proposed)
Iris	1.33	5.33	0.67	4.00	1.33	1.33	2.67	3.33	0.67	2.00	1.33	**0.00**
Heart	2.96	25.19	27.78	14.82	1.85	2.59	10.37	3.70	9.26	2.96	1.85	**0.00**
Wine	2.81	27.88	28.13	13.40	1.70	1.68	3.38	2.83	4.48	1.68	1.14	**0.00**
Australian	5.64	20.98	17.80	4.61	4.91	5.33	2.88	2.58	6.79	5.19	4.60	**0.00**
Climate	5.92	12.76	7.77	6.10	1.84	1.66	4.07	2.96	8.33	9.62	0.73	**0.00**
Hepatitis	7.08	31.71	18.14	12.20	3.88	7.08	8.91	2.53	3.27	4.45	3.97	**0.00**
Waveform	5.92	6.73	8.74	11.66	5.85	0.15	14.09	3.62	8.78	6.98	5.94	**0.00**
Parkinsons	19.35	17.33	5.07	7.15	17.96	7.06	13.70	6.05	10.09	17.84	6.46	**0.00**
Forest	3.59	6.48	0.71	8.00	2.81	7.25	7.79	6.25	5.49	3.39	2.44	**0.00**
Ionosphere	5.97	8.82	3.10	2.31	4.84	6.56	3.40	10.31	0.86	5.39	4.56	**0.00**
Spambase	4.28	18.26	5.61	2.30	3.89	2.98	19.47	15.24	6.39	5.89	4.15	**0.00**
Sonar	13.53	19.70	0.04	9.12	13.50	13.91	12.95	12.97	**0.00**	7.25	11.56	2.79
Accumulated Difference from Max Accuracy	78.38	201.16	123.54	95.67	64.36	57.60	103.69	72.38	64.40	72.63	48.71	**2.79**

88.18 ± 5.68, which is higher than any of the other state-of-the-art approaches (Table VII). In Table VII, the best performance is shown in bold face. For all but one dataset, WFDSF outperforms other methods. The result demonstrates that the WFDSF based fusion system yields the highest accuracy on real-world datasets compared to 11 other models. To better show the improvement obtained with the proposed WFDSF, we also calculate the difference between each classifier's average performance and the max average performance for each dataset (Table VIII). Consequently, we can accumulate those differences across the twelve datasets to quantify the relative performance of

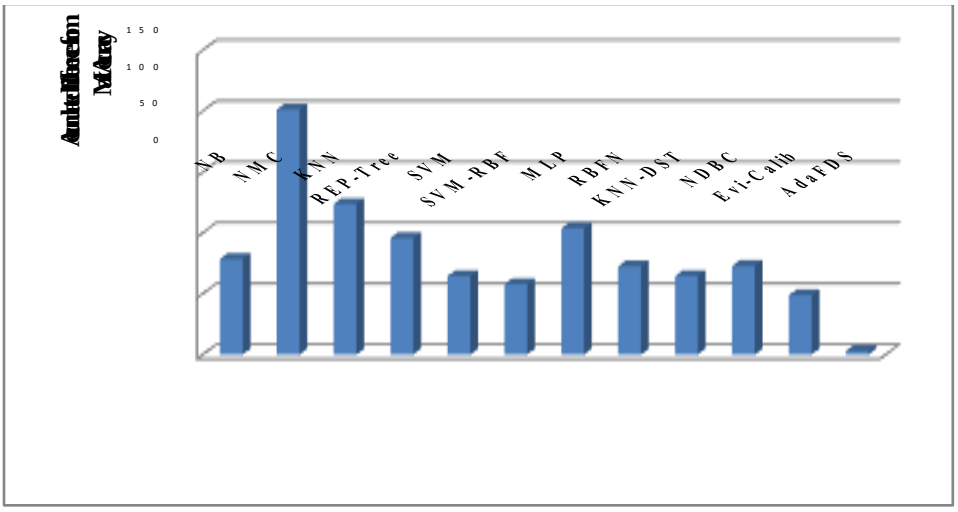

Fig. 5. Accumulated Difference from Max Accuracy, where the blue bar represents the accumulated difference from max accuracy across twelve UCI datasets.

each classifier across the 12 data sets (Fig. 5). For the WFDSF-based fusion system, the total accumulated difference (%) across twelve datasets is only 2.79 since the WFDSF achieves the best performance in eleven of the twelve cases and is only slightly worse than the top performing classifiers in the last data set. Compared to eight state-of-the-art classifiers and three D-S theory based fusion methods, the total difference from best performance for WFDSF is approximately 17.5 times smaller than the next best model, Evi-Calib.

To determine whether the improvement resulting from employing WFDSF is statistically significant or not, experimental results were subjected to analysis of variance (One-way ANOVA) followed by the Tukey-Kramer multiple comparisons test [47] to calculate a *p*-value for each comparative model on twelve UCI datasets. Using a one-way ANOVA for each data set independently, the result of the ANOVA analysis on each dataset shows significant differences for all compared models (all $p < 0.05$). The Tukey-Kramer multiple comparisons test is then conducted (post hoc analysis) to reveal that the performance of the proposed WFDSF is significantly better than many of the compared groups. The results of a post-hoc Tukey-Kramer test are shown in Table IX. Models indicated with a "+" showed significant differences from the performance

TABLE IX
ANOVA-TUKEY-KRAMER MULTIPLE COMPARISONS TEST ON TWELVE UCI DATASETS

Dataset	Classifiers								D-S Theory Based Fusion Methods		
	NB	NMC	k-NN	REP Tree	SVM	SVM-RBF	MLP	RBFN	kNN-DST	Evi-Calib	NDBC
Iris		+									
Heart		+	+	+			+		+		
Wine		+	+	+							
Australian	+	+	+	+	+	+			+	+	+
Climate	+	+	+	+			+		+	+	
Hepatitis		+	+	+			+				
Waveform	+	+	+	+	+		+	+	+	+	+
Parkinsons	+	+			+		+		+	+	
Forest		+		+		+	+	+	+		
Ionosphere		+						+			
Spambase	+	+	+		+		+	+	+	+	+
Sonar	+	+			+	+	+	+		+	

Symbol + refers to a situation in which the average accuracy of the proposed WFDSF is statistically different compared to a comparative model suggested by post-hoc comparisons. (ANOVA-Tukey-Kramer test, p-value < 0.05). In 5 out of 6 data sets, WFDSF was the highest performing classifier when considering the raw average performance. In all 6 data sets, WFDSF was not significantly different from the top performing classifier.

achieved by WFDSF. Although the average accuracy of WFDSF is worse than k-NN and kNN-DST on sonar data, the statistical analysis does not show a significant difference in comparison with these two models. Thus, the Tukey-Kramer multiple comparisons test indicates that our WFDSF is either the top performing classifier or not significantly different from the top performing classifier for all twelve datasets.

V. EXPERIMENT – REAL PHYSIOLOGICAL DATA

In this section, we consider a practical BCI experiment involving fusion of multi-modal

information arising from various physiological signals, such as electroencephalography (EEG) and eye movement, for assessment of driver's cognitive state during a driving task. In every instant of our busy lives, myriad data can be acquired through many different sensors, and different sensors have their own limitation and associated uncertainty. Therefore, the fusion of information from such sources is expected to yield better decisions with a reduction in overall uncertainty. In this study, EEG as well as eye movement data are simultaneously exploited for a recognition task to demonstrate the performance and effectiveness of the proposed WFDSF approach. Since different levels of uncertainty are embedded within disparate EEG and eye movement data, we expect that WFDSF would be able to deal with these uncertainties to obtain superior decisions.

A. Virtual-reality based Highway Driving Paradigm

The cognitive states of drivers significantly affect driving safety; in particular, fatigue driving, or drowsy driving, endangers both the driver and the public. In this study, to investigate the changes in the drivers' cognitive states, a dynamic motion simulator was built with virtual-reality (VR) technologies [48], [49]. This driving simulator included a real car mounted on a dynamic 6-axis motion platform and 360o highway-driving scenes was rendered by seven LCD projectors. All scenes in this study simulated the participants driving on a four-lane divided highway at a constant speed of 100 km/hr at night. During the task, lane perturbation events were randomly introduced to cause the virtual vehicle to drift from the center of the cruising lane. The participants, who had been trained prior to the experiment, were required to drive the vehicle back to the center of the cruising lane as soon as possible after becoming aware of the deviation. The response time (RT) represents the time period between the onset of the deviation and the onset of the response, and is used as an objective measure of the drowsiness level during each lane-departure event. Specifically, the RT would be longer if the driver is drowsier during driving task. Two cognitive

states, alertness and drowsiness, are classified via RT performance of each participant in this study.

Continuous EEG data are recorded from 23 healthy participants (20-28 years of age; mean age: 23.5 years; females and males). All the participants are asked to read and sign an informed consent form before participating in the EEG experiments. The EEG signals are captured from 62 Ag/AgCl scalp electrodes with a unipolar reference at the right earlobe (Compumedics Ltd., VIC, Australia). Subsequently, the raw EEG data are then subjected to a 1-Hz high-pass and a 50-Hz low-pass infinite impulse response (IIR) filter and then downsampled to 250 Hz from the sample recording rate of 1000 Hz used during the hardware phase. Artificial contaminations, e.g., muscle artifacts, oversize eye movement and electrical interference from equipment, are further removed from EEG signals using independent component analysis (ICA). Finally, for each channel of interest of the cerebral cortex, the mean delta- (: 1-3 Hz), theta- (: 4-7 Hz), alpha- (: 8-12 Hz), and beta- (: 13-20 Hz) band powers are used for power spectra analysis and feature extraction.

In the literature, the efficacy of EEG-based brain-computer interfaces (BCIs) in recognition tasks has been limited by low resolutions [50]. The integration of coincident data streams from multiple physiological (e.g., EEG, eye movements and cardiac activity) sources is seen as having the potential to dramatically enhance the interpretation and use of brain-based recordings. Compared with previous studies that utilized only EEG signals for recognition tasks [49], this study further exploited eye tracking to analyze visual information. Participants' eye movement are recorded using the eye tracker (SMI iView X, Senso Motoric Instruments). The eye trackers measured the position of the pupils in the eyeballs with respect to the corneal reflex. The position of the Participants' eyes position is recorded at a rate of 500 Hz. In this study, the average times of the participants' saccade and fixation intervals are selected as features of interest to assess driver's

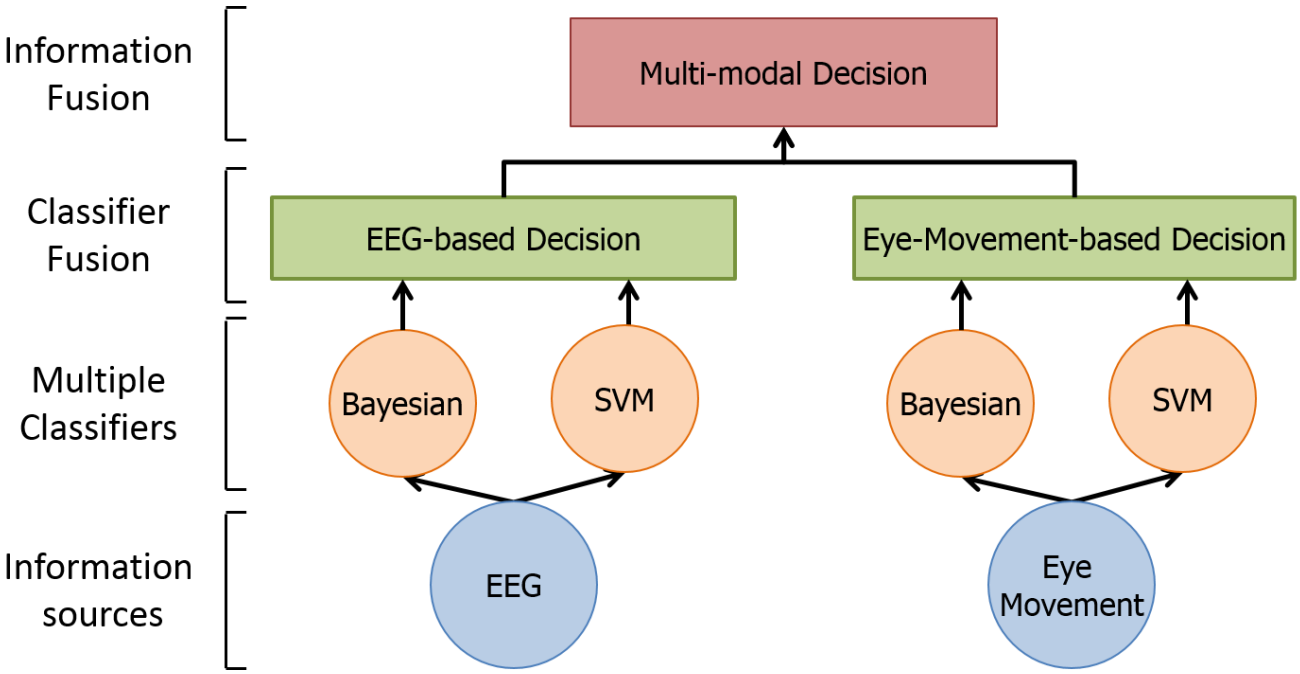

Fig. 6. The structure of multi-modal information fusion based on the proposed WFDSF.

cognitive states.

B. Evaluation Results

In this study, the efficacy of the proposed WFDSF in multi-modal information fusion is demonstrated in a typical BCI application, which recognizes the cognitive states of participants as alert or drowsy during a realistic lane-departure driving task. To establish multi-view estimation, the simultaneously recorded EEG and eye movement signals are used to build an ensemble of Bayesian classifiers and support vector machines (SVM), respectively. A structure of multi-modal information fusion based on the proposed system is shown in Fig. 6. This structure exploits a two-stage hierarchical mechanism for information integration, including classifier fusion and information fusion. For each modality (EEG or Eye-movement data), two classifiers are designed and then WFDSF is used to aggregate the evidences coming from these classifiers.

TABLE X
ACCURACY WITH PERCENTAGE (%) OF COMPARATIVE MODELS ON MULTI-MODAL DATA

Modality	EEG		Eye Movement		Fusion
	Bayesian	SVM	Bayesian	SVM	WFDSF (Proposed)
Average	61.29	67.44	60.97	65.91	**74.22**
STD	1.99	1.58	0.97	2.55	**0.64**

Table X depicts the classification results of different comparative models obtained by five-fold cross-validation. In Table X, the best performance is shown in bold face. The average accuracies of the Bayesian classifier and the SVM using EEG signals alone are 61.29±1.99% and 67.44±1.58%, respectively. On the other hand, the classification performance of these two classifiers using eye movement signal alone are 60.97±0.97% and 65.91±2.55%, respectively. The average classification accuracy (%) obtained using the multi-modal data by the WFDSF-based fusion system is 72.22±0.64, which is the highest accuracy compared with the other classifiers before the fusion stage (Table X). These results suggest that the use of multi-modal information using the proposed WFDSF approach can effectively enhance the performance of BCI.

VI. CONCLUSION

With the increasing prevalence of sensors (e.g. smartphones, traffic/security cameras, and social media) in the environment and in our everyday lives, classification problems have emerged in a wide range of domains. Developing novel techniques to combine multi-modal information to produce highly accurate classification is vital. The D-S theory is one of the useful and popular approaches in the discipline of data fusion. However, the BPA functions in the D-S theory architecture are usually designed heuristically based on the characteristic of the collected dataset.

In this paper, we propose a novel fusion system called WFDSF to boost the accuracy in classification problems. Our WFDSF-based fusion system uses a novel compound model for BPA

assignments, which combines the generative-type and the discriminative-type approaches. Unlike most of the existing methods, the proposed method applies a plausible mathematical structure to deploy the weights of evidences. In addition, an intuitionistic belief assignment is employed to capture uncertainties between classes, which can directly represent uncertainties and imprecision during the classification process. The proposed WFDSF model has been evaluated extensively using twelve real-world datasets from the UCI machine learning repository and its performance has been compared with eleven different methods. The performance of the WFDSF is found to be statistically superior compared to those of the 11 methods. In addition, the WFDSF model has also been evaluated to integrate multi-modal physiological signals for assessing the cognitive states of a driver during a driving task. The experimental results show that the proposed fusion model can produce a reliable and robust decision compared with that produced by single models before the fusion stage. These experimental results suggest that our weighted regulatory mechanism to aggregate evidences from different types of approaches for BPA calculation can bring a significant improvement in performance. The major benefit gained from the WFDSF is that the system can be furnished with information of high quality concerning, possibly, certain aspects of the environment which cannot be sensed directly by any individual sensor operating independently. The performance of the proposed fusion system indicates that the WFDSF model for multi-modal information integration is reasonable and effective.

The proposed WFDSF is expected to be particularly useful in data fusion applications where centralized decisions based on data coming from disparate sensor sources need to be aggregated in order to achieve a joint decision. Our future plan includes 1) developing an effective learning mechanism to dynamically determine weights used in the weighted regulatory architecture to enable online adaptation, 2) creating an appropriate framework to allocate confidence for different

sensory sources and 3) establishing a comprehensive approach to handle different kinds of uncertainty.

ACKNOWLEDGEMENTS

This work was supported in part by the Aiming for the Top University Plan of National Chiao Tung University sponsored by the Ministry of Education, Taiwan, under Grant Number: 104W963, in part by the UST-UCSD International Center of Excellence in Advanced Bio-engineering sponsored by the Taiwan National Science Council I-RiCE Program under Grant Number: MOST 103-2911-I-009-101 and MOST 103-2627-E-009-001, in part by the Cognition and Neuroergonomics Collaborative Technology Alliance Annual Program Plan sponsored by the Army Research Laboratory under Cooperative Agreement Number: W911NF-10-2-0022, and in part by the Office of the Secretary of Defense Autonomy Research Pilot Initiative program MIPR DWAM31168.

The authors would like to thank Prof. Jyh-Yeong Chang and all members at Brain Research Center, National Chiao Tung University, Taiwan.

REFERENCES

[1] R.-C. Luo and C.-C. Chang, "Multisensor fusion and integration: A review on approaches and its applications in mechatronics," *IEEE Trans. Ind. Informatics*, vol. 8, no. 1, pp. 49–60, 2012.

[2] M. Tabassian, R. Ghaderi, and R. Ebrahimpour, "Combining complementary information sources in the Dempster-Shafer framework for solving classification problems with imperfect labels," *Knowledge-Based Syst.*, vol. 27, pp. 92–102, 2012.

[3] P. Xu, Y. Deng, X. Su, and S. Mahadevan, "A new method to determine basic probability assignment from training data," *Knowledge-Based Syst.*, vol. 46, pp. 69–80, 2013.

[4] Y. Xia, H. Leung, and E. Bossé, "Neural Data Fusion Algorithms Based on a Linearly Constrained Least Square Method," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 320 – 329, 2002.

[5] O. Basir, F. Karray, and H. Zhu, "Connectionist-based Dempster-Shafer evidential reasoning for data fusion," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1513–1530, 2005.

[6] L. Gupta, B. Chung, M. D. Srinath, D. L. Molfese, and H. Kook, "Multichannel fusion models for the parametric classification of differential brain activity," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1869–1881, 2005.

[7] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[8] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[9] A. P. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Ann. Math. Stat.*, vol. 38, no. 2, pp. 325–339, 1967.

[10] E. Barrenechea, J. Fernandez, M. Pagola, F. Chiclana, and H. Bustince, "Construction of interval-valued fuzzy preference relations from ignorance functions and fuzzy preference relations. Application to decision making," *Knowledge-Based Syst.*, vol. 58, pp. 33–44, 2014.

[11] A. Laha, N. R. Pal, and J. Das, "Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1633 – 1641, 2006.

[12] Y. Deng and F. T. S. Chan, "A new fuzzy dempster MCDM method and its application in supplier selection," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9854–9861, 2011.

[13] J. Ghasemi, R. Ghaderi, M. R. Karami Mollaei, and S. A. Hojjatoleslami, "A novel fuzzy Dempster-Shafer inference system for brain MRI segmentation," *Inf. Sci.*, vol. 223, pp. 205–220, 2013.

[14] D. Hou, H. He, P. Huang, G. Zhang, and H. Loaiciga, "Detection of water-quality contamination events based on multi-sensor fusion using an extented Dempster–Shafer method," *Meas. Sci. Technol.*, vol. 24, no. 5, p. 055801, 2013.

[15] Z. I. Petrou, V. Kosmidou, I. Manakos, T. Stathaki, M. Adamo, C. Tarantino, V. Tomaselli, P. Blonda, and M. Petrou, "A rule-based classification methodology to handle uncertainty in habitat mapping employing evidential reasoning and fuzzy logic," *Pattern Recognit. Lett.*, vol. 48, no. 15, pp. 24–33, 2013.

[16] A. Talavera, R. Aguasca, B. Galván, and A. Cacereño, "Application of Dempster-Shafer theory for the quantification and propagation of the uncertainty caused by the use of AIS data," *Reliab. Eng. Syst. Saf.*, vol. 111, pp. 95–105, 2013.

[17] T. Denoeux, N. El Zoghby, V. Cherfaoui, and A. Jouglet, "Optimal Object Association in the Dempster-Shafer Framework," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2521 – 2531, 2014.

[18] R.-C. Luo and C.-C. Lai, "Multisensor fusion-based concurrent environment mapping and moving object detection for intelligent service robotics," *IEEE Trans. Ind. Electron.*, vol. 61, no. 8, pp. 4043–4051, 2014.

[19] T. Denœux, "Construction of predictive belief functions using a frequentist approach," in *Proceedings of IPMU*, 2006, pp. 1412–1419.

[20] R. R. Yager, "An intuitionistic view of the Dempster-Shafer belief structure," *Soft Comput.*, vol. 18, no. 11, pp. 2091–2099, 2014.

[21] Y. M. Zhu, "Automatic determination of mass functions in Dempster-Shafer theory using fuzzy c-means and spatial neighborhood information for image segmentation," *Opt. Eng.*, vol. 41, no. 4, pp. 760–770, 2002.

[22] M.-H. Masson and T. Denoeux, "ECM: An evidential version of the fuzzy c-means algorithm," *Pattern Recognit.*, vol. 41, no. 4, pp. 1384–1397, 2008.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[24] T. Denoeux, "A neural network classifier based on Dempster-Shafer theory," *IEEE Trans. Syst. Man Cybern. Part A-Syst. Humans*, vol. 30, no. 2, pp. 131 – 150, 2000.

[25] N. R. Pal and S. Ghosh, "Some classification algorithms integrating Dempster-Shafer theory of evidence with the rank nearest neighbor rules," *IEEE Trans. Syst. Man Cybern. Part A-Syst. Humans*, vol. 31, no. 1, pp. 59–66, 2001.

[26] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Trans. Syst. Man Cybern.*, vol. 25, no. 5, pp. 804 – 813, 1995.

[27] A. O. Boudraa, A. Bentabet, F. Salzenstein, and L. Guillon, "Dempster-Shafer's Basic Probability Assignment Based on Fuzzy Membership Functions," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 4, no. 1, pp. 1–9, 2004.

[28] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.

[29] T. Denœux and P. Smets, "Classification using belief functions: Relationship between case-based and model-based approaches," *IEEE Trans. Syst. Man Cybern. Part B-Cybern.*, vol. 36, no. 6, pp. 1395–1405, 2006.

[30] C. Zhang, Y. Hu, F. T. S. Chan, R. Sadiq, and Y. Deng, "A new method to determine basic probability assignment using core samples," *Knowledge-Based Syst.*, vol. 69, pp. 140–149, 2014.

[31] M. Lichman, "UCI Machine Learning Repository." 2013.

[32] J. Ghasemi, R. Ghaderi, M. R. Karami Mollaei, and S. A. Hojjatoleslami, "A novel fuzzy Dempster-Shafer inference system for brain MRI segmentation," *Inf. Sci.*, vol. 223, no. 20, pp. 205–220, 2013.

[33] Y. Tang, W. Pan, H. Li, and Y. Xu, "Fuzzy Naive Bayes classifier based on fuzzy clustering," in *IEEE International Conference on Systems, Man and Cybernetics*, 2002, vol. 5, pp. 6–9.

[34] C. J. Veenman and M. J. T. Reinders, "The nearest subclass classifier: A compromise between the nearest mean and nearest neighbor classifier," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1417–1429, 2005.

[35] R. R. Yager and N. Alajlan, "Multicriteria Decision-Making With Imprecise Importance Weights," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 4, pp. 882–891, 2014.

[36] P. Smets, “Decision making in the TBM: The necessity of the pignistic transformation,” *Int. J. Approx. Reason.*, vol. 38, no. 2, pp. 133–147, 2005.

[37] B. G. Hu, “What are the differences between bayesian classifiers and mutual-information classifiers?,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 249–264, 2014.

[38] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[39] D. Dubois and H. Prade, “A review of fuzzy set aggregation connectives,” *Inf. Sci.*, vol. 36, no. 1–2, pp. 85–121, 1985.

[40] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.

[41] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21 – 27, 1967.

[42] Y. Freund and L. Mason, “The alternating decision tree learning algorithm,” in *International Conference on Machine Learning*, 1999, vol. 99, pp. 124–133.

[43] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011.

[44] C. L. Castro and A. P. Braga, “Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 888–899, 2013.

[45] S. Chen, C. F. N. Cowan, and P. M. Grant, “Orthogonal least squares learning algorithm for radial basis function networks,” *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, 1991.

[46] P. Xu, F. Davoine, H. Zha, and T. Denoeux, “Evidential calibration of binary SVM classifiers,” *Int. J. Approx. Reason.*, vol. 72, pp. 55–70, 2015.

[47] J. W. Tukey, “Comparing Individual Means in the Analysis of Variance,” *Biometrics*, vol. 5, no. 2, pp. 99–114, 1949.

[48] C.-H. Chuang, L.-W. Ko, T.-P. Jung, and C.-T. Lin, “Kinesthesia in a sustained-attention driving task,” *Neuroimage*, vol. 91, no. 1, pp. 187–202, 2014.

[49] Y.-T. Liu, Y.-Y. Lin, S.-L. Wu, C.-H. Chuang, and C.-T. Lin, "Brain Dynamics in Predicting Driving Fatigue Using a Recurrent Self-Evolving Fuzzy Neural Network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 347–360, 2016.

[50] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. 1–13, 2007.