

# Evaluation of Fuzzy Measures Using Dempster–Shafer Belief Structure: A Classifier Fusion Framework

Pratik Bhowal , Subhankar Sen, Jin Hee Yoon , *Member, IEEE*, Zong Woo Geem , *Senior Member, IEEE*, and Ram Sarkar , *Senior Member, IEEE*

**Abstract**—This article studies the high complexity of the calculation of fuzzy measures that can be used in fuzzy integrals to combine the decisions of different learning algorithms. To this end, this article proposes an alternative low-complexity method for the calculation of fuzzy measures that have been applied to a Choquet integral for the fusion of deep learning models across different application domains for increasing the accuracy of the overall model. This article shows that the Dempster–Shafer (DS) belief structure provides partial information about the fuzzy measures associated with a variable, and this article devises a method to use this partial information for the calculation of fuzzy measures. An infinite number of fuzzy measures are associated with the DS belief structure. This article proposes a theorem to calculate the general form of a specific set of fuzzy measures associated with the DS belief structure. This specific set of fuzzy measures can be expressed as a weighted summation of the basic assignment function of the DS belief structure. The main advantage of expressing the fuzzy measures in this format is that the monotonic condition that needs to be maintained during the calculation of the fuzzy measure can be avoided, and only the basic assignment function needs to be evaluated. The calculation of the basic assignment function is formulated using a method inspired by the Monte Carlo approach used to calculate value functions in the Markov decision process.

**Index Terms**—Choquet integral, classifier combination, deep learning, Dempster–Shafer (DS) theory, fuzzy measure.

Manuscript received 6 March 2022; revised 17 July 2022; accepted 5 September 2022. Date of publication 14 September 2022; date of current version 3 May 2023. This work was supported by the National Research Foundation of Korea under Grant 2020R1A2C1A01011131 funded by the Korea Government (Ministry of Science and ICT). (Corresponding authors: Zong Woo Geem; Jin Hee Yoon.)

Pratik Bhowal is with the Department of Instrumentation and Electronics Engineering, Jadavpur University, Kolkata 700032, India (e-mail: pratikbhowal1999@gmail.com).

Subhankar Sen is with the Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur 303007, India (e-mail: subhankarsen2001@gmail.com).

Jin Hee Yoon is with the Department of Mathematics and Statistics, Sejong University, Seoul 05006, South Korea (e-mail: jin9135@sejong.ac.kr).

Zong Woo Geem is with the College of IT Convergence, Gachon University, Seongnam 13120, South Korea (e-mail: geem@gachon.ac.kr).

Ram Sarkar is with the Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India (e-mail: rsarkar@ieee.org).

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TFUZZ.2022.3206504>.

Digital Object Identifier 10.1109/TFUZZ.2022.3206504

## I. INTRODUCTION

**F**USION learning is the aggregation of the decisions of multiple classifiers, learning algorithms, or experts to enhance the ability of the overall learning model. The maximum advantage of fusion learning can be expected when the models selected for forming the ensemble are diverse. This is intuitive because when the classifiers make different errors, a strategic combination of the results may reduce the overall error.

Aggregation operators or functions like weighted arithmetic mean, geometric mean, median, arithmetic mean, etc. are some widely used approaches to fusion learning. However, many problems exist with these commonly used aggregation operators. Some of them are as follows: 1) outliers have a high influence on the mean calculation and 2) these aggregation functions do not consider the correlation among the decisions of the different classifiers or experts.

Keeping the above facts in mind, a more complex aggregation function, called fuzzy integrals, has been proposed in the literature. Fuzzy integrals like Choquet integral [1] and Sugeno integral [2] take into consideration the interaction among the decisions of the different classifiers. In doing so, fuzzy integrals generally assign weights to not only the individual classifiers but also to the subset of classifiers. These weights are known as fuzzy measures.

Before explaining how we are going to use fuzzy measures in the decision-making process, we look at how we mathematically define fuzzy measures [2].

**Definition 1 (Fuzzy measure):** Let  $(X, \Sigma)$  be a measurable space. Then,  $\mu : \Sigma \rightarrow [0, 1]$  is a fuzzy measure if it satisfies

- 1)  $\mu(\phi) = 0, \mu(X) = 1,$
- 2) if  $Q \subset W$ , then  $\mu(Q) \leq \mu(W)$ , where  $Q, W \in \Sigma.$  (1)

The condition shown in the above definition is monotonic; therefore, a fuzzy measure is sometimes called a “monotone measure.”

Now, we explain how we use fuzzy measures in our decision-making process. Let  $X$  be the universe of discourse of classifiers and  $P_j \subset X$  ( $j = 1, 2, 3, \dots$ ) be the different subsets of classifiers.  $\mu(P_j)$  indicates the relative importance of the different subsets of classifiers. In other words, when we have  $\mu(P_1) > \mu(P_2)$ , it indicates that the subset  $P_1$  of classifiers conveys much more relevant information as compared with

the subset  $P_2$  of classifiers. Thus, fuzzy measures, which can be intuitively thought of as weights in the Choquet integral, provide more importance to subsets of classifiers providing higher relevant information as compared with subsets providing lower relevant information. The other constraint, namely the monotonic constraint, is also justified as, whenever,  $P_2 \subset P_1$ ;  $P_1$  would intuitively convey more information than  $P_2$ , as it has a larger number of classifiers as compared to  $P_2$ .

Even though we understand how fuzzy measures can be used in the decision-making process, the calculation of these fuzzy measures is, however, an NP-hard problem. An additive fuzzy measure was proposed to deal with this problem. However, additive fuzzy measures are considered a too simplified approach to handle practical problems. Another popular approximation is the  $\lambda$ -fuzzy measure [2]. However, the  $\lambda$ -fuzzy measure is an approximation. Also, it becomes increasingly difficult to determine the value of  $\lambda$  as the number of classifiers and, hence, the number of fuzzy measures increase. In this article, we propose an approach, which is not NP-hard, to calculate a set of fuzzy measures without any form of approximation, wherein we take the help of the Dempster–Shafer (DS) theory. To explain the set of fuzzy measures that can be associated with the DS belief structure, we elaborate on the DS theory [3].

*Definition 2 (DS belief structure):* Let  $X$  be a universe of discourse. A DS belief structure is a set mapping  $m : 2^X \rightarrow [0, 1]$  satisfying following properties:

$$\begin{aligned} 1) & m(\phi) = 0 \\ 2) & \sum_{B \in 2^X} m(B) = 1. \end{aligned} \quad (2)$$

Here,  $m$  is referred to as the *basic assignment function*, while we refer to the subsets  $B_i$  such that  $m(B_i) > 0$  as *focal elements*. Let  $B_i$  be the subsets of classifiers in the set of classifiers  $X$ . Now, we assign a value to the mass functions of the subset of classifiers depending on how relevant those subsets of classifiers are in the decision-making process. In other words, a higher mass function value should be assigned to a subset of classifiers conveying more relevant information. However, it can be observed that the monotonic condition associated with the fuzzy measures is not considered in the case of the assignment of the mass functions. Another important thing that can be noted here is that when  $B_i \cap P_j \neq \phi$ , the relevant information conveyed by  $P$  can be associated in some ways with the relevant information conveyed by the subset  $B_i$ . Thus,  $m(B_i)$ s would contribute in some ways in the calculation of  $\mu(P_j)$  provided  $B_i \cap P_j \neq \phi$ . In other words, the DS belief structure can provide partial information about the fuzzy measures associated with the subset of classifiers.

In this article, we propose a general form of a set of fuzzy measures that can be associated with both the fuzzy measure theory and the DS belief structure. We have shown that the fuzzy measures can be expressed using the formula  $\mu(P) = \sum_{i=1}^p w_{B_i}^P m(B_i)$ . However,  $w_{B_i}^P$ s have to satisfy some specific properties for  $\mu(P_j)$ s to be fuzzy measures. We have established the properties in the theorem proposed in this article. Furthermore, we have logically chosen a set of weights for evaluating

our method on different datasets belonging to various domains. This set of weights satisfies all the properties mentioned in our theorem.

After we logically set the weights, we calculate the basic assignment functions. In the later sections, we have shown how this calculation is performed by adapting the Monte Carlo (MC) method used for solving the Markov decision process (MDP) in a specific way. In other words, the MC method for solving the MDP has inspired us in developing this novel scoring system.

We prove the efficacy of the evaluated fuzzy measures by applying them in the domain of fuzzy ensemble, by using the Choquet integral, which is used to combine the decisions of the various deep learning models. We use datasets of three diverse domains, namely breast cancer histology classification, chest X-ray (CXR) image classification, and human action recognition.

### A. Motivation and Contributions

As mentioned above, the calculation of the fuzzy measures is an NP-hard problem. Hence, in this article, we aim to find a set of fuzzy measures that are easier to calculate and properly incorporate the uncertainty involved with the variable. The contributions of our work are as follows.

- 1) We propose a novel generalized set of fuzzy measures, which is associated with both the DS belief structure and the fuzzy measure theory. We prove a theorem that provides us with the conditions for these sets of fuzzy measures to exist. Of these fuzzy measures, we select one logical set of fuzzy measures that can easily be calculated and properly used for fusion learning.
- 2) For calculating the basic assignment functions of the DS theory, we have developed a novel scoring system inspired by the MC method for evaluating value functions in MDPs.

The rest of this article is organized as follows. Section II provides a literature survey of the state-of-the-art fuzzy ensemble techniques and advancements in the DS theory. In Section III, we describe the proposed classifier fusion framework. In Section IV, we report the dataset description, experimental results, comparison with the state-of-the-art methods, and error analysis. Finally, Section V concludes this article and provides some future directions.

## II. LITERATURE SURVEY

### A. Fuzzy Ensemble Techniques

The Choquet integral of fuzzy-number-valued functions based on  $\sigma - \lambda$  rules was formulated by Li et al. [4] followed by the proposal of genetic-algorithm-based optimization for computing fuzzy measures on fuzzy-number-valued data. Beliakov and Wu [5] introduced  $k$ -interactivity, which reduces the number of variables and constraints and reduces the complexity of learning fuzzy measures. The fuzzy measures were learned using a linear programming problem. The devised method facilitated the efficient learning of the fuzzy measures, and it significantly outperformed the fuzzy measures that are calculated using  $k$ -additive and  $k$ -maxitive fuzzy measures. For learning monotone models, Tehrani et al. [6] demonstrated the

use of the Choquet integral as an aggregator operator in machine learning problems. Murillo et al. [7] found that hierarchical least mean square (HLMS), a gradient-based algorithm for the identification of fuzzy measures, had several convergence issues. As a result, they suggested a revamped HLMS implementation that enhanced the convergence by improving the formula for the iterative estimation of fuzzy measure coefficients and the monotonicity check.

Mesiar [8] proposed two types of generalizations of  $k$ -order additive discrete fuzzy measures given by Grabisch [9]. The authors proposed a formula for evaluating Choquet-like integrals on finite spaces based on the Mobius-like transform of the discrete fuzzy measure, followed by the proposal of  $k$ -order additive fuzzy measures defined on an arbitrary space. Finally, the two generalizations were combined to obtain a  $k$ -order pan-additive fuzzy measure on a general space. Liu and Kao [10] devised a mathematical programming approach for the derivation of fuzzy measures based on the correlation coefficient. Grabisch [9] conducted another research study that introduced a gradient algorithm for identifying fuzzy measures from empirical data. The paper presented the use of a mix of standard optimization algorithms, such as Lemke's method, and utility theory for identifying the  $2^n$  coefficients based on semantical considerations.

### B. DS Theory

Beynon et al. [11] described the potential offered by the DS theory of evidence as an alternative approach to multicriteria decision modeling. This article describes the DS theory as a generalization of classical probability theory, followed by the incorporation of this theory in a modified version of the analytic hierarchy process (AHP), termed the DS/AHP method. The DS/AHP approach tackles the inherent drawbacks of the standard AHP and allows opinions on sets of decision alternatives. Although this method did not necessarily facilitate obtaining the highest ranked decision alternative, its major contribution laid in the reduction of the number of serious contenders and computational complexity. Yang et al. [12] borrowed the design idea of Schubert's degree of falsity and proposed a distance-based degree of disagreement through which discounting factors can be generated for discounting combinations of unreliable evidence.

Yager [13] inferred that the DS belief structure provided partial information about the underlying fuzzy measure associated with an uncertain variable. Hence, they proposed a class of fuzzy measures that can be seen as possible completions of a DS belief structure. By the introduction of a measure of entropy on a fuzzy measure, they demonstrated that the three noted completions have the same entropy. However, their studies did not explore the general characterizations of the entropy associated with the class of completions proposed in their paper. Based on the distance of basic belief assignments, Martin et al. [14] proposed some conflict measures of a group of experts. Essentially, the conflict is evaluated for one expert against the rest of the group. These measures of conflict were further used for an *a posteriori* estimation of the relative reliability between the sources of information.

The presented measures of conflict and the associated reliability were evaluated and debated on random basic belief assignments and found their application in real radar target recognition tasks.

Wang et al. [15] incorporated ambiguity measure and DS theory of evidence to enhance the fuzzy soft set-based decision-making capabilities. This approach reduced the uncertainty caused by people's subjective cognition and raised the choice decision level. Yu et al. [16] developed the concepts of the supporting probability function and its distance to demonstrate the degree of correlation between evidence in a body of evidence (BOE) and the distances between the BOEs. In addition, a combination method for the conflicting BOEs based on the supporting probability distance was introduced, in which the basic probability mass of conflicting evidence was corrected by the credibility measure calculated via supporting probability distance before using the Dempster combination rule. Zhao et al. [17] proposed an improved combination method for the conflicting evidence based on inconsistent measurements. They introduced a new approach for measuring the conflict between two pieces of evidence, followed by the revision of the conflicting evidence by the selection of discount coefficients.

### III. PROPOSED WORK

Even though we have introduced the required terms in Section I, we define them here again for the convenience of the readers.  $X$  is the universe of discourse of classifiers.  $B_i$ s are the subset of classifiers. Fuzzy measures  $\mu(P)$ s, assigned to each subset of classifiers, determine the importance of that subset of classifier in the decision-making process. In other words, if a higher value is assigned to a subset of classifiers, it specifies that the specific subset provides more relevant information in the decision-making process than the subset that has been assigned a lower value. Similarly,  $m(B_i)$ s also provide us with information regarding the importance of a subset of classifiers with higher values assigned to subsets of classifiers that provide more relevant information and vice versa. As already mentioned in Section I, the DS belief structure can be thought to provide partial information about the fuzzy measures that we aim to calculate here. Moreover, the monotonicity constraint needs to be satisfied by the fuzzy measures, but the mass functions do not need to satisfy it.

The information about the true class label that is conveyed by the subset  $B_i$  of classifiers in the form of mass function, namely  $m(B_i)$ , provides information about the relevant information conveyed by the subset  $P_j$  of classifiers in the form of  $\mu(P_j)$ , provided  $P_j \cap B_i \neq \phi$ , and on the other hand, if  $P_j \cap B_i = \phi$ , then no information is conveyed about  $\mu(P_j)$  by  $m(B_i)$ .

Now, once we have understood that only the *focal elements* with  $P \cap B_i \neq \phi$  can provide partial information about the fuzzy measure, we need to find a way to combine these elements. However, the values derived after the combination must follow the properties of the fuzzy measures. We argue that we can combine the measures using the formula

$$\mu(P) = \sum_{i=1}^p w_{B_i}^P m(B_i). \quad (3)$$

Here,  $P \subseteq X$ ,  $p$  is the number of focal elements,  $m(B_i)$  is the basic assignment function for the focal elements, and  $w_{B_i}^P$  is the weight that determines the relative importance of the focal elements  $B_i$  in conveying information similar to that conveyed by the subset  $P$  of classifiers.

The weights,  $w_{B_i}^P$ , however, cannot be assigned arbitrary values as  $\mu(P)$ , calculated, must satisfy the properties of the fuzzy measures. We calculate the set of such values of  $w_{B_i}^P$  and use the same for our calculation.

However, besides the two constraints of the fuzzy measures, two more constraints are imposed by the DS theory.  $B_i$  may be totally or partially contained in  $P$ . When  $B_i$  is partially contained in  $P$ , we have not decided what portion of  $m(B_i)$  to allocate to  $\mu(P)$ . In (3), the two extreme cases of the allocation method are as follows.

- 1) If  $B_i \subseteq P$ , then we allocate entire  $m(B_i)$  to  $\mu(P)$ . In other words,  $w_{B_i}^P = 1$  only when  $B_i \subseteq P$ , else  $w_{B_i}^P = 0$ .
- 2) If  $B_i \cap P \neq \phi$ , then we allocate  $m(B_i)$  to  $\mu(P)$ . In other words,  $w_{B_i}^P = 1$  only when  $B_i \cap P \neq \phi$ , else  $w_{B_i}^P = 0$ .

Thus, Scheme 1 of allocation is the strictest version of allocating  $m(B_i)$  to  $\mu(P)$ , while Scheme 2 of allocation is the most lenient version of allocation of  $m(B_i)$  to  $\mu(P)$ . These two conditions give rise to the following constraint:

$$\sum_{B_i \subseteq P} m(B_i) \leq \mu(P) \leq \sum_{B_i \cap P \neq \phi} m(B_i). \quad (4)$$

We now derive a set of weights such that  $\mu(P)$  satisfies the three rules to be followed by the fuzzy measures and the two imposed by the DS belief structure mentioned above.

*Theorem 1:* A fuzzy measures  $\mu$  can be defined as

$$\mu(P) = \sum_{i=1}^p w_{B_i}^P m(B_i) \quad (5)$$

provided that the weights  $w_{B_i}^P$  have the following properties.

- 1) The weights must be fuzzy measures themselves.
- 2)  $\sum_{j=1}^n w_{B_j}^P m(B_j) \leq \sum_{i=1}^{p-n} (1 - w_{B_i}^P) m(B_i)$ , where  $n = |\{B_i : B_i \cap P \neq \phi\}|$ .
- 3)  $\sum_{i=1}^k (1 - w_{B_j}^P) m(B_j) \leq \sum_{j=1}^{p-k} w_{B_j}^P m(B_j)$ , where  $k = |\{B_i : B_i \in P\}|$ .

*Proof:* To prove 1), we need to satisfy the following three criteria of the fuzzy measure theory.

- i) If  $\mu(\phi) = 0$ , then  $\sum_{i=1}^p w_{B_i}^\phi m(B_i) = 0$ . Therefore,  $w_{B_i}^\phi = 0$  for all  $i = 1, 2, 3, \dots, p$ .
- ii) If  $\mu(X) = 1$ , then  $\sum_{i=1}^p w_{B_i}^X m(B_i) = 1$ . However, we know that  $\sum_{B_i \subseteq X} m(B_i) = 1$ . Then,  $w_{B_i}^X = 1$  for all  $i = 1, 2, 3, \dots, p$ .
- iii) If  $Q \subset W$ , then  $\mu(Q) \leq \mu(W)$ . Then,  $\sum_{i=1}^p w_{B_i}^Q m(B_i) \leq \sum_{i=1}^p w_{B_i}^W m(B_i)$ , where  $w_{B_i}^Q$  and  $w_{B_i}^W$  are the weights associated with the decomposition of  $\mu(Q)$  and  $\mu(W)$ , respectively. Here,  $\sum_{i=1}^p m(B_i) [w_{B_i}^W - w_{B_i}^Q] \geq 0$ , so  $w_{B_i}^W \geq w_{B_i}^Q$  whenever  $Q \subseteq W$ .

Based on above three criteria, we can see that  $w_{B_i}^\phi = 0$  and  $w_{B_i}^X = 1$ . In addition,  $w_{B_i}^W \geq w_{B_i}^Q$  whenever  $Q \subseteq W$ . Thus,

these indicate that  $w_{B_i}^P$  must be a set of fuzzy measures for  $\mu(P)$  to be a fuzzy measure.

Next, we show the other two conditions, which are imposed by the DS theory.

- 2) Since  $\mu(P) \leq \sum_{B_i \cap P = \phi} m(B_i)$ ,  $\sum_{i=1}^{p-n} w_{B_i}^P m(B_i) + \sum_{j=1}^n w_{B_j}^P m(B_j) \leq \sum_{i=1}^{p-n} m(B_i)$ , where  $n$  is as mentioned above. Therefore,  $\sum_{j=1}^n w_{B_j}^P m(B_j) \leq \sum_{i=1}^{p-n} (1 - w_{B_i}^P) m(B_i)$ .
- 3) Since  $\sum_{B_i \subseteq P} m(B_i) \leq \mu(P)$ ,  $\sum_{i=1}^k m(B_i) \leq \sum_{j=1}^{p-k} w_{B_j}^P m(B_j) + \sum_{i=1}^k w_{B_i}^P m(B_i)$ , where  $k$  is as mentioned above. Therefore,  $\sum_{i=1}^{p-k} (1 - w_{B_j}^P) m(B_j) \leq \sum_{j=1}^k w_{B_j}^P m(B_j)$ .

Thus, we see that (5) represents the fuzzy measures, provided that the weights follow the three conditions mentioned in this theorem. ■

It is to be noted that any fuzzy measure that satisfies (4) would be a fuzzy measure associated with both the fuzzy measure theory and the DS theory. Thus, there can be an infinite number of such fuzzy measures. In (5), we calculate a few such sets of fuzzy measures that can be written in that form. However,  $w_{B_i}^P$ s must follow the three conditions mentioned in the theorem for the quantities to be fuzzy measures associated with both the fuzzy measure theory and the DS theory.

We choose a specific set of weights for evaluating our method on different datasets. We define this set of weights as follows:

$$w_{B_i}^P = \frac{|B_i \cap P|}{|P|} \quad (6)$$

where  $|\cdot|$  is the number of elements of a set  $\cdot$ . It can easily be checked that this set of weights satisfies the three conditions in the theorem.  $m(B_i)$  would not have any effect on the value of  $\mu(P)$  if  $B_i \cap P = \phi$  as  $B_i$  does not provide us with any information about whether information about the true class is being conveyed by the subset  $P$  or not. So,  $w_{B_i}^P$  is assigned to zero. On the other extreme, when  $B_i \cap P = B_i$ , we are provided with the detailed knowledge of  $B_i$ , but we are still uncertain about the elements of  $P$ , which are not part of  $B_i$ . So, we assign a value to  $w_{B_i}^P$ , which is higher but not equal to 1. However, if  $B_i \cap P = P$ , knowing  $B_i$  gives us complete knowledge of  $P$ , that is,  $B_i$  is not providing partial information anymore. Hence,  $w_{B_i}^P$  is assigned a value equal to 1 in this case. For all the other cases when  $B_i \cap P \subset B_i$ , the higher the value of  $|B_i \cap P|$ , the more the information conveyed by  $B_i$ , and hence, higher value must be assigned to  $w_{B_i}^P$ . Since (6) is consistent with this logic, we select (6) as a weighting scheme.

For calculating the fuzzy measures that we would be used in this article, we combine (5) and (6). We calculate the fuzzy measures using the following formula:

$$\mu(P) = \sum_{i=1}^p \frac{|B_i \cap P|}{|P|} m(B_i). \quad (7)$$

The only quantities that we need to know to calculate  $\mu(P)$  using (7) are the values of  $m(B_i)$ . However, before demonstrating how we can calculate  $m(B_i)$ , we use some examples to illustrate the proposed Theorem.

### A. Examples to Illustrate the Proposed Theorem

Let  $X = \{1, 2, 3\}$ ,  $P = \{1, 2\}$ ,  $Q = \{1, 2\}$ , and  $W = \{1\}$ .  $B_i$ s can assume values  $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$ . Let  $m(\{1\}) = 0.1$ ,  $m(\{2\}) = 0.2$ ,  $m(\{3\}) = 0.05$ ,  $m(\{1, 2\}) = 0.35$ ,  $m(\{1, 3\}) = 0.03$ ,  $m(\{2, 3\}) = 0.07$ , and  $m(\{1, 2, 3\}) = 0.2$ .

1) *Example 1:* Suppose  $w_{B_i}^\phi \neq 0$  for  $B_i = \{1\}$  (say); then,  $\mu(\phi) = 0.1$  and is not equal to 0.

Similarly,  $w_{B_i}^X = 0.5$  for  $B_i = \{1\}$ ; then,  $\mu(X) = 0.95$ .

Now, suppose for  $B_i = \{2\}$ ,  $w_{B_i}^W > w_{B_i}^Q$ , while for all other  $B_i - s$ , we have  $w_{B_i}^W = w_{B_i}^Q$ ; then, it is clear that  $\mu(W) > \mu(Q)$ , which should not be the case. Hence, the weights must be fuzzy measures, as shown in the theorem.

2) *Example 2:* Let us consider the subset  $W = \{1\}$ . Now, subsets for which  $B_i \cap P = \phi$  are  $\{2\}, \{3\}, \{2, 3\}$ , while the rest belong to  $B_i \cap P \neq \phi$ . Let  $w_{\{2\}}^W = 0.8$ ,  $w_{\{3\}}^W = 0.6$ ,  $w_{\{2,3\}}^W = 0.7$ ,  $w_{\{1\}}^W = 0.75$ ,  $w_{\{1,2\}}^W = 0.8$ ,  $w_{\{1,3\}}^W = 0.8$ , and  $w_{\{1,2,3\}}^W = 0.8$ . It can be seen very clearly that condition 2 is violated using these weights as the right-hand side is equal to 0.081, while the left-hand side is equal to 0.539. It can be seen that as a result, the condition  $\mu(P) \leq \sum_{B_i \cap P = \phi} m(B_i)$  is violated as  $\mu(P) = 0.778$ , while  $\sum_{B_i \cap P = \phi} m(B_i) = 0.32$ .

Similarly, we can devise an example to show that condition 3, if violated, would not generate the fuzzy measure using the mass functions as the generated quantity would violate the primary rules of fuzzy and measure theory.

$m(B_i)$  conveys the extent to which the relevant information about the true class is conveyed by the subset of classifiers  $B_i$ . However, it is to be noted that  $m(B_i)$  does not specify which subset of  $B_i$  conveys the most relevant information in the determination of the true class label. Thus,  $m(B_i)$  gives us the information that the true class label matches the decision made by the subset of classifiers  $B_i$ . Thus, the fuzzy measure for the subset of classifiers  $P$  (i.e.,  $\mu(P)$ ) conveys the relevant information conveyed by the subset of classifiers  $P$  and is expressed as the weighted sum of  $m(B_i)$ s, which, in turn, provides the relevant information of the true class label conveyed by the subsets of classifiers  $B_i$ s.

We calculate the values of  $B_i$ s taking the following criteria in mind.

- 1) How well the subset  $B_i$  of classifiers can classify such that their combined decision matches the true label.
- 2) How well the classifiers in  $B_i$  combine with other classifiers such that the decision made by the resulting subset is the same as the true label.

Thus, we make  $m(B_i)$  dependent on not only the accuracy of the decision made by this subset of classifiers but also how compatible they are with the other classifiers being used.

To take into consideration the above two points while calculating the values of  $m(B_i)$ , we take the help of the rewarding system that we developed inspired by the MC method for calculating value function in an MDP.

Before explaining how we adopt a method that is close to the MC prediction method for MDPs in our work, we explain it in brief. In the MC prediction technique, the state-value function

is evaluated with respect to a specific policy. In this technique, given a specific policy, the value function of a specific state with respect to the policy is evaluated. Episodes are generated with respect to the specific policy, and the average return (expected return) of the specific state with respect to the policy is considered as the value function of that state. In our work, we adapt this MC prediction technique to calculate the values of the  $m(B_i)$ s.

It is important to note that in our work, we do not make a decision or try to evaluate a value function for a specific policy; rather, we evaluate the average of the value functions for a given number of different policies. For determining  $m(B_i)$ s, we take the average of several value functions calculated with respect to different policies that we define. We do this to get an intuition of how good is the state with respect to all the policies, all of which capture an important aspect of the subset of classifiers. The number of value functions whose average is to be taken depends on  $|B_i|$  and is equal to  $|X| - |B_i| + 1$ . The policies with respect to which these value functions would be calculated are defined as follows. The first policy would be to classify based on the information conveyed by the classifiers included in  $B_i$ . The second policy would be to randomly add one of the classifiers from the remaining set of classifiers and then classify based on the information conveyed by the new set of classifiers, namely the classifiers present in  $B_i$  and the new classifier randomly added, while the third policy is to add two classifiers randomly and then use the cumulative information conveyed by the new set of classifiers. New policies are added till all the classifiers present in the set  $X$  of classifiers are taken into consideration for the decision-making process. This makes the number of policies equal to  $|X| - |B_i| + 1$ , as mentioned above.

Each of these policies generates a value function based on the reward structure. Before we define the reward, we explain why this average of these value functions gives us the value of  $m(B_i)$ . The value function generated from the policy where no additional classifier is included in the subset and the decision is taken solely based on the decisions of the classifiers in the subset  $B_i$  indicates how accurate the classifiers in the subset  $B_i$  are. The value functions for all the other policies where we randomly add one or more classifiers that are not part of  $B_i$  and then make the decision indicate how compatible the classifier subset  $B_i$  is with other classifiers in the set of classifiers,  $X$ .

We now define the actions and the states. There are two actions, namely, making a classification and adding a classifier randomly. Let us assume that there are five classifiers. Then, there are 32 different states, consisting of the different combinations of classifiers. We define the rewards for the MDP as follows: +1 for a correct decision, -1 for a wrong decision, and 0 for adding other classifiers randomly.

Just like in MC prediction for solving the MDP, we divide the entire task into episodes, where the classification of each item (or the decision made by the classifier subset per item in the training set) is considered as one episode. Then, we take the average of the returns (expected returns) in each episode to evaluate the state-value functions for different policies.  $m(B_i)$ s are equal to the average of state-value functions with respect to the different policies, as mentioned above.

TABLE I  
EXPERIMENTAL RESULTS FOR FOUR-CLASS BREAST CANCER CLASSIFICATION  
PROBLEM, WHEN EVALUATED ON THE BACH DATASET

Model	VA	TA	P (avg)	R (avg)
VGG16	97	86	0.87	0.86
VGG19	98	83	0.84	0.83
Xception	99	91	0.91	0.91
Inception V3	99	90	0.90	0.90
InceptionResnet V2	99	91	0.91	0.91
<b>DS-based fuzzy ensemble</b>	<b>99.4</b>	<b>95</b>	<b>0.95</b>	<b>0.95</b>

VA stands for Validation Accuracy (%), TA stands for Test Accuracy (%), P stands for Precision and R stands for Recall.

Furthermore, this problem holds the Markov property because, in this problem, at any particular instant, the present state summarizes the entire history. This is because, at any state, the decision is taken based on the classifiers present in that state, and it does not depend on whether that classifier was originally a part of  $B_i$  or even on when that classifier was added. Thus, the decision is influenced by only the present state and not by history.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

##### A. Dataset Description

In the following section, we have discussed three different kinds of publicly available image datasets that have been used for the performance evaluation of the ensemble learning methodology proposed in this article.

- 1) *ICIAr 2018 Grand Challenge on Breast Cancer Histology (BACH) images*: The ICIAr 2018 Grand Challenge [18] on BACH dataset images is a large public dataset, composed of 400 microscopy images and 30 whole-slide annotated images of hematoxylin and eosin (H&E) stained breast lesions. The BACH dataset, which can be freely accessed at [19], is publicly available under the CC BY-NC-ND license. Our research study is conducted on Part A of the dataset, which consists of breast histology microscopy images categorized into the following four classes: normal, benign, *in situ* carcinoma, and invasive carcinoma, with 100 images present in each class. Each image, of resolution  $2048 \times 1536$  pixels, is in the RGB color mode. Each microscopy image, stored in the Tagged Image File Format file extension, has a magnification factor of 200 and pixel scale of  $0.42 \times 0.42 \mu\text{m}$ .
- 2) *Novel COVID-19 Chestxray Repository*: It is a public CXR image repository, available on Kaggle [20], comprising CXR images of frontal and lateral radiographic views. The dataset consists of three classes, namely, COVID-19, Pneumonia, and Normal (healthy) classes, with a total of 752, 1584, and 1639 CXR images, respectively. The CXR images are in the RGB color mode and each image has a resolution of  $512 \times 512$  pixels.
- 3) *UTD Multimodal Human Action Dataset (UTD-MHAD)*: The UTD-MHAD [21] is a large public human action recognition dataset consisting of four data modalities namely RGB videos, and depth videos, skeleton sequences, and inertial signals captured with the aid of

Microsoft Kinect camera and a wearable inertial sensor. It is a comprehensive and well-annotated dataset consisting of 27 human action classes performed by eight subjects, and each action is repeated four times by each subject. The dataset is publicly available at [22].

##### B. Experimental Setup

To implement the proposed method, we have used Python programming language with the aid of the Keras package with Tensorflow as the deep learning framework on Nvidia GeForce GTX 1080 Ti GPU with 11-GB RAM, 3584 CUDA cores, and a memory clock of 5505 MHz.

##### C. Results

- 1) *BACH dataset*: Under image preprocessing protocols, the H&E stained breast microscopy images are downsized to a resolution of  $512 \times 512$  pixels using bicubic interpolation, followed by the application of Macenko stain normalization [23]. Five fine-tuned deep convolutional neural network (DCNN) architectures, pretrained on the ImageNet dataset, namely VGG16, VGG19, Xception, InceptionV3, and InceptionResnetV2, are used for the extraction of deep features from specific convolutional layers after performing a thorough analysis of the activation maps of each layer in the DCNN model. These extracted features are then fused to generate deep image descriptors. These image descriptors are further fed into a multilayered perceptron (MLP) classifier, which yields the four-class classification performance. A confidence matrix is generated, by combining the confidence scores obtained across the five DCNN models used, which is then passed through the proposed DS-based fuzzy ensemble strategy for decision score aggregation. In the four-class breast cancer classification, the ensemble method is used to combine the confidence scores of each classifier for each of the four classes, and the final combined score is computed. The results are depicted in Table I. It can be observed that the accuracy after the ensemble is superior compared to the accuracy of the individual classifiers both in the case of test and validation sets. In fact, in the case of the test set, the accuracy of the proposed DS-based fuzzy ensemble model outperforms the best accuracy of the individual classifier by 4%.
- 2) *Novel COVID-19 Chestxray Repository*: Essential data preprocessing protocols, such as image resizing, color space translation, image denoising filters, and division of the CXR dataset, are applied to the CXR radiographs. Offline image augmentation techniques are applied to the training and validation sets. The CXR images are then fed as input into a DCNN architecture, pretrained on the ImageNet dataset. VGG16, Xception, and InceptionV3 are employed as transfer learning DCNN models, which are then fine-tuned. The global average pooling function is applied to the convolution layers of the DCNN model, having salient activation maps, which reduces the feature maps into 1-D vectors. These feature vectors are then

TABLE II  
EXPERIMENTAL RESULTS OBTAINED FOR THREE-CLASS CXR CLASSIFICATION  
PROBLEM, WHEN EVALUATED ON THE NOVEL COVID-19 CHESTXRAY  
REPOSITORY DATASET

Model	VA	TA	P (Avg)	R (Avg)	AUC
VGG16	96.71	91.22	0.92	0.92	0.92
Xception	97.02	92.98	0.93	0.93	0.92
Inception V3	97.49	93.48	0.94	0.94	0.94
<b>DS-based fuzzy ensemble</b>	<b>99</b>	<b>95.8</b>	<b>0.96</b>	<b>0.96</b>	<b>0.96</b>

VA stands for Validation Accuracy (%), TA stands for Test Accuracy (%), P stands for Precision, and R stands for Recall.

concatenated into a single feature embedding. This feature embedding is then passed through an MLP network, and then, the entire architecture is trained for 100 epochs for obtaining a refined deep feature embedding. As shown in (8), we have used a combination of cross-entropy loss ( $\mathcal{L}_{CE}$ ) and contrastive loss ( $\mathcal{L}_{con}$ ) as our overall training objective. This loss function facilitates the regularization of the deep feature embedding and helps in addressing the issue of domain gap while dealing with heterogeneous CXR data sources, which hence results in robust joint training

$$\mathcal{L}_{overall} = \mathcal{L}_{CE} + \alpha \mathcal{L}_{con} \quad (8)$$

where  $\alpha$  is a hyperparameter.

The refined feature embedding is then fed into a classifier with a 512-neuron fully connected layer, with an activation of the rectified linear unit, and a dropout rate of 0.5 to curtail the problem of overfitting. For addressing the three-class CXR classification problem under consideration, there is finally a three-neuron output layer with softmax activation. The three classifiers, corresponding to the three DCNN models used, generate confidence scores of validation and unseen test CXR images, which are combined to form confidence matrices of validation and test sets, respectively. The DS-based fuzzy ensemble method proposed here is used to combine the confidence scores of each classifier. We have tabulated the results in Table II. In this dataset also, we can observe that the accuracy of the proposed ensemble method outperforms the accuracy of all the individual classifiers. Even in the worst-case scenario, the fusion method outperforms the test accuracy of the individual classifier by 2.32%.

3) *UTD-MHAD*: We have extracted three geometric feature representations from the raw skeleton joint sequences provided in the skeleton data folder of the *UTD-MHAD*. These geometric features, namely pairwise joint distance motion ( $JJd_{motion}$ ), pairwise line angle ( $LLa$ ), and pairwise planar angle ( $PPa$ ), are used for capturing various spatiotemporal dynamic information during human actions.

These distilled feature representations are then converted into RGB image encodings, which are fed into DCNN classifiers. We have implemented our proposed DS-based fuzzy ensemble for the aggregation of the decision scores generated by DCNN classifiers. Fig. 1 shows the RGB image encodings for action classes A9 (Draw a circle

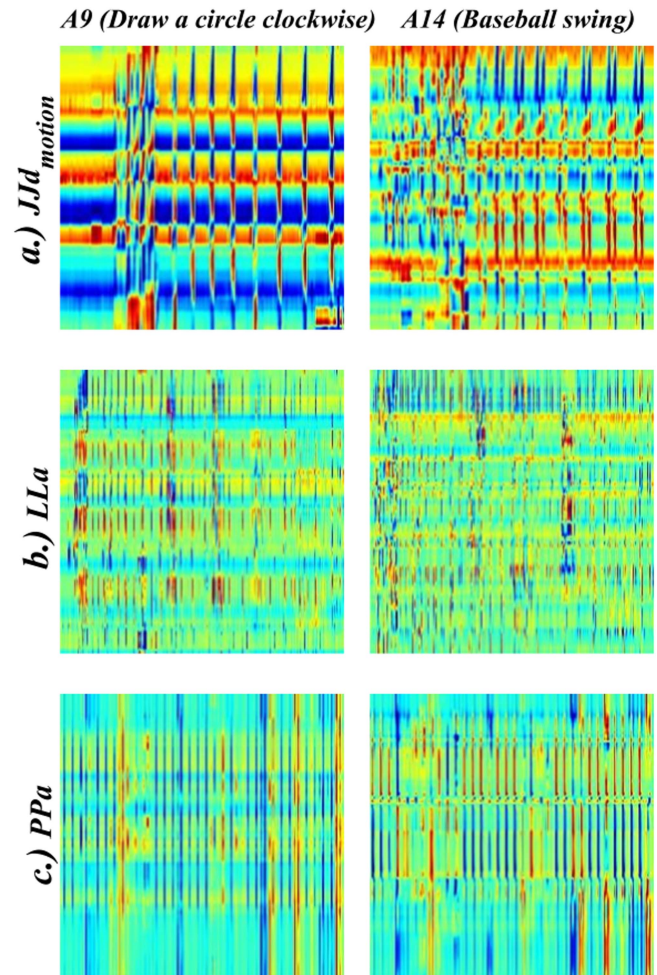


Fig. 1. First and second columns of the figure depict the RGB image encodings obtained for action classes A9 (Draw a circle clockwise) and A14 (Baseball swing), respectively, of the *UTD-MHAD*. Each row corresponds to the type of spatiotemporal feature extracted.

TABLE III  
EXPERIMENTAL RESULTS ON THE TEST IMAGE ENCODINGS OBTAINED AFTER  
THE EXTRACTION OF SPATIOTEMPORAL GEOMETRIC FEATURES FROM  
SKELETON SEQUENCES OF *UTD-MHAD*

Feature used/En	TA	P (Average)	R (Average)
$JJd_{motion}$	91.02	0.95	0.95
$LLa$	93.50	0.93	0.92
$PPa$	92.56	0.93	0.92
<b>DS-based fuzzy ensemble</b>	<b>95.6</b>	<b>0.95</b>	<b>0.95</b>

En stands for Ensemble, TA stands for Test Accuracy (%), P stands for Precision, and R stands for Recall.

clockwise) and A14 (Baseball swing), respectively, of the *UTD-MHAD*, obtained corresponding to each type of feature extraction methods employed. The experimental results are depicted in Table III. It can be observed from the table that the proposed fusion method outperforms the individual classifiers in terms of all the metrics considered, bolstering our claim that the fusion method improves the overall accuracy of the model. A detailed mathematical description of the spatiotemporal feature extraction techniques is given in the Appendix.

TABLE IV  
PERFORMANCE OF THE CHOQUET-INTEGRAL-BASED METHOD ON BACH DATASET USING THE FUZZY MEASURES

Overlap Index or the Power Capacity	Description of overlap Index or the Power Capacity	Accuracy (%)
Power Capacity	$m(A) = \frac{ A }{n}$	94
Overlap index for calculating capacities	$O(A, B) = \max_{i=1}^n \min(A(i), B(i))$	95
	$O(A, B) = 1/n \sum_{i=1}^n \min(A(i), B(i))$	95
	$O(A, B) = 1/n \sum_{i=1}^n \sqrt{A(i)B(i)}$	96
Proposed fuzzy measures		95

TABLE V  
COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON THE BACH DATASET

Work Ref.	Method	Number of classes	Image level accuracy (%)
[30]	DCNN	4	76
[31]	SVM with quadratic kernels	4	79.2
[32]	DCNN	4	83
[33]	DCNN	4	85
[34]	DCNN	4	86
[35]	Convolutional Capsule Network	4	87
[36]	DCNN + Gradient Boosted Trees	4	87.2
[37]	DCNN	4	87.5
[38]	DCNN + Network Fusion	4	88.5
[29]	DCNN + MV	4	90
[28]	EMS-Net	4	91.75
[39]	DCNN	4	92.25
[27]	Guided Soft attention network	4	93
[26]	DenseNet-161	4	94
[25]	Deep Attention	4	96
Proposed	DCNN + DS-based fuzzy ensemble	4	95

#### D. Comparison With Other Fuzzy Measures Derived From Overlap Functions

For comparing the performance of the proposed fuzzy measure with other fuzzy measures, we have implemented the capacities or the fuzzy measures used in the paper [24]. We have used the capacities for evaluating their efficacy on the BACH dataset. We have followed the rules of calculation mentioned in the paper. The results have been shown in Table IV.

#### E. Comparison With Other Methods

1) *BACH Dataset*: We have compared our method with the state-of-the-art methods using the four-class breast cancer classification dataset. Our proposed method has outperformed most of the existing methods in terms of accuracy evaluated on the test set, as demonstrated in Table V.

Here, we briefly explain the existing breast histology image classification methods, which we consider for comparison and which were applied to the said dataset. Fan et al. [25] employed a deep attention network and trained it with a batch size of 16, an stochastic gradient descent (SGD) optimizer, a learning rate of 0.1, and recorded the results up to 50 epochs. DenseNet-161 was utilized with considerable offline data augmentation by Kohl et al. [26]. A data distribution of 80% and 20% for training and validation, respectively, was used. Yang et al. [27] employed a guided attention network with extensive augmentation in their study accompanied by a fourfold cross-validation method. Yang et al. [28] incorporated an ensemble of nine fine-tuned pretrained transfer learning models (architectural variants of DenseNet and ResNet) for the four-class classification problem. Roy et al. [29] used the DCNN along with majority voting (MV), patch extraction, and data augmentation. They divided the dataset into training, validation, and testing sets. 40 of the 400 images were used as the test set for image classification, while the patch was extracted from the rest of the 360 images, 20% of which were

TABLE VI  
COMPARISON OF THE PROPOSED METHOD WITH OTHER METHODS ON THE NOVEL COVID-19 CHESTXRAY REPOSITORY

Work Ref.	Method	Test Accuracy (%)
[40]	DCNN + Choquet Integral	95.49
Proposed	DCNN + DS-based fuzzy ensemble	95.8

TABLE VII  
COMPARISON OF THE PROPOSED METHOD WITH THE EXISTING METHODS ON THE UTD-MHAD USING CROSS-SUBJECT PROTOCOL

Work Ref.	Method	Accuracy (%)	Kinect	RGB	Inertial
JTM [41]	Joint trajectory Map	85.81	✓	×	×
JDM [46]	Joint distance map	88.10	✓	×	×
HDM-BG [47]	Bayesian Hierarchical Dynamic Model	92.80	✓	×	×
Banerjee et al. [48]	DCNN+ fuzzy-integral-based fusion	97.91	✓	×	×
PoseMap [42]	Pose Estimation Maps	94.51	✓	✓	×
Action Machine [43]	13D + 2-D DCNN	92.50	✓	✓	×
STAR-Net [49]	Spatio-Temporal Activation Reprojection	90.00	×	✓	×
Gimme Signals [44]	Multivariate signal encoding	93.33	✓	×	✓
HAMLET [45]	Multimodal self-attention	95.12	✓	✓	✓
Proposed	DCNN + DS-based fuzzy ensemble	95.6	✓	×	×

used for testing patchwise accuracy, 10% for validation, while the rest 70% for patchwise training. As can be observed from Table VI, the proposed method beats almost all the existing methods barring [25], which used a complex attention-based module.

2) *Novel COVID-19 Chestxray Repository*: We have evaluated our proposed DS-based fuzzy ensemble approach on the Novel COVID-19 Chestxray Repository proposed in [40]. We have compared our method with the existing techniques. The results have been tabulated in Table VI. Bhowal et al. [40] developed a method of calculating the fuzzy measures of individual classifiers using coalition game theory and information theory before using the lambda fuzzy approximation for calculating the fuzzy measures of the set of classifiers. They also incorporated three different weighting schemes for calculating the Shapley values. Our proposed DS-based fuzzy ensemble yields leverage of 0.31% for the three-class CXR classification.

3) *UTD-MHAD*: In Table VII, we have tabulated the comparison of our proposed DS-based fuzzy ensemble with the existing approaches evaluated on the UTD-MHAD. Since the joint trajectory map technique, employed by Wang et al. [41], relies majorly on the temporal dynamics, it hence fails to capture discriminative spatial features essential for distinguishing actions such as classes A22 (*Jog*) and A23 (*Walk*). By only utilizing the Kinect modality, our method outperforms existing methods [42], [43], [44], [45], which incorporated a multimodal approach for harnessing salient information to facilitate multimodal feature learning. As illustrated in Table VII, the proposed method beats most of the existing methods, thereby justifying that the proposed method is quite general and works on a variety of deep learning and machine learning models on a variety of datasets.

#### F. Error Analysis

Although the proposed method performs well on varied datasets, we can state that there are two main limitations of the proposed fusion technique, which are listed as follows.

- 1) Besides the MC method that we used for the calculation of  $m(B_i)$ s, other optimization techniques may be used,



each of which may produce a separate set of basic assignment functions and a separate set of fuzzy measures as a result. Which algorithm to use has to be decided by the programmer using the code.

- 2) The second major problem is that even though we find the general formula for a set of fuzzy measures that can be associated with the DS theory and can be expressed in the form  $\mu(P) = \sum_{i=1}^p w_{B_i}^P m(B_i)$ , there are infinite sets of fuzzy measures that can be associated with the DS belief structure, and it cannot be expressed in this form. Thus, it is not feasible to consider all the possible sets of fuzzy measures associated with the DS belief structure and select the best possible one.

## V. CONCLUSION AND FUTURE DIRECTIONS

In this article, our contribution is twofold. First, we provide a specific generalized set of fuzzy measures associated with both the fuzzy measure theory and the DS theory. We also prove a theorem to provide us with the conditions such that the fuzzy measures can be generated using a predecided weighting scheme and the basic assignment functions of the DS theory. We further use a method inspired by the MC method for the calculation of the value function to calculate the basic assignment functions of the DS theory.

To prove the efficacy of the fuzzy measures derived here, we employ the fuzzy measures to the Choquet integral in order to fuse the decisions of deep learning models experimented on three varied kinds of datasets, namely, the BACH, Novel COVID-19 Chestxray Repository, and UTD-MHAD.

However, the method has two specific limitations. First, the values of  $m(B_i)$ s can be calculated using a number of different algorithms besides ours, which must be decided by the user. Though we evaluate the generalized formula for the set of fuzzy measures associated with the DS theory, which can be expressed as  $\mu(P) = \sum_{i=1}^p w_{B_i}^P m(B_i)$ , there are infinite sets of fuzzy measures that cannot be expressed in this form but can be associated with the DS belief structure. Hence, it may happen that the best possible set, giving the best result in a dataset, may not be of the form that we have derived. In our future work, we would try to make our theorem more generalized, so that we may represent more fuzzy measures associated with the DS theory using our theorem. We will evaluate our method with different algorithms and datasets to assess which algorithm performs best on a vast set of datasets.

## APPENDIX

### SPATIOTEMPORAL GEOMETRIC FEATURE EXTRACTION TECHNIQUES FOR 3-D HUMAN SKELETON ACTION RECOGNITION

Consider a video portraying a particular human activity or movement performed by a subject. This video comprises of  $F$  number of frames, each frame consisting of  $J$  number of key joints. Each joint  $i$ , where  $i \in J$ , is a position vector  $p_i = \{x_i, y_i, z_i\}$ . For the  $f$ th frame, where  $f \in F$ , the position vector across all the joints is given as  $P_f = \{p_1, p_2, p_3, \dots, p_{J-1}, p_J\}$ , and across the  $F$  frames, we thus

have a sequence  $P = \{P_1, P_2, P_3, \dots, P_{F-1}, P_F\}$ . A detailed description of the spatiotemporal feature extraction techniques is given as follows.

- 1) *Pairwise joint distance motion (JJd<sub>motion</sub>)*: In every frame, for  $i, j \in J, i \neq j$ , the pairwise joint distance metric of  $p_i$  and  $p_j$  is equal to the Euclidean distance between them and is given as

$$JJd_{ij}^f = \|p_i^f - p_j^f\|_2. \quad (9)$$

In the UTD-MHAD, for the  $f$ th frame having  $J = 20$  key joints,  $\binom{20}{2} = 190$  pairwise joint distance features are extracted. Across  $F$  frames of the entire video sequence, the joint–joint distance features are given as

$$JJd = \{JJd^0, JJd^1, JJd^2, \dots, JJd^{F-1}, JJd^F\}. \quad (10)$$

$JJd_{\text{motion}}$  is a feature representation that captures the kinematic motion patterns of skeleton joint sequences over time. It is computed by finding the temporal difference over the extracted pairwise joint distance features. For a given frame interval  $N$  and for  $f \in F, f < F - N$ , the joint–joint distance motion feature sequence ( $DM$ ) is given as

$$DM^f = JJd^f - JJd^{f+N}. \quad (11)$$

Across the entire video, we obtain the following feature sequence:

$$JJd_{\text{motion}} = \{DM^0, DM^1, \dots, DM^{F-N-1}, DM^{F-N}\}. \quad (12)$$

Here, we have taken frame interval ( $N$ ) equal to 10.

- 2) *Pairwise line angle (LLa)*: Pairwise joint vector ( $JJv$ ) is given as the difference between the two position vectors  $p_i$  and  $p_j$ , where  $i, j \in J, i \neq j$ :

$$JJv_{ij}^f = p_i^f - p_j^f. \quad (13)$$

Pairwise joint orientation ( $JJo$ ) is given as

$$JJo_{ij}^f = JJv_{ij}^f / JJd_{ij}^f. \quad (14)$$

Pairwise line angle ( $LLa$ ) is defined as the angle, in radians ( $0 < \theta < \pi$ ) between lines  $L_{J_i \rightarrow J_j}$  and  $L_{J_p \rightarrow J_q}$ . The change in the features in the vector sequence is swifter than the distance features, which facilitates in capturing discriminating spatial information corresponding to the angular movements, such as Draw circle, Baseball swing, Wave, Throw, Basketball shoot, and so on, performed by the subjects

$$LLa_{ijpq}^f = \arccos \left( JJo_{ij}^f \odot JJo_{pq}^f \right) \quad (15)$$

where  $i, j, p, q \in J$  are the joint indices, where  $(i, j) \neq (p, q)$  in (15). In the UTD-MHAD, for the  $f$ th frame having 34 lines in total,  $\binom{34}{2} = 561$  pairwise joint distance features are extracted. Across  $F$  frames of the entire video sequence, the line–line angle features are given as

$$LLa = \{LLa^0, LLa^1, LLa^2, \dots, LLa^{F-1}, LLa^F\}. \quad (16)$$

- 3) *Pairwise planar angle (PPa)*: Pairwise planar angle (PPa) is defined as the angle, in radians ( $0 < \theta < \pi$ ), between the normal vectors of planes  $P_{J_i \rightarrow J_j}$  and  $L_{J_p \rightarrow J_q}$

$$PPa_{ij, k, p, q, r}^f = \arccos \left( JJo_{ij}^f \otimes JJo_{ik}^f \odot JJo_{pq}^f \otimes JJo_{pr}^f \right) \quad (17)$$

where  $i, j, k, p, q, r \in J$  are the joint indices, where  $(i, j, k) \neq (p, q, r)$  in (17). In the UTD-MHAD, for the  $f$ th frame having five planes in total,  $\binom{5}{2} = 10$  pairwise planar angle features are extracted. Across  $F$  frames of the entire video sequence, the line–line angle features are given as

$$PPa = \{PPa^0, PPa^1, PPa^2, \dots, PPa^{F-1}, PPa^F\}. \quad (18)$$

The obtained feature sequence is then converted to its corresponding three-channel RGB image encoding by using the encoding scheme, as shown in (19), proposed by Du et al. [50]

$$p = \text{floor} \left( \frac{p - c_{\min}}{c_{\max} - c_{\min}} * (255 - 0) \right) \quad (19)$$

where  $p$  indicates the pixel intensity and  $c_{\max}$  and  $c_{\min}$  are the maximum and minimum of all the joint coordinates, respectively. The obtained image encoding is then resized to a dimension of  $256 \times 256$  pixels using bicubic interpolation. As illustrated in Fig. 1, the spatial structure and the temporal dynamics are depicted along the width and height of each image encoding, respectively. The image encodings are then passed through the Alexnet model, trained from scratch, to generate decision scores for each feature vector.

## REFERENCES

- [1] G. Choquet, "Theory of capacities," *Ann. l'institut Fourier*, vol. 5, pp. 131–295, 1954.
- [2] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Tokyo Inst. Technol., Tokyo, Japan, 1974.
- [3] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
- [4] C. Li, G. Zeng-tai, and D. Gang, "Genetic algorithm optimization for determining fuzzy measures from fuzzy data," *J. Appl. Math.*, vol. 2013, pp. 1–11, 2013.
- [5] G. Beliakov and J.-Z. Wu, "Learning fuzzy measures from data: Simplifications and optimisation strategies," *Inf. Sci.*, vol. 494, pp. 100–113, 2019.
- [6] A. F. Tehrani, W. Cheng, K. Dembczyński, and E. Hüllermeier, "Learning monotone nonlinear models using the Choquet integral," *Mach. Learn.*, vol. 89, no. 1/2, pp. 183–211, 2012.
- [7] J. Murillo, S. Guillaume, E. Tapia, and P. Bulacio, "Revised HLMS: A useful algorithm for fuzzy measure identification," *Inf. Fusion*, vol. 14, no. 4, pp. 532–540, 2013.
- [8] R. Mesiar, "Generalizations of k-order additive discrete fuzzy measures," *Fuzzy Sets Syst.*, vol. 102, no. 3, pp. 423–428, 1999.
- [9] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 1995, vol. 1, pp. 145–150.
- [10] S.-T. Liu and C. Kao, "Fuzzy measures for correlation coefficient of fuzzy numbers," *Fuzzy Sets Syst.*, vol. 128, no. 2, pp. 267–275, 2002.
- [11] M. Beynon, B. Curry, and P. Morgan, "The Dempster–Shafer theory of evidence: An alternative approach to multicriteria decision modelling," *Omega*, vol. 28, no. 1, pp. 37–50, 2000.
- [12] Y. Yang, D. Han, and C. Han, "Discounted combination of unreliable evidence using degree of disagreement," *Int. J. Approx. Reason.*, vol. 54, no. 8, pp. 1197–1216, 2013.
- [13] R. R. Yager, "A class of fuzzy measures generated from a Dempster–Shafer belief structure," *Int. J. Intell. Syst.*, vol. 14, no. 12, pp. 1239–1247, 1999.
- [14] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proc. 11th Int. Conf. Inf. Fusion*, 2008, pp. 1–8.
- [15] J. Wang, Y. Hu, F. Xiao, X. Deng, and Y. Deng, "A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster–Shafer theory of evidence: An application in medical diagnosis," *Artif. Intell. Med.*, vol. 69, pp. 1–11, 2016.
- [16] C. Yu, J. Yang, D. Yang, X. Ma, and H. Min, "An improved conflict-evidence combination approach based on a new supporting probability distance," *Expert Syst. Appl.*, vol. 42, no. 12, pp. 5139–5149, 2015.
- [17] Y. Zhao, R. Jia, and P. Shi, "A novel combination method for conflict-evidence based on inconsistent measurements," *Inf. Sci.*, vol. 367, pp. 125–142, 2016.
- [18] G. Aresta et al., "BACH: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, 2019.
- [19] G. Aresta, T. Araújo, and S. Kwok, "ICIAr 2018 grand challenge on breast cancer histology images." Accessed: Jan. 1, 2022. [Online]. Available: <https://iciar2018-challenge.grand-challenge.org/>
- [20] S. Sen, P. Bhowal, and R. Sarkar, "Novel COVID-19 Chestxray Repository." Accessed: Jan. 1, 2022. [Online]. Available: <https://www.kaggle.com/subhankarsen/novel-covid19-chestxray-database>
- [21] C. Chen, R. Jafari, and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 168–172.
- [22] C. Chen, R. Jafari, and N. Kehtarnavaz, "UT Dallas multimodal human action dataset (UTD MHAD)." Accessed: Jan. 1, 2022. [Online]. Available: <https://personal.utdallas.edu/kehtar/UTD-MHAD.html>
- [23] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro*, 2009, pp. 1107–1110.
- [24] D. Paternain, H. Bustince, M. Pagola, P. Sussner, A. Kolesárová, and R. Mesiar, "Capacities and overlap indexes with an application in fuzzy rule-based classification systems," *Fuzzy Sets Syst.*, vol. 305, pp. 70–94, 2016.
- [25] M. Fan, T. Chakraborti, I. Eric, C. Chang, Y. Xu, and J. Rittscher, "Fine-grained multi-instance classification in microscopy through deep attention," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, 2020, pp. 169–173.
- [26] M. Kohl, C. Walz, F. Ludwig, S. Braunewell, and M. Baust, "Assessment of breast cancer histology using densely connected convolutional networks," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 903–913.
- [27] H. Yang, J.-Y. Kim, H. Kim, and S. P. Adhikari, "Guided soft attention network for classification of breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1306–1315, May 2020.
- [28] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-NET: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images," *Neurocomputing*, vol. 366, pp. 46–53, 2019.
- [29] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for classification of breast histology images using deep learning," *Comput. Med. Imag. Graph.*, vol. 71, pp. 90–103, 2019.
- [30] C. A. Ferreira et al., "Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet V2," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 763–770.
- [31] A. Sarmiento and I. Fondón, "Automatic breast cancer grading of histological images based on colour and texture descriptors," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 887–894.
- [32] Y. Wang, L. Sun, K. Ma, and J. Fang, "Breast cancer microscope image classification based on CNN with image deformation," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 845–852.
- [33] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 837–844.
- [34] N. Brancati, M. Frucci, and D. Riccio, "Multi-classification of breast cancer histology images by using a fine-tuning strategy," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 771–778.
- [35] T. Iesmantas and R. Alzbutas, "Convolutional capsule network for classification of breast cancer histology images," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 853–860.
- [36] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 737–744.

- [37] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 914–922.
- [38] A. Mahbod, I. Ellinger, R. Ecker, Ö. Smedby, and C. Wang, "Breast cancer histological image classification using fine-tuned deep network fusion," in *Proc. Int. Conf. Image Anal. Recognit.*, 2018, pp. 754–762.
- [39] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Breast cancer diagnosis with transfer learning and global pooling," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2019, pp. 519–524.
- [40] P. Bhowal, S. Sen, J. H. Yoon, Z. W. Geem, and R. Sarkar, "Choquet integral and coalition game-based ensemble of deep learning models for COVID-19 screening from chest X-ray images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 12, pp. 4328–4339, Dec. 2021.
- [41] P. Wang, Z. Li, Y. Hou, and W. Li, "Action recognition based on joint trajectory maps using convolutional neural networks," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 102–106.
- [42] M. Liu and J. Yuan, "Recognizing human actions as the evolution of pose estimation maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1159–1168.
- [43] J. Zhu et al., "Action machine: Rethinking action recognition in trimmed videos," 2018, *arXiv:1812.05770*.
- [44] R. Memmesheimer, N. Theisen, and D. Paulus, "Gimme signals: Discriminative signal encoding for multimodal activity recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 10394–10401.
- [45] M. M. Islam and T. Iqbal, "HAMLET: A hierarchical multimodal attention-based human activity recognition algorithm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 10285–10292.
- [46] C. Li, Y. Hou, P. Wang, and W. Li, "Joint distance maps based action recognition with convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 624–628, May 2017.
- [47] R. Zhao, W. Xu, H. Su, and Q. Ji, "Bayesian hierarchical dynamic model for human action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7733–7742.
- [48] A. Banerjee, P. K. Singh, and R. Sarkar, "Fuzzy integral based CNN classifier fusion for 3D skeleton action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 6, pp. 2206–2216, Jun. 2021.
- [49] W. McNally, A. Wong, and J. McPhee, "Star-Net: Action recognition using spatio-temporal activation reprojection," in *Proc. 16th Conf. Comput. Robot Vis.*, 2019, pp. 49–56.
- [50] Y. Du, Y. Fu, and L. Wang, "Skeleton based action recognition with convolutional neural network," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit.*, 2015, pp. 579–583.