IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 5, 2020, accepted February 24, 2020, date of publication February 28, 2020, date of current version March 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977108

Fuzzy Logic System Application for Detecting SNP-SNP Interaction

CHENG-HONG YANG[1,2,3], (Senior Member, IEEE), LI-YEH CHUANG[4,5], AND YU-DA LIN[1], (Member, IEEE)

[1]Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[2]Ph. D. Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[3]Drug Development and Value Creation Research Center, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[4]Department of Chemical Engineering, I-Shou University, Kaohsiung 84001, Taiwan
[5]Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung 84001, Taiwan

Corresponding authors: Li-Yeh Chuang (chuang@isu.edu.tw) and Yu-Da Lin (yudalinemail@gmail.com)

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant 108-2221-E-992-031-MY3, Grant 108-2221-E-214-019-MY3, and Grant 108-2811-E-992-502.

ABSTRACT The identification of interactions between single-nucleotide polymorphisms (SNP–SNP interactions) is crucial for determining human genetic disease susceptibility. With rapid technological advancements, multiobjective multifactor dimensionality reduction (MOMDR) measurements have achieved high detection success rates. However, the classification of high- or low-risk groups is central to MOMDR and has yet to be extensively studied. To address limitations in binary classification, we propose an improved fuzzy sigmoid (FS) approach that uses membership degrees in MOMDR, thus denoting it as FSMOMDR. For determining the interval of membership, our improved FS approach assesses the distance between the i^{th} multifactor class and outcome (cases and controls). Thus, the improved FS approach enables MOMDR algorithms to determine the membership degrees of high- and low-risk groups in each multifactor class because the two-element set is extended to a specified membership interval. Moreover, the improved FS approach can handle uncertain information, which thus enables the effective detection of the *m*-locus combinations with similar distributions. FSMOMDR measurements can also distinguish similar frequencies among genotype combinations, thus enabling the detection of more significant SNP–SNP interactions. On the basis of the classification accuracy rate of MOMDR and results obtained from the analysis of several test data sets, we determined FSMOMDR to be superior to other MDR-based methods with respect to detection success rate. The results indicate that binary and fuzzy classifications involving MOMDR can provide insight into uncertainty in risk classification. Thus, FSMOMDR could successfully detect SNP–SNP interactions in coronary artery disease in a large data set obtained from the Wellcome Trust Case Control Consortium. We could successfully reduce uncertain information in MDR and thus suggest that membership based on the improved sigmoid function can be used to identify SNP–SNP interactions as well as obtain content knowledge.

INDEX TERMS Fuzzy set, SNP-SNP interaction, classification, case-control study.

I. INTRODUCTION

The genome-wide association study (GWAS) has been extensively used to detect associations between complex genes [1]; the approach aims to reveal factors associated with a particular disease. These factors include single nucleotide polymorphisms (SNPs) and other DNA-related factors. However, if researchers only use a single factor for disease identification in a GWAS, then they do not identify other factors that are significantly associated with a particular disease [2]. With respect to variability associations between complex genes, SNP–SNP interactions may explain the absence of inheritance [3]. SNP–SNP interaction is a major factor for identifying many genetic diseases [4], [5]. Consequently, SNP–SNP interaction detection has become important in multifactorial disease analysis [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal.

Effective calculation is a powerful tool for improving the recognition of SNP–SNP interactions in genetic association research [7]– [9].

Many methods have been proposed for the detection of SNP–SNP interactions. One such method is the Bayesian epistasis correlation map (BEAM) [10]. A BEAM introduces a Bayesian mark partitioning model and a Markov chain Monte Carlo sampling approach to maximize model posterior probability. Another method is AntEpiSeeker, which introduces a two-stage ant colony optimization to detect SNP–SNP interactions [11]. Similarly, Wan *et al.* introduced a Boolean Operation–based Screening and Testing (BOOST) approach for examining all pairwise interactions in genome-wide case–control studies [12]. Another method is SNPRuler, which introduces a branch and bound algorithm to determine the chi-square–based maximum rule utility metric for detecting SNP–SNP interaction [13]. Similarly, Ritchie *et al.*, introduced a multifactor dimensionality reduction (MDR) method, based on statistical evaluation, for detecting SNP–SNP interaction. MDR can characterize nonadditive interactions between discrete factors in case–control studies [14].

Unlike traditional statistical approaches, such as logistic regression, MDR uses nonparametric and genetic model data in case–control studies. MDR reduces the dimensionality of multifactor information by distinguishing genotype combinations into high- and low-risk groups. This process detects nonlinear or nonadditive interactions between the original variables. MDR uses the k-fold cross-validation (CV) approach to avoid overfitting in MDR-based predictions of disease status. Several MDR-based extension methods have been proposed [15], including MDR-ER [16], particle swarm optimization-based MDR (PBMDR) [17], class-based MDR (CMDR) [18], MOMDR [19], IMDR [20], and the empirical fuzzy MDR (EFMDR) [21]. Recently, MDR has had demonstrably superior implementation in the SNP–SNP-interaction detection of cardiovascular diseases [22], breast cancer [23], and facial emotion perceptions [24].

Let A be a classic binary set; its membership function yields outputs of only 1 or 0, depending on whether x belongs to A. Zadeh [25] proposed fuzzy set theory, representing a class of objects with continuous rank membership. A fuzzy set A in a universal space X is a set of ordered pairs $\{(x, \mu_A(x))|x \in X\}$, where $\mu_A(x)$ in [0, 1] represents the membership degree of x in the fuzzy set A. Accordingly, $\mu_A(x)$ is reduced to û, the indicator function $I_A(x)$ of set A. A classical set is thus considered a special fuzzy set where the indicator function is a membership function. This fuzzy logic extension of classical set theory has been used in various fields, including bioinformatics and medicine [21], [26]. In the detection of SNP–SNP interactions, classification into high- and low-risk groups is a problem of uncertainty. Accordingly, fuzzy logic is an approximation approach based on the representation of linguistic knowledge. It entails the use of fuzzy rules to address uncertainties [27]–[29]. When fuzzy logic is used, MDR can better distinguish high- and low-risk groups and thus increase detection success rates in SNP–SNP interaction detection processes. Both fuzzy set-based generalized MDR [30] and EFMDR [21] are fuzzy–based MDR approaches. Moreover, FGMDR detects SNP–SNP interaction using fuzzy set–based generalized linear models for improved covariate adjustment. When using FGMDR, selecting suitable parameters is difficult. Thus, EFMDR uses the MDR-based empirical fuzzy set that does not require the selection of suitable parameters. EFMDR is effective for identifying intergene SNP–SNP interactions in particular diseases. Moreover, a quicker version of EFMDR has been proposed [31]. Despite the recently increasing focus and investment of resources in MDR classification, studies on this topic have been limited.

This paper proposes an improved fuzzy logic system that is based on MOMDR to estimate the membership degree for the epistasis detection data set. The fuzzy sigmoid (FS) is favorable for SNP–SNP interaction detection because it allows local features to belong to multiple groups. Many studies have successfully applied the FS to improve algorithm performance [21], [32]–[34]. In particular, FSMOMDR has been used on coronary artery disease data sets. The results were obtained through simulation using a real big data set from the Wellcome Trust Case Control Consortium (WTCCC). The results demonstrated the superiority of FSMOMDR relative to other algorithms with respect to success detection rate.

The remainder of this paper is organized as follows. The relevant approach is summarized in Section II, where we define an MO function based on fuzzy membership degrees and present FSMOMDR. Experimental evaluations and result analyses are provided in Section III. In Section IV, we discuss the advantages of the FSMOMDR algorithm. Finally, Section V concludes this paper.

II. METHODS

A. MDR PROCESS

MDR detects SNP–SNP interactions by evaluating each m-locus combination using the distribution of cases and controls [14]. Specifically, the m-locus combination is such that an SNP–SNP interaction is represented by a set $\{s_1, \ldots, s_m | s \in \text{SNPs}, s_i \neq s_j\}$. Because each SNP contains the three genotypes, an m-locus combination has 3^m genotype combinations. In MDR, each genotype combination is called a multifactor class. A dimension reduction approach is introduced in MDR for converting a high dimension into a 2×2 confusion matrix in which the actual class contains cases and controls and the predicted class contains high- and low-risk groups. Subsequently, a k-fold CV operation generates k CV subsets. In each CV operation, a CV subset is used as a testing data set and other $k - 1$ CV subsets are combined to form a training data set. The purpose of the testing data set is to evaluate the trained model, which is trained using the corresponding training data set. An optimal trained model (denoted as an i-fold CV model where $i = 1, 2, \ldots, k$) is selected according the highest correct classification rate in each CV operation. Thus, the k-fold CV models can be obtained, and the CV

consistency (CVC) operation is used to count the occurrence frequency of a fold CV model among k-fold CV models. The model with the highest CVC is regarded as the best model in an MDR implementation. MDR comprises the follow steps: 1) perform the k-fold CV operation, 2.1) generate the training and testing data sets according to the k-fold CV operation, 2.2) generate all m-locus combinations, 3.1) assign cases and controls into multifactor classes, 3.2) calculate the ratio between cases and controls within each multifactor class of the m-locus combination, 3.3) classify all multifactor classes into a high-risk group and a low-risk group, 3.4) evaluate the m-locus combination using the correct classification rate (CCR), 3.5) select the best model with the highest CCR in each CV operation, and 4) perform CVC operation.

B. EFMDR PROCESS

MDR entails the application of binary classification to determine membership to high- or low-risk groups using the frequencies of multiple genotypes in cases and controls. Binary classification methods cannot address uncertainty, which results in the loss of key information [26]. Empirical fuzzy MDR (EFMDR) is an extension of MDR using the empirical fuzzy (EF) approach to address the limitations of binary classification [21]. MDR differs from EFMDR in Steps 3.2, 3.3, and 3.4 in the aforementioned MDR process. In EFMDR, the EF approach is used to evaluate the membership degrees of high- and low-risk groups within each multifactor class [denoted as H(w_H) and L(w_L), respectively] through Step 3.2; however, Step 3.3 of the MDR process is omitted from EFMDR. In Step 3.4, the CCR is evaluated on the basis of H(w_H) and L(w_L). EFMDR comprises the following steps: 1) perform a k-fold CV operation, 2.1) generate training and testing data sets according to the k-fold CV operation, 2.2) generate all m-locus combinations, 3.1) assign cases and controls to multifactor classes, 3.2) calculate the membership degrees of high-risk H(w_H) and low-risk L(w_L) groups within each multifactor class of the m-locus combination, 3.3) evaluate the m-locus combination using the CCR based on H(w_H) and L(w_L) (denoted as CCR_{fuzzy}), 3.4) select the best model with the highest CCR_{fuzzy} in each CV operation, and 4) perform a CVC operation.

C. MOMDR PROCESS

Using the Pareto set operation, MOMDR was introduced by Yang *et al.* in 2018 [19]. MDR-based methods use a single classification measure (usually the CCR) as an objective function to detect SNP–SNP interactions, whereas MOMDR uses multiple objective functions. MOMDR introduces the maximized multiobjective (MO) function as follows:

$$\text{maximize} \begin{cases} f_1 = LR \\ f_2 = CCR, \end{cases} \tag{1}$$

where functions f_1 and f_2 are the likelihood rate (LR) [35] and CCR [14] measures, respectively. In the Pareto set operation, if $f_i(x_1) \geq f_i(x_2)$ for all objective functions, then x_1 dominates another solution x_2. In the Pareto set X^*, others do not dominate each $x^* \in X^*$. The Pareto set and Pareto set filter operators record and determine nondominated SNP–SNP interactions. For a k-fold CV, the number k of Pareto sets (X^*) is generated in the evaluations of all m-locus combinations. Finally, optimal SNP–SNP interactions can be determined through the CVC operation. The MOMDR comprises the follow steps: 1) perform k-fold CV operation, 2.1) generate training and testing data sets according to the k-fold CV operation, 2.2) generate all m-locus combinations, 3.1) assign cases and controls to multifactor classes, 3.2) calculate the ratio between cases and controls within each multifactor class of the m-locus combination, 3.3) classify all multifactor classes into a high-risk group and a low-risk group, 3.4) evaluate the m-locus combination using the MO function, 3.5) perform the Pareto set operation in each CV operation, and 4) perform the CVC operation.

Algorithm 1 FSMOMDR Pseudo-Code

```
01: Divide data into K subsets
02: For k = 1 to K subsets
03:   assign k^th CV subset as the testing data and the other
      CV subsets as the training data
04: training data:
05:   While (stop when all m-locus combinations are
      evaluated)
06:   For i = 1to the number of multifactor class
07:         the membership (H(w_{i,H}) group and L(w_{i,L})
            group) of i^th multifactor class of m-locus
            combination is measured by improved FS
            approach
08:   End i
09:     compute the TP_f, FP_f, FN_f, and TN_f
10:     compute the LR_fuzzy
11:     compute the CCR_fuzzy
12:   End while
13:   Pareto set operation collects the optimal models
      according to the LR_fuzzy and CCR_fuzzy
14: End training data
15: testing data:
16:   compute the LR_fuzzy and CCR_fuzzy of the best m-locus
      combination using testing data
17: End testing data
18: End k
19: perform the CVC operation
```

D. FSMOMDR PROCESS

FSMOMDR extends the MOMDR by using an improved FS approach to extend binary classification into a fuzzy classification; it is based on membership degree in the MO measure. FSMOMDR is similar to EFMDR in that both are used to calculate the membership degree in the high-risk and low-risk groups within each multifactor class, H(w_H) and L(w_L). Thus, an MO function, based on the H(w_H) and L(w_L)

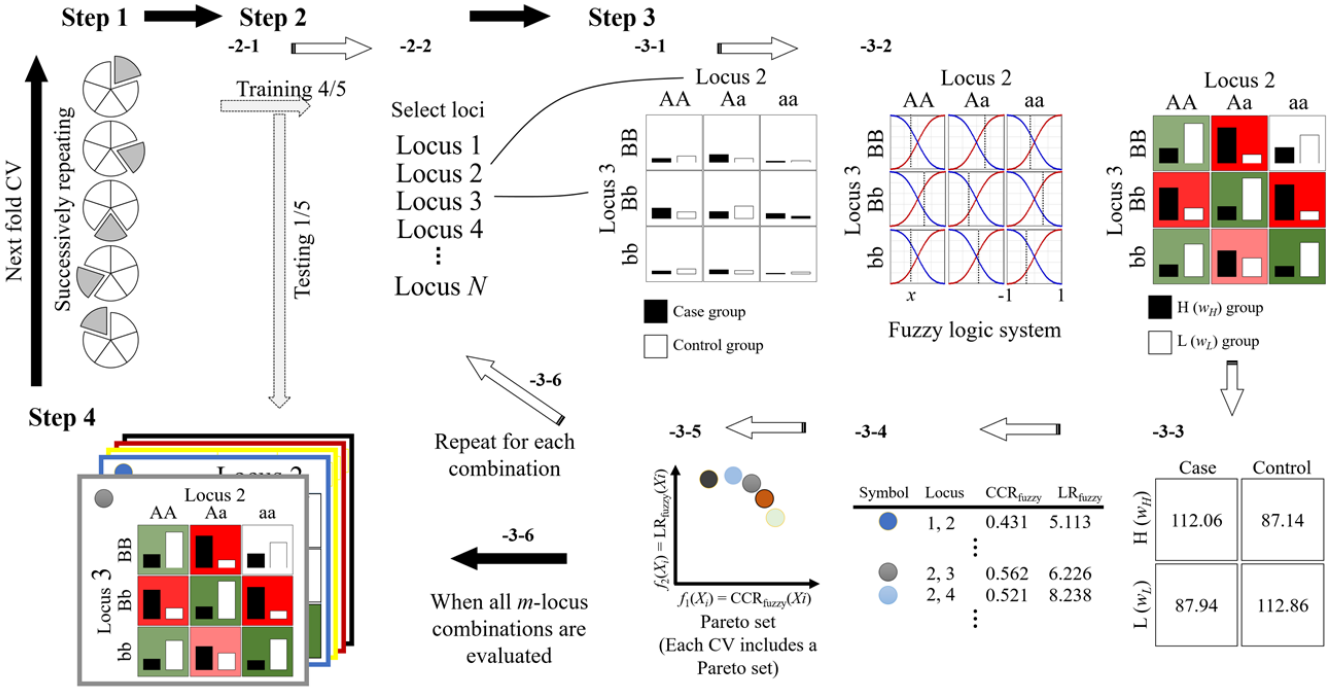

FIGURE 1. Diagram of FSMOMDR. All steps are detailed in the section on the FSMOMDR process.

membership degrees, can be formulated as follows.

$$\text{maximize} \begin{cases} f_1 = LR_{Fuzzy} \\ f_2 = CCR_{Fuzzy}, \end{cases} \tag{2}$$

where the functions f_1 and f_2 are fuzzy CCR and LR measures, respectively, that are both based on the membership degree. FSMOMDR is illustrated in Fig. 1; it comprises four steps (Algorithm 1) as follows.

Step 1: Perform the k-fold CV operation

1-1: Randomly sort the data set. All cases (samples for a given disease) and controls (samples for the normal population) are randomly shuffled.

1-2: Stratified random k-fold. The ratio between cases and controls is calculated, and k CV subsets comprising cases and controls are generated according to the ratio between cases and controls.

Step 2: Generate the training and testing data and all m-locus combinations

2-1: In each CV operation, a CV subset is used as testing data to evaluate the best model, and other CV subsets are combined to form the training data to determine the best model (defined by the m-locus combination having the highest value of measure).

2-2: Generate all m-locus combinations. Among all SNPs, all m-locus combinations are generated and assigned to a set.

Step 3: Evaluation of m-locus combination

3-1: The m-locus combination can generate 3^m multifactor classes according to all combinations of genotypes. Each multifactor class contains case and control groups. Each sample in the training data is assigned to a particular multifactor class. When a sample matches a given multifactor class, then it is assigned to the case group if it belongs to a case and to the control group otherwise.

3-2: The membership degree of the multifactor class is measured using the improved FS approach. In FSMOMDR, each sample can have a partial membership degree for both H(w_H) and L(w_L) groups. The case group to control group ratio within the i^{th} multifactor class is transformed into the interval $[-1, 1]$ by using (3).

$$x_i = \frac{2n_{i1}}{n_{i1} + n_{i0}} - 1, \tag{3}$$

where n_{i1} and n_{i0} are the sample frequencies matching the i^{th} multifactor class in the case group and the control group, respectively. An improved FS approach is introduced to evaluate the H(w_H) group and L(w_L) group within the i^{th} multifactor class and is formulated as follows:

$$H(w_{i,H}) = \begin{cases} 0, & \text{if } x_i = -1 \\ \dfrac{1}{1 + \left(\frac{x_i - 1}{x_i + 1}\right)^2}, & \text{other} \\ 1, & \text{if } x_i = 1 \end{cases} \tag{4}$$

$$L(w_{i,L}) = 1 - H(w_{i,H}) \tag{5}$$

3-3: A 2×2 contingency table is generated. Four fuzzy units of true positive (TP_f), false positive (FP_f), false negative (FN_f), and true negative (TN_f) are composed of both the membership degrees [H (w_H) and L (w_L)] and the sample frequencies of the i^{th} multifactor class (i.e., n_{i1} and n_{i0}).

In all multifactor classes, the TP_f value is the sum of H(w_H) with frequency n_1, and the FN_f value is the sum of L(w_L) with frequency n_1. Similarly, both the sum of H(w_H) with frequency n_0 and the sum of L(w_L) with frequency n_0 are the values of FP_f and TN_f, respectively. The formulae for TP_f, FP_f, FN_f, and TN_f are as follows:

$$\begin{aligned} TP_f &= \sum_i n_{i1} \times H(w_{i,H}) \\ FP_f &= \sum_i n_{i0} \times H(w_{i,H}) \\ FN_f &= \sum_i n_{i1} \times L(w_{i,L}) \\ TN_f &= \sum_i n_{i0} \times L(w_{i,L}) \end{aligned} \tag{6}$$

$$\begin{aligned} LR_{Fuzzy} &= 2 \sum \text{Observed} \log \left[\frac{\text{Observed}}{\text{Expected}} \right] \\ &= 2 \left[TP_f \times log \left(\frac{TP_f}{A^*} \right) + FP_f \times log \left(\frac{FP_f}{B^*} \right) \right. \\ &\quad \left. + FN_f \times log \left(\frac{FN_f}{C^*} \right) + TN_f \times log \left(\frac{TN_f}{D^*} \right) \right] \end{aligned}$$

$$s.t. \begin{cases} A^* = \dfrac{(TP_f + FN_f)(TP_f + FP_f)}{TP_f + FP_f + FN_f + TN_f} \\ B^* = \dfrac{(FP_f + TN_f)(TP_f + FP_f)}{TP_f + FP_f + FN_f + TN_f} \\ C^* = \dfrac{(TP_f + FN_f)(FN_f + TN_f)}{TP_f + FP_f + FN_f + TN_f} \\ D^* = \dfrac{(FP_f + TN_f)(FN_f + TN_f)}{TP_f + FP_f + FN_f + TN_f} \end{cases} \tag{7}$$

Consequently, the dimensions of 3^m multifactor classes are reduced into 2×2 dimensions by considering the membership degrees for the high-risk and low-risk groups.

3-4: The m-SNP combination is evaluated using the MO function, which is based on the H(w_H) and L(w_L) membership degrees. LR_{Fuzzy} and CCR_{fuzzy} are calculated on the basis of the two-way contingency table from step 3-3.

1) Objective function 1: LR_{Fuzzy} consists of observed frequencies in the 2×2 contingency table, including expected frequencies under the null hypothesis of no association [35]. LR_{Fuzzy} is formulated according to (7).

2) Objective function 2: CCR_{fuzzy} assesses the proportion of correctly classified individuals with an m-locus combination. CCR_{fuzzy} is formulated as per (8).

$$CCR_{fuzzy} = 0.5 \times \left(\frac{TP_f}{TP_f + FN_f} + \frac{TN_f}{FP_f + TN_f} \right) \tag{8}$$

3-5: Pareto set operation. The Pareto set operation determines candidates ($X_j^* = (x_1^*, \ldots, x_i^*)$) and adds each candidate into Pareto set j, where $j \in \{1, \ldots, k\}$ and k is the number of CVs. No candidate in the Pareto set dominates any another. Suppose an m-locus combination x_q is currently evaluated in the j^{th} CV and x_q is compared with all x^* in X_j^*; if x_q is not dominated by any x^*, then it is added to X_j^*. When x_p in X_j^* is dominated by x_q such that $f_1(x_q) \geq f_1(x_p)$ and $f_2(x_q) \geq f_2(x_p)$, then x_p is omitted from X_j^*.

3-6: Steps 3-1 to 3-5 are repeated until all m-locus combinations are evaluated. Steps 3-1 to 3-5 explain FSMOMDR data training when selecting candidates.

Step 4: All candidates in the Pareto set j, where $j \in \{1, \ldots, k\}$, are evaluated using the j^{th} testing data in the CV. In each CV operation, all m-SNP combinations are evaluated using step 3 to generate the Pareto set. Ultimately, k Pareto sets are obtained, in which each candidate is counted according to its number of occurrences (denoted as CVC) in the k Pareto sets. The highest CVC of all candidates represents the optimal SNP–SNP interactions, in which the medians of the objective values of the testing data are SNP–SNP interaction measures (i.e., Step 4).

III. RESULTS

A. EPISTATIC MODELS WITHOUT MARGINAL EFFECTS

The data sets were simulated using 40 epistatic models without marginal effects, according to the multilocus penetrance [13]. In the 40 epistatic models, heritability (h^2) was set between 0.025 and 0.2 to control the phenotypic variation of the epistatic model. In the data set, the specific target (optimal SNP–SNP interaction) was generated through a minor allele frequency (*MAF*) of either 0.2 or 0.4 [36], and other SNPs were generated through an *MAF* uniformly selected from [0.05, 0.5]. GAMETES software was used to generate data sets according to the aforementioned settings [37]. In each data set, only one specific target exists. We randomly simulated 100 data sets in each epistatic model. The detection success rate of an epistatic model was calculated by counting the number of specific targets detected by the algorithm in 100 data sets.

We compared FSMOMDR with AntEpiSeeker [11], BOOST [12], BEAM [10], SNPRuler [13], MDR [16], PBMDR [17], CMDR [18], MOMDR [19], IMDR [20], and EFMDR [21] across epistatic models without marginal effects (Fig. 2). For epistatic models 1 to 10 ($h^2 \geq 0.2$), all methods had a strong ability to accurately detect the specific targets in each data set. For epistatic models 11 to 40 ($h^2 \leq 0.1$), FSMOMDR exhibited a superior detection success rate compared with those of MDR and EFMDR. However, the performance of FSMOMDR was inferior to that of SNPRuler (for model 34), BOOST (for models 25, 33, 34, and 40), and CMDR (for model 25). The performance of FSMOMDR in 40 epistatic models was evaluated using the Wilcoxon signed-rank test. A p value of <0.05 indicated significantly superior performance of FSMOMDR compared with its ten counterparts. Thus, as indicated in Table 1, FSMOMDR offers superior performance to that of its counterparts. Although $p > 0.05$ for FSMOMDR compared with BOOST, a trend of superiority was evident when comparing both methods. As for computation time, FSMOMDR took an average of 32.7 s to run a complete process in 40 epistatic models, including 1000 SNPs with 400 samples.

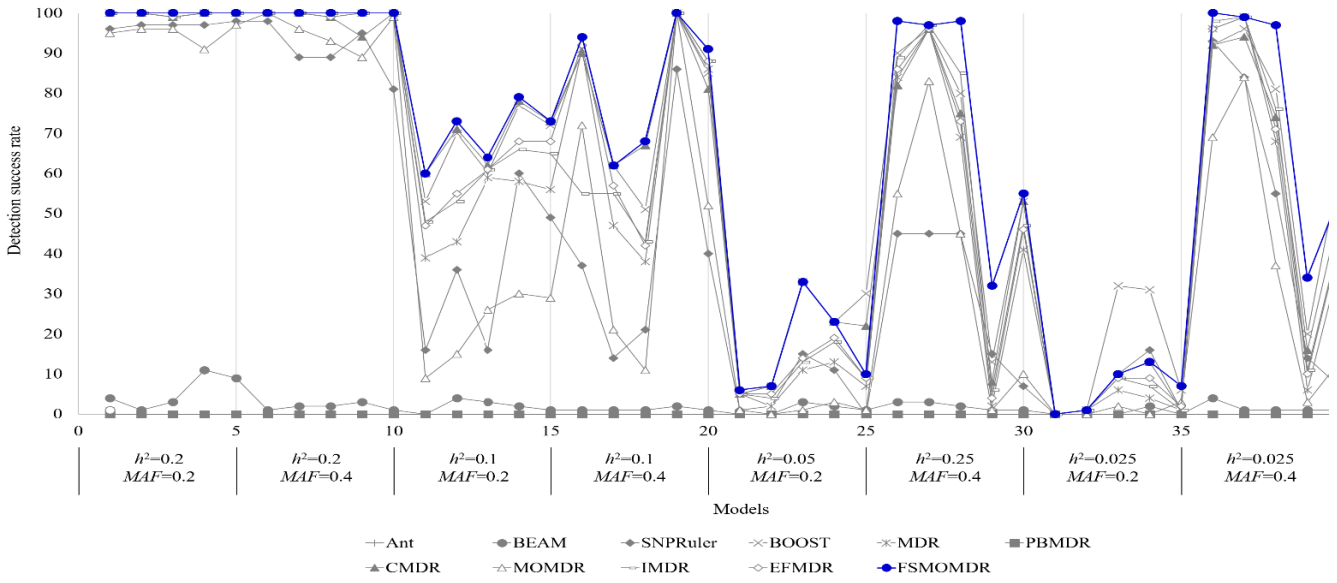

FIGURE 2. Comparison of detection success rates of AntEpiSeeker (Ant), BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, EFMDR, and FSMOMDR in disease models without marginal effects. Each dataset included 1000 SNPs, and sample sizes were 400 (200 cases and 200 controls). Detection success rate was in the proportion of 100 data sets in which specific disease-associated SNP–SNP interaction was detected.

TABLE 1. Comparison of Antepiseeker, BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, EFMDR, and FSMOMDR on 40 SNP–SNP interaction models using the Wilcoxon Signed-Rank test.

FSMOMDR *vs.*	R^+	R^-	$R^=$	*P* value
Antepiseeker	39	0	1	<0.001
BEAM	39	0	1	<0.001
SNPRuler	36	1	3	<0.001
BOOST	19	4	17	0.063
MDR	28	0	12	<0.001
PBMDR	39	0	1	<0.001
CMDR	21	1	18	<0.001
MOMDR	37	0	3	<0.001
IMDR	26	0	14	<0.001
EFMDR	28	0	12	<0.001

R^-: the degree to which FSMOMDR is inferior to the algorithm, R^+: the degree to which FSMOMDR is superior to the algorithm, $R^=$: the degree to which FSMOMDR is equal to the algorithm.

TABLE 2. Comparison of Antepiseeker, BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, EFMDR, and FSMOMDR on six SNP–SNP interaction models using the Wilcoxon Signed-Rank test.

FSMOMDR *vs.*	R^+	R^-	$R^=$	*P* value
Antepiseeker	6	0	0	0.028
BEAM	6	0	0	0.027
SNPRuler	6	0	0	0.027
BOOST	5	0	1	0.042
MDR	6	0	0	0.027
PBMDR	6	0	0	0.028
CMDR	5	1	0	0.249
MOMDR	6	0	0	0.027
IMDR	6	0	0	0.027
EFMDR	6	0	0	0.028

R^-: the degree to which FSMOMDR is inferior to the algorithm, R^+: the degree to which FSMOMDR is superior to the algorithm, $R^=$: the degree to which FSMOMDR is equal to the algorithm.

B. EPISTATIC MODELS WITH MARGINAL EFFECTS

The six multilocus penetrances were used to simulate epistatic models with marginal effect (models 1–6) [38]. We used GAMETES software [37] to simulate 100 data sets in each epistatic model, with the MAF evenly set at [0.05, 0.5]. In 100 data sets, the detection success rate was calculated by counting the number of specific targets detected by the algorithm.

Fig. 3 illustrates the detection success rates of AntEpiSeeker, BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, EFMDR, and FSMOMDR in six epistatic models. FSMOMDR was superior to other algorithms in six epistatic models with marginal effects. For the six epistatic models with marginal effects, a Wilcoxon signed-rank test indicated significant superiority in the detection success rate of FSMOMDR relative to that of the other nine algorithms (Table 2). A trend of superiority was evident in comparisons of FSMOMDR with CMDR. Our results suggest that the improved FS approach effectively enhanced MOMDR with respect to considering the uncertainty in the H/L classification of disease sites. As for computation time, FSMOMDR spent an average of 41.1 s running a complete process for each of the six epistatic models, including 1000 SNPs with 400 samples.

C. EXPERIMENT WITH REAL DATA

A real coronary artery disease (CAD) data set from the WTCCC database was used to evaluate the ability of FSMOMDR to detect SNP–SNP interactions. The WTCCC database was constructed in 2005 by 50 British research

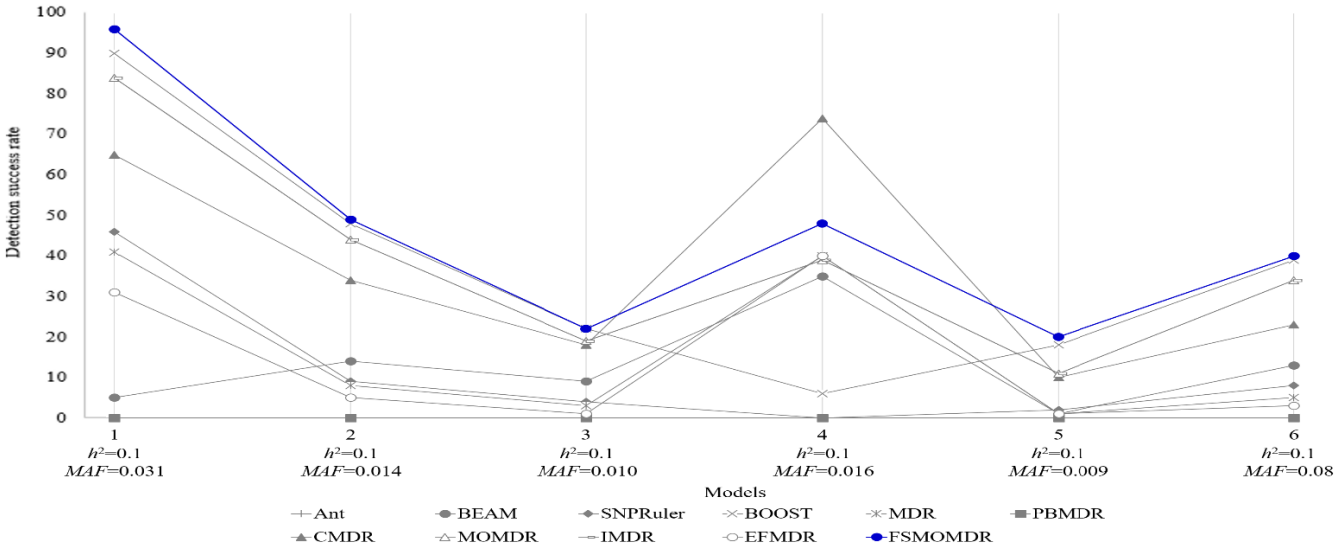

FIGURE 3. Comparison of detection success rates of AntEpiSeeker (Ant), BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, EFMDR and FSMOMDR in disease models with marginal effects. Each data set included 1000 SNPs, and sample sizes were 400 (200 cases and 200 controls). Detection success rate was calculated as the proportion for 100 data sets in which specific disease-associated epistasis was detected.

TABLE 3. FSMOMDR results for coronary artery disease on WTCCC data.

Location	SNP Groups	Related Genes	LR_{fuzzy} [*1]	CCR_{fuzzy} [*2]	CVC
Chr1	rs41399650, rs17163057	UNKNOWN, UNKNOWN	272.135	0.784	5
Chr2	rs41509345, rs41453947	UNKNOWN, UNKNOWN	242.958	0.783	5
Chr3	rs41367044, rs10866051	GTF2E1, LOC105376942	391.105	0.837	5
Chr4	rs41426946, rs41529544	PPA2, UNKNOWN	270.066	0.797	5
Chr5	rs41493746, rs41421845	UNKNOWN, LINC02107	91.429	0.674	5
Chr6	rs3006172, rs41489047	WDR27, ADGRB3	240.729	0.779	5
Chr7	rs41437948, rs41468749	POU6F2, GALNT17	55.271	0.639	5
Chr8	rs35120859, rs17480050	UNKNOWN, CSGALNACT1	139.631	0.722	4
Chr9	rs41354745, rs2891142	KANK1, SLC24A2	161.29	0.736	5
Chr10	rs41370151, rs2944490	FAM107B, TCERG1L	267.022	0.779	5
Chr11	rs41535846, rs41518446	UNKOWN, MAML2	47.620	0.627	2
Chr12	rs16926425, rs7299571	SOX5, UNKNOWN	712.602	0.957	5
Chr13	rs7328649, rs9540728	FAM155A, PCDH9	275.813	0.805	5
Chr14	rs41324950, rs41453247	LOC105370603, UNKNOWN	234.386	0.783	5
Chr15	rs41418744, rs41418548	UNKNOWN, SHC4	80.943	0.664	3
Chr16	rs235633, rs41483646	UNKNOWN, UNKNOWN	186.539	0.752	5
Chr17	rs3785579, rs1870998	CACNG1, UNKNOWN	494.358	0.871	2
	rs4969207, rs3785579	DNAH17, CACNG1	483.102	0.868	2
Chr18	rs4799934, rs3794931	CELF4, ZNF516	191.401	0.732	5
Chr19	rs375299, rs41370444	UNKNOWN, UNKNOWN	31.169	0.601	5
Chr20	rs2748666, rs41405046	UNKNOWN, UNKNOWN	440.051	0.872	5
Chr21	rs41378546, rs41451052	POU6F2, UNKNOWN	23.811	0.585	5
Chr22	rs41437948, rs41468749	POU6F2, GALNT17	52.509	0.634	4
ChrX	rs1419930, rs41500547	UNKNOWN, DMD	56.305	0.637	5

Chr: Chromosome; [*1] LR_{fuzzy} value is based on membership degree of fuzzy sigmoid approach. [*2] CCR_{fuzzy} value is based on membership degree of fuzzy sigmoid approach.

teams [39]. The CAD data set comprises 23 sub-datasets (chromosomes 1 to 22 and X), with 500569 SNPs in total. Each sample comprised 1988 patients with coronary heart disease and 1500 healthy people in the United Kingdom. All SNPs were genotyped using the Affymetrix gene chip 500 K mapping array.

Table 3 displays the FSMOMDR-detected SNP–SNP interactions. SNP information was determined by dbSNP at the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/snp/). Any SNP not on a gene was labeled "UNKNOWN." The chromosome had more than one SNP–SNP interaction, according to its multiobjective characteristic. We used raw data sets to evaluate the significance level of an SNP–SNP interaction, with the p values obtained from a chi-square test (χ^2). Among the 23 chromosomes, all SNP–SNP interactions detected by FSMOMDR yielded a p value of <0.0001, indicating a significant SNP–SNP interaction. The CVC indicated the

degree to which optimal SNP–SNP interaction was detected across a 5-fold CV, with CVC = 5 reflecting the highest degree [40]. In MDR, a high CCR value (>0.5) potentially reduces the frequency of chance [41]. A high LR value can reduce uncertainty in the disease model. High CCR and LR values indicated a strong contrast between cases and controls. As presented in Table 3, CCR_{fuzzy} values were in the range of 0.585–0.957, with an average value of 0.747 and a standard deviation (SD) of 0.096. LR_{fuzzy} values were in the range of 23.811–712.602, with an average value of 226.760 (SD = 171.880). The four SNP–SNP interactions exhibited high CCR_{fuzzy} (>0.8), LR_{fuzzy} (>250), and CVC (=5) values in addition to high significance ($p < 0.0001$), indicating that these interactions were potentially related to CAD. Further analysis of the implicated gene polymorphisms and their functional relevance is necessary. Table 3 lists the duration of operation of FSMOMDR for the WTCCC data set. We noted that the duration of operation increased proportionally with the number of SNPs.

IV. DISCUSSION

MDR was demonstrated to be a nonparametric approach for detecting nonlinear interaction between SNPs. MDR transforms the multifactor nonlinear combination from a high dimension to a low dimension. This may be an explanation for why 3^m multifactor genotypes could be transformed by binary classification into 2×2 contingency tables to improve the evaluation of SNP–SNP interaction [14]. Binary classification determines membership to high- or low-risk groups using the frequencies of multiple genotypes in cases and controls. However, binary classification may result in the loss of key information due to uncertainty [26]. Assuming that the balanced data set (i.e., where the classification threshold is 1) in the 2-SNP combination consists of nine multifactor genotypes, it is divided into a high-risk group (dominance ≥ 5.5) and low-risk group (dominance < 5.5). For a low multifactorial genotype, the membership degree is 2.5. MDR cannot distinguish between the two multifactorial genotypes. Although other studies have explored the shortcomings of MDR [21], [26], research in this field remains limited.

We determined the effectiveness of FSMOMDR by evaluating its performance in several epistatic models compared with the performance of other algorithms. For 40 epistatic models without marginal effects, FSMOMDR outperformed AntEpiSeeker in 39 models, BEAM in 39 models, SNPRuler in 36 models, BOOST in 19 models, MDR in 28 models, PBMDR in 39 models, CMDR in 21 models, MOMDR in 37 models, IMDR in 26 models, and EFMDR in 28 models. Furthermore, for six epistatic models with marginal effects, FSMOMDR outperformed AntEpiSeeker, BEAM, SNPRuler, MDR, PBMDR, MOMDR, IMDR, and EFMDR in all six models and outperformed BOOST and CMDR in five models. The results indicate that FSMOMDR has superior detection ability to AntEpiSeeker, BEAM, SNPRuler, BOOST, MDR, PBMDR, CMDR, MOMDR, IMDR, and EFMDR. Regarding the core principles and theoretical advantages of FSMOMDR, our algorithm handles uncertainty information through the FS approach, and fuzzy logic enables MOMDR to assign two membership degrees in each multifactor class because the two-element set {0, 1} is extended to the membership interval [0, 1]. For the interval [0, 1], our improved FS approach assesses the distance between the i^{th} multifactor class and outcome (cases and controls). This effectively improves the membership degrees of high- and low-risk groups in each multifactor class. This enables the effective detection of the m-locus combinations with similar distributions, thus enabling the detection of more significant SNP–SNP interactions. Jung *et al.* [26] introduced the original fuzzy-based MDR. The limitation of fuzzy-based MDR lies in its selection of the sigmoid function's parameters. Leem *et al.* introduced an EF function without such parameter selection to overcome the limitations of fuzzy-based MDR [21]. FSMOMDR used a ratio of cases to controls to map any region to the interval $[-1, 1]$. This strategy can reduce imbalance between the cases and controls [16], [42]. Thus, FSMOMDR does not select the parameter value of the fuzzy set. Moreover, in the 2×2 contingency table, our improved FS approach was superior to EFMDR with respect to the difference in the four cells (i.e., TP_f, FP_f, FN_f, and TN_f). Simulation experiments demonstrated that FSMOMDR has a higher detection success rate than EFMDR does. Moreover, the FSMOMDR can be extended by the neutrosophic set [43].

In addition to its retention of MDR's advantages, FSMOMDR has three other characteristics. First, it applies multiobjective measurement, which is based on the improved FS approach, to increase the distinction between multifactor classes for improved detection of potential SNP–SNP interactions. Second, to understand the distribution of multifactor classes associated with a particular disease, FSMOMDR can use the membership degree to graphically represent SNP–SNP interactions. Third, FSMOMDR has no need for selecting the parameters of the fuzzy sets.

The computation time of FSMOMDR is $k \times \binom{n}{m} \times s \times 3^m$ for the evaluation of the m-locus combinations between k-fold CV subsets in n SNPs with s samples. Specifically, for 100 data sets containing 1000 SNPs and 400 samples, the average computation times for MDR, EFMDR, and FSMOMDR are approximately 26, 28, and 31 s, respectively. For large data sets used in a GWAS, FSMOMDR has an approximate computation time of 29.14 h for 23 chromosomes. We recommend choosing from a large existing suite of computational methods, including parallel operation [44], graphics processing units-based MDR [45], the greedy search strategy [46], and differential evolution-based MDR [18], to improve FSMOMDR runtime.

V. CONCLUSION

We designed a powerful FS approach, FSMOMDR, for detecting SNP–SNP interactions. FSMOMDR is based on the improved sigmoid function, which allows it to yield better

information on MOMDR uncertainty. Each multifactor class can be evaluated with respect to its membership degree in the high-risk and low-risk groups, thus enabling FSMOMDR to detect more potential SNP–SNP interactions. FSMOMDR was demonstrated to have satisfactory power on real GWAS data sets; it can be used for SNP–SNP interaction detection. The findings of the present study suggest that the membership, based on the improved sigmoid function, can be used to identify SNP–SNP interactions in addition to obtaining content knowledge, thus reducing the information uncertainty in MDR.

REFERENCES

[1] W. S. Bush and J. H. Moore, "Genome-wide association studies," *PLoS Comput. Biol.*, vol. 8, no. 12, 2012, Art. no. e1002822.

[2] J. H. Moore, F. W. Asselbergs, and S. M. Williams, "Bioinformatics challenges for genome-wide association studies," *Bioinformatics*, vol. 26, no. 4, pp. 445–455, Jan. 2010.

[3] E. E. Eichler, J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore, and J. H. Nadeau, "Missing heritability and strategies for finding the underlying causes of complex disease," *Nature Rev. Genet.*, vol. 11, no. 6, pp. 446–450, Jun. 2010.

[4] H. J. Cordell, "Detecting gene–gene interactions that underlie human diseases," *Nature Rev. Genet.*, vol. 10, Jun. 2009, Art. no. 392.

[5] T. F. Mackay and J. H. Moore, "Why epistasis is important for tackling complex human disease genetics," *Genome Med.*, vol. 6, no. 6, p. 125, 2014.

[6] T. F. C. Mackay, "Epistasis and quantitative traits: Using model organisms to study gene–gene interactions," *Nature Rev. Genet.*, vol. 15, no. 1, pp. 22–33, Dec. 2014.

[7] C.-H. Yang, Y.-K. Kao, L.-Y. Chuang, and Y.-D. Lin, "Catfish Taguchi-based binary differential evolution algorithm for analyzing single nucleotide polymorphism interactions in chronic dialysis," *IEEE Trans. Nanobiosci.*, vol. 17, no. 3, pp. 291–299, Jul. 2018.

[8] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, and H.-W. Chang, "Analysis of high-order SNP barcodes in mitochondrial D-loop for chronic dialysis susceptibility," *J. Biomed. Informat.*, vol. 63, pp. 112–119, Oct. 2016.

[9] C.-H. Yang, S.-H. Moi, Y.-D. Lin, and L.-Y. Chuang, "Genetic algorithm combined with a local search method for identifying susceptibility genes," *J. Artif. Intell. Soft Comput. Res.*, vol. 6, no. 3, pp. 203–212, Jul. 2016.

[10] Y. Zhang and J. S. Liu, "Bayesian inference of epistatic interactions in case-control studies," *Nature Genet.*, vol. 39, no. 9, pp. 1167–1173, Aug. 2007.

[11] Y. Wang, X. Liu, K. Robbins, and R. Rekaya, "AntEpiSeeker: Detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm," *BMC Res. Notes*, vol. 3, no. 1, Apr. 2010, Art. no. 117.

[12] X. Wan, C. Yang, Q. Yang, H. Xue, X. Fan, N. L. S. Tang, and W. Yu, "BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies," *Amer. J. Hum. Genet.*, vol. 87, no. 3, pp. 325–340, Sep. 2010.

[13] X. Wan, C. Yang, Q. Yang, H. Xue, N. L. S. Tang, and W. Yu, "Predictive rule inference for epistatic interaction detection in genome-wide association studies," *Bioinformatics*, vol. 26, no. 1, pp. 30–37, Oct. 2010.

[14] M. D. Ritchie, L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl, and J. H. Moore, "Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer," *Amer. J. Hum. Genet.*, vol. 69, no. 1, pp. 138–147, Jul. 2001.

[15] D. Gola, J. M. Mahachie John, K. van Steen, and I. R. König, "A roadmap to multifactor dimensionality reduction methods," *Briefings Bioinf.*, vol. 17, no. 2, pp. 293–308, Jun. 2016.

[16] C.-H. Yang, Y.-D. Lin, L.-Y. Chuang, J.-B. Chen, and H.-W. Chang, "MDR-ER: Balancing functions for adjusting the ratio in risk classes and classification errors for imbalanced cases and controls using multifactor-dimensionality reduction," *PLoS ONE*, vol. 8, no. 11, Nov. 2013, Art. no. e79387.

[17] C.-H. Yang, H.-S. Yang, and L.-Y. Chuang, "PBMDR: A particle swarm optimization-based multifactor dimensionality reduction for the detection of multilocus interactions," *J. Theor. Biol.*, vol. 461, pp. 68–75, Jan. 2019.

[18] C.-H. Yang, L.-Y. Chuang, and Y.-D. Lin, "CMDR based differential evolution identifies the epistatic interaction in genome-wide association studies," *Bioinformatics*, vol. 33, no. 15, pp. 2354–2362, May 2017.

[19] C.-H. Yang, L.-Y. Chuang, and Y.-D. Lin, "Multiobjective multifactor dimensionality reduction to detect SNP–SNP interactions," *Bioinformatics*, vol. 34, no. 13, pp. 2228–2236, Feb. 2018.

[20] L. Y. Chuang, Y. D. Lin, and C. H. Yang, "Improved classification method for detecting potential interactions between genes," in *Intelligent Computing*. Cham, Switzerland: Springer, 2018, pp. 394–403.

[21] S. Leem and T. Park, "An empirical fuzzy multifactor dimensionality reduction method for detecting gene-gene interactions," *BMC Genomics*, vol. 18, no. S2, Mar. 2017, Art. no. 115.

[22] J. Gui, J. H. Moore, S. M. Williams, P. Andrews, H. L. Hillege, P. van der Harst, G. Navis, W. H. Van Gilst, F. W. Asselbergs, and D. Gilbert-Diamond, "A simple and computationally efficient approach to multifactor dimensionality reduction analysis of gene-gene interactions for quantitative traits," *PLoS ONE*, vol. 8, no. 6, Jun. 2013, Art. no. e66545.

[23] O.-Y. Fu, H.-W. Chang, Y.-D. Lin, L.-Y. Chuang, M.-F. Hou, and C.-H. Yang, "Breast cancer-associated high-order SNP-SNP interaction of CXCL12/CXCR4-related genes by an improved multifactor dimensionality reduction (MDR-ER)," *Oncol. Rep.*, vol. 36, no. 3, pp. 1739–1747, Jul. 2016.

[24] L.-Y. Chuang, H.-Y. Lane, Y.-D. Lin, M.-T. Lin, C.-H. Yang, and H.-W. Chang, "Identification of SNP barcode biomarkers for genes associated with facial emotion perception using particle swarm optimization algorithm," *Ann. Gen. Psychiatry*, vol. 13, no. 1, 2014, Art. no. 15.

[25] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, Jun. 1965.

[26] H.-Y. Jung, S. Leem, S. Lee, and T. Park, "A novel fuzzy set based multifactor dimensionality reduction method for detecting gene–gene interaction," *Comput. Biol. Chem.*, vol. 65, pp. 193–202, Dec. 2016.

[27] Z. Liu, C. Chen, and Y. Zhang, "Decentralized robust fuzzy adaptive control of humanoid robot manipulation with unknown actuator backlash," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 3, pp. 605–616, Jun. 2015.

[28] J. Huang, M. Ri, D. Wu, and S. Ri, "Interval type-2 fuzzy logic modeling and control of a mobile two-wheeled inverted pendulum," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2030–2038, Aug. 2018.

[29] Y.-J. Liu, Y. Gao, S. Tong, and Y. Li, "Fuzzy approximation-based adaptive backstepping optimal control for a class of nonlinear discrete-time systems with dead-zone," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 1, pp. 16–28, Feb. 2016.

[30] H.-Y. Jung, S. Leem, and T. Park, "Fuzzy set-based generalized multifactor dimensionality reduction analysis of gene-gene interactions," *BMC Med. Genomics*, vol. 11, no. S2, Apr. 2018, Art. no. 32.

[31] S. Leem and T. Park, "EFMDR-fast: An application of empirical fuzzy multifactor dimensionality reduction for fast execution," *Genomics Informat.*, vol. 16, no. 4, Dec. 2018, Art. no. e37.

[32] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy C-Means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, Oct. 2018.

[33] Z. Moslehi, M. Taheri, A. Mirzaei, and M. Safayani, "Discriminative fuzzy C-Means as a large margin unsupervised metric learning algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3534–3544, Dec. 2018.

[34] C. Liu, W. Huang, F. Sun, M. Luo, and C. Tan, "LDS-FCM: A linear dynamical system based fuzzy C-means method for tactile recognition," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 72–83, Jan. 2019.

[35] W. S. Bush, T. L. Edwards, S. M. Dudek, B. A. McKinney, and M. D. Ritchie, "Alternative contingency table measures improve the power and detection of multifactor dimensionality reduction," *BMC Bioinf.*, vol. 9, no. 1, 2008, Art. no. 238.

[36] R. J. Urbanowicz, J. Kiralis, N. A. Sinnott-Armstrong, T. Heberling, J. M. Fisher, and J. H. Moore, "GAMETES: A fast, direct algorithm for generating pure, strict, epistatic models with random architectures," *BioData Mining*, vol. 5, no. 1, Oct. 2012, Art. no. 16.

[37] J. Shang, J. Zhang, X. Lei, W. Zhao, and Y. Dong, "EpiSIM: Simulation of multiple epistasis, linkage disequilibrium patterns and haplotype blocks for genome-wide interaction analysis," *Genes Genomics*, vol. 35, no. 3, pp. 305–316, Jan. 2013.

[38] J. Namkung, K. Kim, S. Yi, W. Chung, M.-S. Kwon, and T. Park, "New evaluation measures for multifactor dimensionality reduction classifiers in gene–gene interaction analysis," *Bioinformatics*, vol. 25, no. 3, pp. 338–345, Jan. 2009.

[39] T. Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–678, Jun. 2007.
[40] A. A. Motsinger and M. D. Ritchie, "The effect of reduction in cross-validation intervals on the performance of multifactor dimensionality reduction," *Genetic Epidemiol.*, vol. 30, no. 6, pp. 546–555, 2006.
[41] C. S. Coffey, P. R. Hebert, M. D. Ritchie, H. M. Krumholz, J. M. Gaziano, and P. M. Ridker, "An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: The importance of model validation," *BMC Bioinf.*, vol. 5, no. 1, 2004, Art. no. 49.
[42] C.-H. Yang, Y.-D. Lin, and L.-Y. Chuang, "Class balanced multifactor dimensionality reduction to detect gene–gene interactions," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 1, pp. 71–81, Jan. 2020.
[43] F. Smarandache, "Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, spherical fuzzy set, and q-rung orthopair fuzzy set, while neutrosophication is a generalization of regret theory, grey system theory, and three-ways decision (revisited)," *J. New Theory*, no. 29, pp. 1–31, 2019.
[44] W. S. Bush, S. M. Dudek, and M. D. Ritchie, "Parallel multifactor dimensionality reduction: A tool for the large-scale analysis of gene-gene interactions," *Bioinformatics*, vol. 22, no. 17, pp. 2173–2174, Jun. 2006.
[45] C. S. Greene, N. A. Sinnott-Armstrong, D. S. Himmelstein, P. J. Park, J. H. Moore, and B. T. Harris, "Multifactor dimensionality reduction for graphics processing units enables genome-wide testing of epistasis in sporadic ALS," *Bioinformatics*, vol. 26, no. 5, pp. 694–695, Jan. 2010.
[46] C.-H. Yang, Y.-D. Lin, C.-S. Yang, and L.-Y. Chuang, "An efficiency analysis of high-order combinations of gene–gene interactions using multifactor-dimensionality reduction," *BMC Genomics*, vol. 16, no. 1, Jul. 2015, Art. no. 489.

CHENG-HONG YANG (Senior Member, IEEE) received the Ph.D. degree in computer engineering from North Dakota State University, in 1992. He was the President of the National Kaohsiung University of Applied Science, Taiwan, from 2012 to 2016. He is currently the Chair Professor with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. He has authored/coauthored over 380 refereed publications and a number of book chapters. He is an Editorial Board Member of multiple international journals. His main areas of research are fuzzy control, evolutionary computation, bioinformatics, machine learning, and data analysis. He is a Fellow of the Institution of Engineering and Technology and the American Biographical Institute.

LI-YEH CHUANG received the M.S. degree from the Department of Chemistry, University of North Carolina, in 1989, and the Ph.D. degree from the Department of Biochemistry, North Dakota State University, in 1994. She is currently a Professor with the Department of Chemical Engineering and the Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan. She has authored/coauthored over 300 refereed publications. Her main areas of research are bioinformatics, biochemistry, and genetic engineering.

YU-DA LIN (Member, IEEE) received the M.S. and Ph.D. degrees from the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan, in 2011 and 2015, respectively. He is currently a Postdoctoral Fellow with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. He is also a Software Engineer and an Adjunct Assistant Professor. His main research interests include artificial intelligence, biomedical informatics, bioinformatics, and computational biology. He has authored/coauthored over 90 refereed publications. He is a member of the IEEE Tainan Section, the IEEE Young Professionals, and the IEEE Computational Intelligence Society Membership.