Applied Intelligence (2023) 53:29433–29447
https://doi.org/10.1007/s10489-023-05080-8

Multimodal fuzzy granular representation and classification

Fenggang Han[1] · Xiao Zhang[2] · Linjie He[2] · Liru Kong[2] · Yumin Chen[2]

Accepted: 30 September 2023 / Published online: 31 October 2023

Abstract

In a complex classification task, samples are represented by various types of multimodal features, including structured data, text, images, video, audio, etc. These data are usually high dimensionally, large-sized, structurally complex, and semantically inconsistent. The representation, translation, alignment, fusion and co-learning of multimodal data are core technical challenges to traditional classification tasks. Kernel functions are applied in dealing with multimodal data for extracting some nonlinear information. However, they cannot consider the aspects of complex structures and uncertain semantics in a multimodal classification task. Fuzzy granular computing emerges as a powerful vehicle to handle the structured and uncertain multimodal data. In this paper, we propose a framework of multimodal classification based on kernel functions and fuzzy granular computing. First, a fuzzy granulation based on kernel functions is introduced to extract nonlinear features for the multimodal classification. Then, a model of multimodal fuzzy classification including fuzzy granular representation, fusion and learning for multimodal data is constructed. Finally, we design an efficient fuzzy granular classification algorithm for big multimodal data based on the proposed model. Experimental results demonstrate the effectiveness of our proposed model and its corresponding algorithm.

Keywords Multimodal learning · Fuzzy sets · Fuzzy granules · Granular computing · Multimodal classification

1 Introduction

With the advent of the big data era, it is a fact that heterogeneous and multimodal data widely appear [1]. In real application systems, data and information from multifarious hardware and software sources exhibit heterogeneity and multimodality [2]. Those sources encompass a variety of hardware sensors such as radar, infrared, accelerometers; people's sense of touch, hearing, vision, and smell; software-collected information media, including audio, images, video, and text; and pattern recognition results, such as cars, pedestrians, and license plates, etc. In a medical diagnosis system, there exists gigantic multimodal data [3] including categorical data from detection results, numerical data from urine and blood tests, and images from X-ray, B-ultrasound, CT, and MRI. Modality refers to a certain type of information or the representation format in which information is stored [4]. Multimodality is the integration of multiple types of information and data. In the future internet of things, multimodal data will be produced through a mixture of text, images, video, audio, data readings from devices, perceptual signals from sensors, and recognition results from intelligent systems. Individuals can easily make decisions based on such complex multimodal information, but intelligent software and machines still struggle with it difficultly. Thus, it becomes a challenging task to develop an effective representation and evaluation model for processing these multimodality data.

In multimodal learning, there are three core challenges: representation, fusion and learning. Representation involves the study of how to effectively represent and summarize multimodal data in away that leverages complementarity

✉ Liru Kong
k599651599@163.com

Fenggang Han
hanfg@xmut.edu.cn

Xiao Zhang
zhangxiaomyz@gmail.com

Linjie He
linjiehe_jason@foxmail.com

Yumin Chen
cym0620@163.com

1 School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024, China

2 School of Computer & Information Engineering, Xiamen University of Technology, Xiamen 361024, China

and redundancy [4]. It serves as foundation for both fusion [5] and learning [6]. A well-crafted representation can generate structured and fusible information that leads to an effective classifier. There are two typical methods for representing multimodal data: the first method separates features into different classifiers to data create various joints, data which are then integrates into an ensemble classifier. The second method aggregates the features into a long vector for learning linear projections that are maximize correlation. Joint representation is sensitive to the choice of classifiers, while coordinated representation ultrahigh dimensionality and neglects the structural information within the data. Furthermore, extracting the fuzzy information from multimodal data remains a challenging task.

Multimodal data originating from various information sources usually results in a nonlinear distribution. To enhance classification accuracy, multi-kernel learning has been proposed to transform heterogeneous features into a unified representation framework [7, 8]. Kernel functions are employed to measure the similarity between samples described by different types of features. A linear kernel is used for linearly distributed data, while the Gaussian kernel is applied to data with a nonlinear distribution. A string kernel assesses the similarity of two strings in gene analysis [9] kernel, and a histogram intersection kernel evaluates the similarity of two images [10]. In recent years, numerous studies have been conducted on multi-kernel learning [11], feature selection [12, 13], and classification and recognition [14–16].

Zadeh proposed fuzzy sets and fuzzy computing [17] in 1979. Fuzzy computing is based on fuzzy set theory and simulates the inaccurate and nonlinear information processing capabilities of human brain, including fuzzy inference systems [18], fuzzy logic [19], and fuzzy systems [20] etc. It has been widely used in many fields [21–25]. Zadeh further studied the problem of fuzzy information granulation [26–28], and believed that human cognitive ability can be summarized as three main characteristics of granulation, organization and causality. Pedrycz pointed out that the construction of information granule is the key to granular computing. He constructed an information granule from the perspective of fuzzy sets and used it for clustering [29] and classification [30], and proposed a variety of granular classifiers [31, 32]. Other scholars employed granulation and multigranulation technologies for attribute reduction [33, 34], feature selection [35–37], safety analysis [38, 39], three-way decisions [40], regression [41], classification [42, 43] and clustering [44]. Fuzzy granulation is an effective method for structuring various information and data [45, 46]. However, fuzzy granulation technology is rarely applied to multimodal data.

From the above analysis we can draw the conclusion that multimodal learning and classification mainly suffer from two essential shortcomings: 1) For multimodal big data, there is no structural representation for multimodal learning that provides valid fusion and classification, and 2) For uncertain multimodal data, multi-kernel learning is not well-suited for handling fuzzy classification tasks.

In this study, we design a framework of fuzzy granular representation and classification for multimodal data, which aims to develop structural data representation that is beneficial for multimodal learning. The structural data is expressed by granules. Various granules originated from modal data are combined into a granular vector. Thus the multimodal classifier can be founded by a learning process on granular vectors. Furthermore, each modality may be an image, a text, or a voice resulting semantic diversification. The formats of multimodal data are structured, semi-structured and unstructured leading to data diversification. These characteristics of multimodal data reflect the uncertainty and fuzziness. It is necessary to develop a fuzzy granulation method for tackling these vague information from multimodal data. The contributions of this paper have three aspects. First, using fuzzy granulation and kernel functions, we develop a structural representation frame. Second, we further propose some granular operations to induce a fusion technique for multimodal data. Third, we design a granular classification algorithm to handle fuzzy multimodal data based on the proposed frame.

We illustrate the proposed framework of multimodal fuzzy granular representation and classification in Fig. 1. In a multimodal system, it is characterized by numerous multimodal features. These features can take the form of images, texts, acoustic waves, charts, time series or tables. The kernel function determines the similarity matrix among samples, considering the corresponding features of a single modality. The fuzzy granules can be generated from the similarity matrix. Subsequently, these granules are fused into a unified granular vector, which is utilized for training and testing a granular classifier.

The paper is organized as follows. First, in Section 2, we introduce the preliminaries regarding multimodal learning, granular computing and fuzzy sets. Then, in Section 3 we present multimodal fuzzy granular representation. Furthermore, in Section 4, we propose a multimodal fuzzy granular classifier based on this granular representation. In Section 5, we conduct a series of experiments and illustrate their results. Finally, the conclusion and future work are covered in Section 6.

2 Preliminaries

In this section, we will introduce multimodal learning and reveal its specific branch: multi-kernel learning. Additionally, we will review some definitions related to granular computing and fuzzy sets are reviewed.

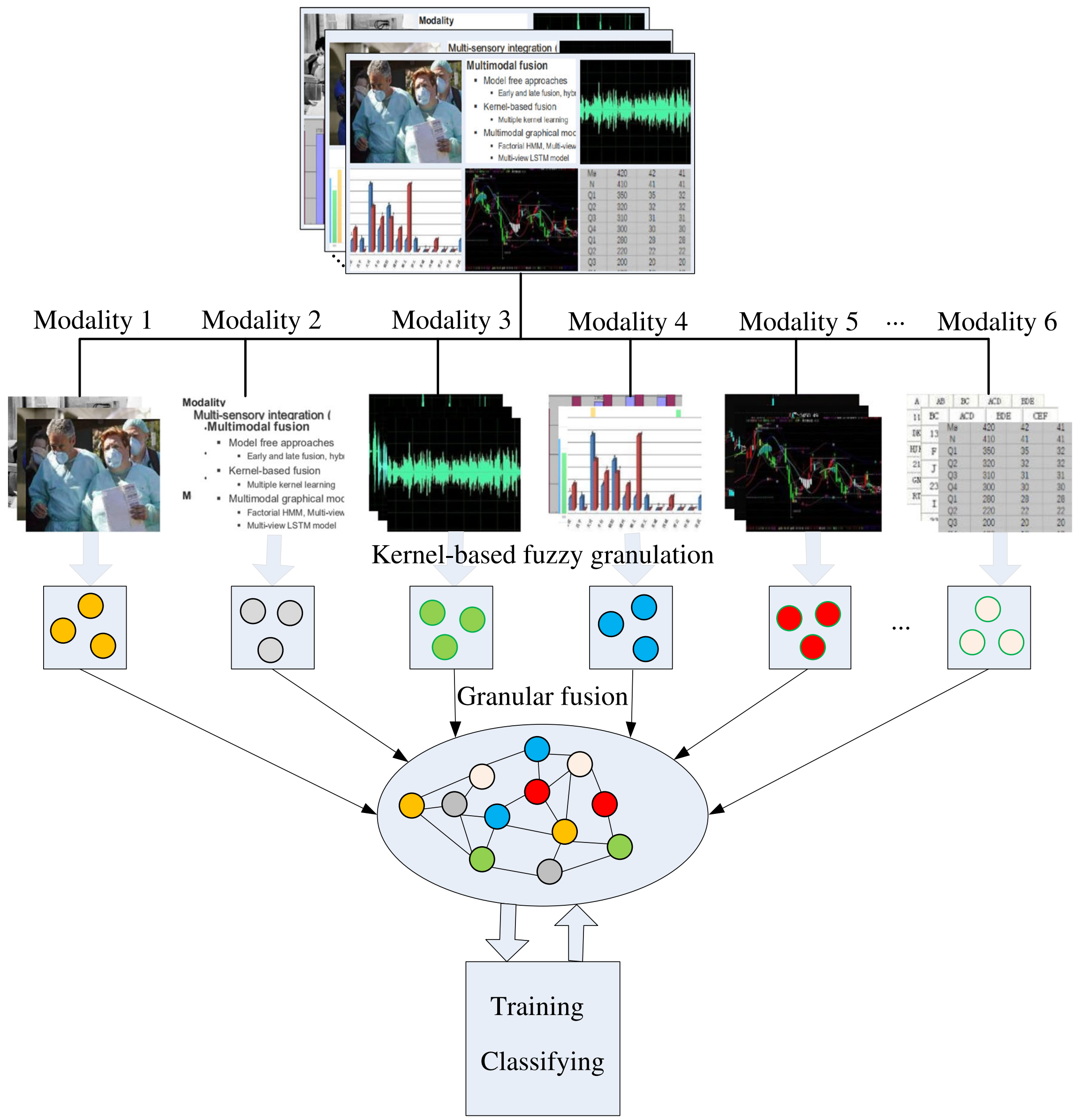


Fig. 1 A framework of multimodal fuzzy granular representation and classification

2.1 Multimodal learning

The world around us encompasses multiple modalities. A research problem or dataset is therefore characterized as multimodal when it includes multiple modalities. Multimodal learning aims to build models that can process and relate information from multiple modalities. A kernel can be used to extract information from a single modality. Multi-kernel learning is developed for multimodal learning, which exploits kernels that naturally correspond to different modalities and combines kernels either linearly or non-linearly to improve learning performance.

Given a training set $T = \{(x_i, y_i)_{i=1}^n\}$, where $y_i \in \{+1, -1\}$ is the label of sample x_i, for a test sample x, the classification decision function of multi-kernel learning is defined as:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K_\beta(x_i, x) + b), \tag{1}$$

where α_i is the parameter corresponding to each sample, b is the bias, and $K_\beta(x_i, x)$ is calculated as:

$$K_\beta(x_i, x) = \sum_{m=1}^{P} \beta_m k_m(x_i^m, x^m), \tag{2}$$

where $k_m(x_i^m, x^m)$ is the base kernel of samples x_i, x on modality m, and β_m is the parameter of modality m.

2.2 Granular computing

Granular computing is a new concept and computing paradigm. Its essence is to solve a complex problem with an appropriate level of granularity, which aims to decrease the redundancy and complexity. Granular computing is an umbrella including rough sets, neighborhood rough sets, fuzzy sets and quotient spaces that are adept in handling inconsistent and uncertain information.

$IS = (U, C)$ is an information system, where U is a sample set, C is a set of condition features. U is partitioned into a family of granules $g_R(x)$ by a relation R derived with the features. There are mainly three relations: equivalence, fuzziness and similarity.

Given $x, y \in U$, for a relation B that is generated by the feature subset $B \subseteq C$, the granule $g_B(x)$ of x is defined as:

$$g_B(x) = \{y|x, y \in U, d_B(x, y) \leq \delta), \tag{3}$$

where $\delta \in R^+$ is a granulation threshold and $d_B(x, y)$ is a similarity or distance metric that satisfies: 1) $d_B(x, y) \geq 0$, 2) $d_B(x, y) = d_B(y, x)$, and 3) for any $x, y, z \in U$, $d_B(x, y) + d_B(y, z) > d_B(x, z)$. The different granulation thresholds bring about various relations resulting in different granules.

If $d_B(x, y)$ is a similarity metric and $\delta = 1$ (or $d_B(x, y)$ is a distance metric and $\delta = 0$), then it induces an equivalence relation and produces equivalence granules. Further, rough sets can be represented by these granules. Given $X \subseteq U$, the lower and upper approximate sets of X on B are defined as:

$$B_*(X) = \cup\{x \in U|g_B(x) \subseteq X\}, \tag{4}$$

$$B^*(X) = \cup\{x \in U|g_B(x) \cap X \neq 0\}. \tag{5}$$

The order couple $< B_*(X), B^*(X) >$ is called as a rough set that utilizes the lower and upper approximate sets to approach the X set.

If $d_B(x, y)$ is a distance metric and $\delta \in R^+$, then it induces a neighborhood (similarity) relation and produces neighborhood granules. Therefore, neighborhood rough sets can be represented by these granules.

2.3 Fuzzy sets

In the real world, there are a lot of fuzzy phenomena and objects. Fuzzy set theory was founded by Zadeh in 1965 to describe fuzzy objects and has been widely used in the field of control, as referenced in [47, 48]. A fuzzy set is used to express the fuzziness of a concept. The appropriate membership function is employed to characterize the fuzziness. The fuzzy object is analyzed by the relevant operation and transformation of the fuzzy set.

Suppose A is a fuzzy set, the $\mu_A()$ is a membership function of A, abbreviated as $A()$. For each $x \in U$, $\mu_A(x)$ is called the membership degree of element x belonging to fuzzy set A. Given a nonempty and finite domain $U = \{x_1, x_2, ..., x_n\}$, the fuzzy set A is expressed as:

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + ... + \frac{\mu_A(x_n)}{x_n}. \tag{6}$$

Suppose A, B are two fuzzy sets on U, for any $x \in U$ and its membership degrees $\mu_A(x)$, $\mu_B(x)$, there are four common operators listed in the follows.

Zadeh operator:

$$\mu_A(x) \vee \mu_B(x) = max\{\mu_A(x), \mu_B(x)\}, \tag{7}$$

$$\mu_A(x) \wedge \mu_B(x) = min\{\mu_A(x), \mu_B(x)\}. \tag{8}$$

Algebraic operator:

$$\mu_A(x) \mp \mu_B(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x), \tag{9}$$

$$\mu_A(x) \cdot \mu_B(x) = \mu_A(x) \times \mu_B(x). \tag{10}$$

Bounded operator:

$$\mu_A(x) \oplus \mu_B(x) = min\{1, \mu_A(x) + \mu_B(x)\}, \tag{11}$$

$$\mu_A(x) \odot \mu_B(x) = max\{0, \mu_A(x) + \mu_B(x) - 1\}. \tag{12}$$

Einstein operator:

$$\mu_A(x) \overset{+}{\epsilon} \mu_B(x) = \frac{\mu_A(x) + \mu_B(x)}{1 + \mu_A(x) \cdot \mu_B(x)}, \tag{13}$$

$$\mu_A(x) \dot{\epsilon} \mu_B(x) = \frac{\mu_A(x) \cdot \mu_B(x)}{1 + (1 - \mu_A(x))(1 - \mu_B(x))}. \tag{14}$$

3 Multimodal fuzzy granular representation

Multimodal data include categorical, numerical, text, images, audio, video, etc. They may be inconsistent, fuzzy, and uncertain, making them difficult to tackle. A fuzzy set is used to express a vague object. Generally, it describes fuzziness by establishing an appropriate membership function and employs its operations and transformations to analyze uncertain objects. In the following, we introduce some kernel functions to granulate these multimodal data fuzzily, and further construct a fuzzy representation for multimodal learning.

3.1 Kernel functions for multimodal learning

A multimodal information system includes various types of data, and their relations are measured by different kernel functions listed as follows.

1. Regarding numerical data, three kernel functions are used to extract features: linear, polynomial and Gaussian [50]

kernels. Suppose x, y are a pair of samples, the three kernel functions are expressed as follows:

Linear kernel:

$$k(x, y) = x^T y; \tag{15}$$

Polynomial kernel:

$$k(x, y) = (x^T y)^d; \tag{16}$$

Gaussian kernel:

$$k(x, y) = e^{-\frac{||x-y||^2}{\delta^2}}; \tag{17}$$

where d is the parameter of polynomial kernel and δ is the parameter of Gaussian kernel.

2. As for categorical data, we use the Hamming distance for defining a match kernel:

Match kernel:

$$k(x, y) = \frac{\sum_{i=1}^{m} h(x_i, y_i)}{m}, \tag{18}$$

where m is the dimension of x and

$$h(x_i, y_i) = \begin{cases} 0 & x_i \neq y_i; \\ 1 & x_i = y_i. \end{cases} \tag{19}$$

3. For images and video data, the histogram intersection kernel [10] is used for measuring the similarity between two images:

$$k(x, y) = \sum_{j=1}^{r} min(x(j), y(j)), \tag{20}$$

where x and y are color histograms extracted from two images; $x(j)$ denotes the count of pixels in the jth bin of the histogram; and $\sum_{j=1}^{r} x(j) = 1$. Histogram intersection kernel calculates the overlap of histograms of two images.

4. Cosine kernel [51] is designed for computing similarity between two word vectors:

$$k(x, y) = \frac{xy^T}{||x||||y||}, \tag{21}$$

where x and y are word vectors extracted from texts or documents. The word vectors can be represented by word vector models, such as Word2vec [52], Glove [53] and TF-IDF [51].

5. Cauchy kernel originated from the Cauchy distribution [54] is employed in mel-frequency cepstral coefficients (MFCCs) features [55] that are extracted from audio data. The Cauchy kernel is expressed as:

$$k(x, y) = \frac{1}{1 + \frac{||x-y||^2}{\delta}}, \tag{22}$$

where x and y are the audio feature samples. The Cauchy kernel is a long-tailed kernel and can be used to give long-range influence and sensitivity over the high dimension space.

The value domain of these kernel functions is [0, 1]. It can be known that these kernels satisfy $k(x, x) = 1$ and $k(x, y) = k(y, x)$. Therefore, these kernel functions are reflexive and symmetrical. The relations induced by them are similarity or fuzzy relations.

3.2 Fuzzy granulation by kernel functions

$MDS = (U, MC, \{d\})$ is a multimodal decision system, where $U = \{x_1, x_2, ..., x_n\}$ is a set of samples or objects; $MC = \{M_1, M_2, ..., M_P\}$ is a set of multimodal condition features containing P different single modal features; the dimensionality of each feature may be different; and $\{d\}$ is a decision feature.

Suppose $MDS = (U, MC, \{d\})$ is a multimodal decision system, for any two samples $x_i, x_j \in U$, a kernel function $k_s(x_i, x_j)$ is computed for a single modal feature $M_s \in MC$ on samples x_i, x_j. For example, the Gaussian kernel is $k_s(x_i, x_j) = e^{-\frac{||v(x_i,s)-v(x_j,s)||^2}{\delta^2}}$, where $v(x_i, s)$ is a value of x_i on the single modal feature M_s.

Definition 1. Let $MDS = (U, MC, \{d\})$ be a multimodal decision system, for any sample $x_i \in U$ and a single modal feature $M_s \in MC$, the fuzzy granule of x_i on M_s is defined as:

$$g_s(x_i) = \sum_{j=1}^{n} \frac{k_s(x_i, x_j)}{x_j} = \frac{k_s(x_i, x_1)}{x_1} + \frac{k_s(x_i, x_2)}{x_2} + ... + \frac{k_s(x_i, x_n)}{x_n}, \tag{23}$$

where $n = |U|$, $+$ is a union of elements, '-' is a separator between a sample and its kernel. It is simply denoted by $g_s(x_i) = \{k_s(x_i, x_1), k_s(x_i, x_2), ..., k_s(x_i, x_n)\}$.

For the decision feature d, since its values are categorical data, the sample $x \in U$ is granulated into a fuzzy decision granule by the match kernel. The fuzzy granule is essentially a set of ordering kernels measuring similarities between samples.

Given a fuzzy granule $g_s(x_i)$ for x_i on M_s, the size of the fuzzy granule is represented as:

$$Size(g_s(x_i)) = |g_s(x_i)| = \sum_{j=1}^{n} k_s(x_i, x_j). \tag{24}$$

Since the value domain of kernel function is [0, 1], the size of a fuzzy granule satisfies: $0 \leq Size(g_s(x_i)) \leq n$.

Given any two fuzzy granules $g_s(x_i)$, $g_s(x_j)$ for x_i, x_j on M_s, the distance between two fuzzy granules is defined as:

$$Dis(g_s(x_i), g_s(x_j)) = Size(|g_s(x_i) - g_s(x_j)|), \tag{25}$$

where $g_s(x_i) - g_s(x_j) = \sum_{t=1}^{n} \frac{k_s(x_i,x_t)-k_s(x_j,x_t)}{x_t}$.

3.3 Multimodal granular vectors

In the area of machine learning, vectors and matrices are excellent representation forms for constructing a classifier. For the convenience of structural data and machine learning, we can combine fuzzy granules into a multimodal granular vector.

Definition 2 Let $MDS = (U, MC, \{d\})$ be a multimodal decision system, for any sample $x_i \in U$, a multimodal feature subset $MP \subseteq MC$, and $MP = \{M_1, M_2, ..., M_m\}$, then the fuzzy granular vector of x_i on the multimodal feature subset MP is defined as:

$$V_P(x_i) = (g_1(x_i), g_2(x_i), ..., g_m(x_i)). \tag{26}$$

The $V_P(x_i)$ is a fuzzy granular vector on the multimodal feature MP, which is composed of fuzzy granules. The $g_1(x_i)$ is a fuzzy granule of the sample x_i on the single modal feature M_1. It is also called an element of the fuzzy granular vector. Therefore, the elements of fuzzy granular vectors are sets. Unlike traditional vectors, the elements of them are real numbers.

Given a fuzzy granular vector $V_P(x_i)$ for x_i on a multimodal feature M_P, the size of the fuzzy granular vector is represented as:

$$Size(V_P(x_i)) = |V_P(x_i)| = \sum_{s=1}^{m} |g_s(x_i)| = \sum_{s=1}^{m}\sum_{j=1}^{n} k_s(x_i, x_j). \tag{27}$$

The size of a fuzzy granular vector is also called the modulus of a fuzzy granular vector. It is easy to know that the size of a fuzzy granular vector satisfies: $0 \leq |V_P(x_i)| \leq m * n$.

Figure 2 shows an example of multimodal representation. Three samples are described using the four modalities 1, 2, 3 and 4, where modality 1 is expressed by categorical data, modality 2 is expressed by numerical data, modality 3 is expressed by audio data and modality 4 is expressed by images. We use match kernel k_1, Gaussian kernel k_2, Cauchy kernel k_3 and histogram intersection kernel k_4 to extract features and granulate these samples. Thus, we can obtain fuzzy

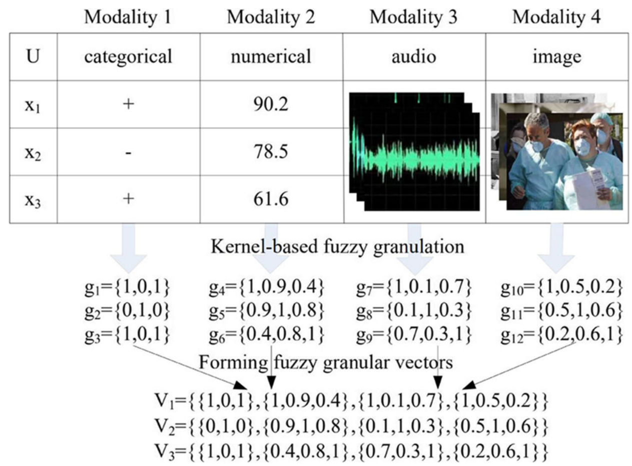


Fig. 2 Multimodal granular representation

granules and form fuzzy granular vectors. For example, we granulate sample x_1.

$k_1(x_1, x_1) = 1$; $k_1(x_1, x_2) = 0$; $k_1(x_1, x_3) = 1$; so the granule of x_1 on modality 1 is $\{1, 0, 1\}$

$k_2(x_1, x_1) = 1$; $k_2(x_1, x_2) = 0.9$; $k_2(x_1, x_3) = 0.4$; so the granule of x_1 on modality 2 is $\{1, 0.9, 0.4\}$

$k_3(x_1, x_1) = 1$; $k_3(x_1, x_2) = 0.1$; $k_3(x_1, x_3) = 0.7$; so the granule of x_1 on modality 3 is $\{1, 0.1, 0.7\}$

$k_4(x_1, x_1) = 1$; $k_4(x_1, x_2) = 0.5$; $k_4(x_1, x_3) = 0.2$. so the granule of x_1 on modality 4 is $\{1, 0.5, 0.2\}$

We combine the four granules into a fuzzy granular vector. Thus, we obtain:

$V_1 = \{\{1, 0, 1\}, \{1, 0.9, 0.4\}, \{1, 0.1, 0.7\}, \{1, 0.5, 0.2\}\}$.

4 Multimodal fuzzy granular classification

By means of multimodal fuzzy granulation, a sample is granulated into some fuzzy granules by kernel functions on different modal features. Each granule represents an extracted feature from a single modality. Determining how to combine these granules poses a significant challenge for classifiers. Initially, we define some operations of fuzzy granules. Then, a fusion method for fuzzy granules is presented by operations of granules, which converts single modal fuzzy granules into a multimodal granule. At the same time, the decision label of the sample is granulated into a fuzzy decision granule. Finally, we minimize the discrepancy of the multimodal granule and the decision granule to induce a multimodal fuzzy granular classification.

4.1 Operations of fuzzy granules

Definition 3. Let $MDS = (U, MC, \{d\})$ be a multimodal decision system, for any sample $x_i \in U$ and two single modal features $a, b \in MC$, $s = g_a(x_i) = \sum_{j=1}^{n} \frac{k_a(x_i,x_j)}{x_j}$, $t = g_b(x_i) = \sum_{j=1}^{n} \frac{k_b(x_i,x_j)}{x_j}$ are two fuzzy granules on a and b, then the intersection, union, addition, subtraction and multiplication operations of the two fuzzy granules are defined as:

$$s \wedge t = g_a(x_i) \wedge g_b(x_i) = \sum_{j=1}^{n} \frac{min(k_a(x_i, x_j), k_b(x_i, x_j))}{x_j}, \quad (28)$$

$$s \vee t = g_a(x_i) \vee g_b(x_i) = \sum_{j=1}^{n} \frac{max(k_a(x_i, x_j), k_b(x_i, x_j))}{x_j}, \quad (29)$$

$$s + t = g_a(x_i) + g_b(x_i) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) + k_b(x_i, x_j)}{x_j}, \quad (30)$$

$$s - t = g_a(x_i) - g_b(x_i) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) - k_b(x_i, x_j)}{x_j}, \quad (31)$$

$$s * t = g_a(x_i) * g_b(x_i) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) * k_b(x_i, x_j)}{x_j}, \quad (32)$$

$$s \oplus t = g_a(x_i) \oplus g_b(x_i) = \sum_{j=1}^{n} \times \frac{max(k_a(x_i, x_j), k_b(x_i, x_j)) - min(k_a(x_i, x_j), k_b(x_i, x_j))}{x_j}, \quad (33)$$

where $\sum$ represents a union of elements, $k_a(x_i, x_j)$ is a kernel function, '-' represents a separator between a sample and the kernel function expression.

Definition 4. Let $MDS = (U, MC, \{d\})$ be a multimodal decision system, for any two samples $x_i, x_k \in U$ and a single modal feature $a \in MC$, $s = g_a(x_i) = \sum_{j=1}^{n} \frac{k_a(x_i,x_j)}{x_j}$, $t = g_a(x_k) = \sum_{j=1}^{n} \frac{k_a(x_k,x_j)}{x_j}$ are two fuzzy granules of x_i, x_k on a, where $k_a(x_i, x_j)$ represents a kernel function of samples x_i, x_j, then the intersection, union, addition, subtraction, multiplication and XOR operations of the two fuzzy granules are defined as:

$$s \wedge t = g_a(x_i) \wedge g_a(x_k) = \sum_{j=1}^{n} \frac{min(k_a(x_i, x_j), k_a(x_k, x_j))}{x_j}, \quad (34)$$

$$s \vee t = g_a(x_i) \vee g_a(x_k) = \sum_{j=1}^{n} \frac{max(k_a(x_i, x_j), k_a(x_k, x_j))}{x_j}, \quad (35)$$

$$s + t = g_a(x_i) + g_a(x_k) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) + k_a(x_k, x_j)}{x_j}, \quad (36)$$

$$s - t = g_a(x_i) - g_a(x_k) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) - k_a(x_k, x_j)}{x_j}, \quad (37)$$

$$s * t = g_a(x_i) * g_a(x_k) = \sum_{j=1}^{n} \frac{k_a(x_i, x_j) * k_a(x_k, x_j)}{x_j}, \quad (38)$$

$$s \oplus t = g_a(x_i) \oplus g_a(x_k) = \sum_{j=1}^{n} \times \frac{max(k_a(x_i, x_j), k_a(x_k, x_j)) - min(k_a(x_i, x_j), k_a(x_k, x_j))}{x_j}. \quad (39)$$

Samples are granulated into fuzzy granules, which are sets of similarity measures. The intersection, union, subtraction and XOR operations between fuzzy granules are intersection, union, subtraction and exclusive or operations between sets. Definition 3 denotes operations of fuzzy granules with a same sample under different modal features, while Definition 4 represents operations of fuzzy granules with different samples under a same modality.

4.2 Fusion of fuzzy granules

In a multimodal decision system, given a fuzzy granule $g_s(x)$ for x on M_s and a weight $w_s = \{w_{1s}, w_{2s}, ..., w_{ns}\}$, the multiplication between the fuzzy granule and the weight is represented as $w_s * g_s(x) = \{w_{s1}k_s(x, x_1), w_{s2}k_s(x, x_2), ..., w_{sn}k_s(x, x_n)\}$. Obviously, the weight w_s is a fuzzy granule on modality M_s.

Given a fuzzy granular vector $W = (w_1, w_2, .., w_m)$ represented the weight of multimodal features, for a sample $x \in U$ and a fuzzy granular vector $V_P(x) = (g_1(x), g_2(x), ..., g_m(x))$ computed with multimodal features, a feature fusion is a dot product between W and $V_P(x)$, which is defined as:

$$W \bullet V_P(x) = \sum_{j=1}^{m} w_j * g_j(x) = w_1 * g_1(x) + w_2 * g_2(x) + ... + w_m * g_m(x) = \sum_{j=1}^{m} w_j * g_j(x), \quad (40)$$

where the $g_j(x)$ is a fuzzy granule of x on modality j and the weight w_j is a fuzzy granule for modality j. A fuzzy granular vector is formed by some granules. For the sample x, the result of dot product between fuzzy granular vectors is a fuzzy granule. Therefore, the fusion of fuzzy granules is a new fuzzy granule.

4.3 Fuzzy granular classifiers for multimodal data

Kernel functions are used to extract features from multimodal attributes for constructing fuzzy granules. Fuzzy granular learning is proposed for handling multimodal features by combining fuzzy granules. The fusion of features can be interpreted as the combination of multimodal fuzzy granules from horizontal and vertical directions.

Given a training set $\{x_i, y_i\}_{i=1}^{n}$, where y_i is the label of sample x_i, the residual granule is:

$$g_e(x_i) = W \bullet V_P(x_i) - g_d(y_i) = \sum_{j=1}^{m} w_j * g_j(x_i) - g_d(y_i). \tag{41}$$

Then, we propose a granular square loss function of multimodal learning, which is computed as:

$$Loss(x_i) = \underset{w}{argmin} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} w_j * g_j(x_i) - g_d(y_i)\right)^2. \tag{42}$$

As for a test sample x, the classification function of multimodal learning is computed as:

$$f(x) = \underset{c}{argmin} \left| \sum_{j=1}^{m} w_j * g_j(x) - g_d(y_c) \right|. \tag{43}$$

The key question is how to achieve the optimal granular weight. We can use a gradient descent method for solving the optimization problem. Hence, a gradient is calculated by the following theorem.

Theorem 1. Suppose a cost function is $J(W) = \frac{1}{2}(g_e(x))^2$, then the difference of $J(W)$ with respect to the shared-weight vector W is: $(g_e(x)*g_1(x), g_e(x)*g_2(x), ..., g_e(x)*g_m(x))$.

Proof. According to the definition of residual granule, we know $g_e(x) = \sum_{j=1}^{m} w_j * g_j(x) - g_d(y)$. Since $W = \{w_1, w_2, ..., w_m\}$, so the difference of $g_e(x)$ with respect to W is $(g_1(x), g_2(x), ..., g_m(x))$. Therefore, the difference of $J(W))$ with respect to W is calculated as follows: $\frac{\nabla J(W)}{\nabla W} = (\frac{\nabla J(W)}{\nabla w_1}, \frac{\nabla J(W)}{\nabla w_2}, ..., \frac{\nabla J(W)}{\nabla w_m}) = (\frac{\nabla J(W)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_1}, \frac{\nabla J(W)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_2}, ..., \frac{\nabla J(W)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_m}) = (\frac{\nabla(1/2(g_e(x))^2)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_1}, \frac{\nabla(1/2(g_e(x))^2)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_2}, ..., \frac{\nabla(1/2(g_e(x))^2)}{\nabla g_e(x)} * \frac{\nabla g_e(x)}{\nabla w_m}) = (g_e(x) * \frac{\nabla(w_1 * g_1(x) - g_d(y)))}{\nabla w_1}, g_e(x) * \frac{\nabla(w_2 * g_2(x) - g_d(y))}{\nabla w_2}, ..., g_e(x) * \frac{\nabla(w_m * g_m(x) - g_d(y)))}{\nabla w_m}) = (g_e(x) * g_1(x), g_e(x) * g_2(x), ..., g_e(x) * g_m(x))$. The theorem is proved.

Algorithm 1 Training for the multimodal fuzzy granular learning (MFGL).

Input: A training set $MDS = (U, MC, \{d\})$, where $m = |MC|$; n is the number of local granular classifiers; the number of iterations t and a learning rate r.
Output: The n local granular classifiers. The weight vectors for these classifiers are $\{W_1, W_2, ..., W_n\}$, where $W_i = (w_1, w_2, ..., w_m)$.
1: Randomly select part of samples and features for forming n training sample sets;
2: For $i = 1$ to n parallel loop in the steps (3)-(12):
3: Achieve a training set U_i, where $|U_i| = N_i$;
4: Perform (5) - (12) steps for each training sample $x \in U_i$:
5: Granulate the training sample by a kernel function on each modality, inducing fuzzy granules $g_j(x)$;
6: Construct a fuzzy granular vector $V_P(x) = (g_1(x), g_2(x), ..., g_m(x))$ of x;
7: Perform the fuzzy granulation on the label, inducing a fuzzy decision granule $g_d(y)$;
8: Start an iteration process in steps (9)-(12):
9: The residual granule is obtained by the formula: $g_e(x) = W \bullet V_P(x) - g_d(y) = \sum_{j=1}^{m} w_j * g_j(x) - g_d(y)$;
10: Modify the granular weight vector according to Theorem 1: $w_j = w_j - r * g_e(x) * g_j(x)$;
11: Error accumulation: $e = e + |g_e(x)|$;
12: If the error ratio e/e_{max} converges or reaches the maximal iter number, then the iteration halts;
13: Output the n local granular classifiers.

Algorithm 2 Classification for the multimodal fuzzy granular learning (MFGC).

Input: The n local granular classifiers, a test sample t.
Output: A label of the test sample t.
For $i = 1$ to n parallel loop in the steps (2)-(5):
2: The test sample is fuzzily granulated as $g_j(t)$ in a local granular classifier on the j modality;
Construct a fuzzy granular vector $V_P(t) = (g_1(t), g_2(t), ..., g_m(t))$ of t;
4: Perform the fuzzy granulation on different labels, inducing several fuzzy decision granules: $g_d(y_c)$;
A new fuzzy granule can be induced by a forward computing process, and compare it with the fuzzy decision granules, then achieve the classification label: $f(t) = \underset{c}{argmin} \left| \sum_{j=1}^{m} w_j * g_j(t) - g_d(y_c) \right|$;
6: Apply a voting method to get the predict label of the sample t;
Output the label of the test sample t.

4.4 Local granulation and multi-classifier integration

In order to improve the efficiency, we use a local fuzzy granulation method. In the granulation process, some samples are randomly selected from a training set, and some features are randomly selected from multimodal features. Then, some local granular classifiers are constructed. Finally, these classifiers are integrated. The strategy of integration classification adopts the method of voting. These local granular classifiers are independent, so they can be granulated and calculated in parallel. In the classification process, a test sample is sent to all local granular classifiers for granulation and classification. Finally, the voting strategy is adopted to determine the label of the test sample.

4.5 The learning algorithm for multimodal data

The principle of fuzzy granular classifiers is presented in the above subsection. Therefore, we can design local granular classifiers based on a multimodal granular representation. These local granular classifiers can be calculated in parallel, which are divided into multimodal fuzzy granular learning (training) (MFGL) and multimodal fuzzy granular classification (MFGC). The specific algorithms are designed as follows.

From the detailed description, we know that a training algorithm is an iterative process for constructing many local classifiers. The testing process involves voting with these local classifiers to obtain the predict label. Furthermore, the training and testing algorithms can be carried out using parallel computing.

5 Experimental analysis

The effectiveness of Multimodal Fuzzy Granular (MFG) method is demonstrated through some experiments. First, we illustrate the convergence of MFG on some UCI data sets. Then, we compare MFG with other traditional algorithms on these data sets. Furthermore, aiming to reveal its multimodal capacity we design some multimodal classification jobs running on multimodal data sets.

We can notice that for normal UCI data sets, each feature is considered as a modality. Therefore, we use the single kernel and randomly select some of the features to perform fuzzy granulation tasks. However, for UCI multimodal data sets, we use the multi-kernel for granulating a set of features as a modality. Here, the set of features is considered as a single attribute.

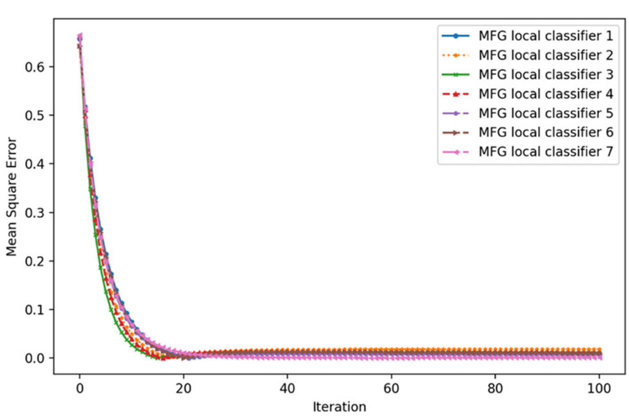


Fig. 3 MSE of different classifiers with linear kernels on Wine data set

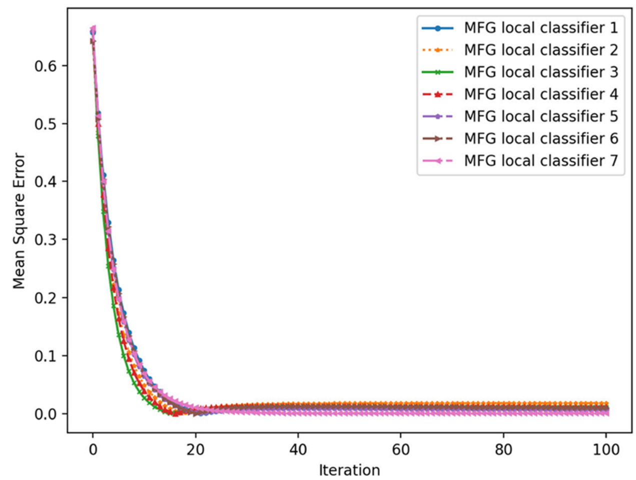


Fig. 4 MSE of different classifiers with Gaussian kernels on Heart data set

5.1 Convergence analysis

In this subsection, we use a single kernel function on the Wine data set, and a multi-kernel function on the Heart data set. All the samples from both data sets are employed to train and detect whether the classifier iterations converge or not. The Mean Square Error (MSE) is adopted to evaluate the convergence. It is expressed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2. \tag{44}$$

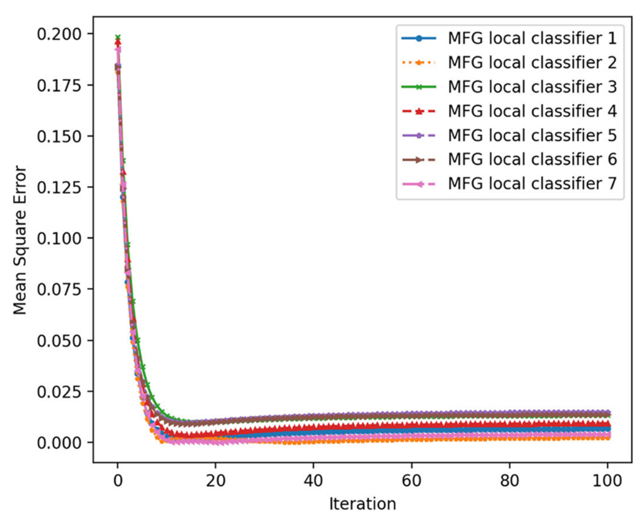


Fig. 5 MSE of different classifiers with linear kernels on Heart data set

Table 1 Description and classification accuracy of three normal UCI data sets (Mean%±Std%)

Data sets	F	N	C	KNN	SVM_L	SVM_G	MLP	MFG
Wine	13	178	3	95.34±3.29	95.56±2.54	95.00±3.89	94.78±2.10	**96.06±2.70**
Glass	9	214	7	68.14±5.91	60.70±1.06	68.30±8.93	64.65±8.86	**69.30±4.75**
Breast Cancer	30	569	2	96.57±1.38	96.75±1.30	93.77±2.86	**97.46±1.49**	95.03±2.65

The bold entries in the table signify that the current algorithm achieved a higher accuracy rate on this dataset compared to other algorithms

The fuzzy granulation randomly selects part of samples and attributes for constructing local MFG classifiers. In this experiment, we randomly select seven times to form seven classifiers on each data set, which are labeled as local MFG classifiers 1-7 in Figs. 3, 4, and 5. Figure 3 shows the iterative trend of MSE of different classifiers using linear kernels on the Wine data set. Figures 4-5 illustrate the iterative trend of MSE of different classifiers on Heart data set using Gaussian kernels and linear kernels respectively.

The experimental results are shown in Figs. 3, 4, and 5. Both of these figures demonstrate that the MSE of local classifiers for each data set can reach the lowest point in less than 20 iterations and then rises slightly. As the number of iterations continues to increase, it is eventually approaches zero. So, we can draw a conclusion that the MFG shows a fast convergence.

5.2 Normal UCI data sets

In the following experiment, we evaluate the classification performance of our proposed method on three normal UCI data sets. These data sets are listed in Table 1. The column F represents the number of features; the column N represents the number of samples; and the column C represents the number of labels.

In this subsection, we use different samples and features to construct fuzzy granules using the single kernel function which is reflexive and symmetrical. In order to test the performance of our proposed method on UCI data sets, we use the average accuracy of randomly selecting different samples and features. We employ classical classification algorithms such as KNN, SVM (SVM_L, SVM_G) as comparisons. SVM_L means using a linear kernel and SVM_G means using a Gaussian kernel. Additionally, we also compare with the classical neural network classification MLP (using 2-layers multilayer perceptron). The results with 10-fold cross validation are listed in Table 1.

In Table 1, the predictive results of our proposed method outperform classical classifiers such as KNN, SVM_L and SVM_G on Wine and Glass data sets, but are slightly inferior on Breast Cancer data set. but The same result applies to the MLP classifier.

As the local classifiers are constructed by randomly selecting features, we also test the influence of feature selection on classification, which is illustrated in Figs. 6, 7, and 8. From Figs. 6, 7, and 8, we can observe that when the number of selected features reaches approximately half of the total number of features, the classification accuracy reaches its highest point. Additional features will not only potentially decrease accuracy but also result in an increase in time complexity.

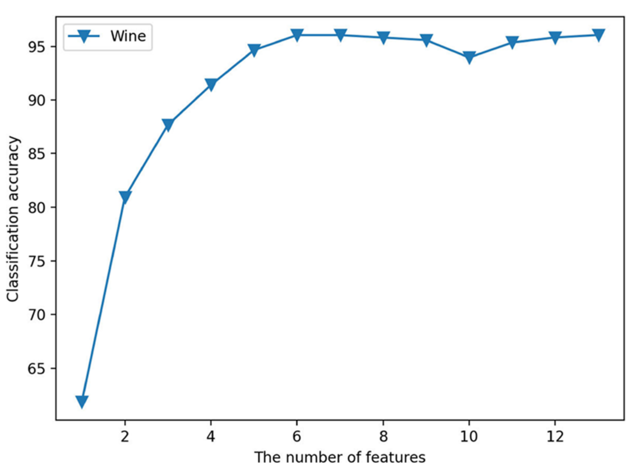


Fig. 6 Classification accuracy of randomly selecting different features for Wine data set

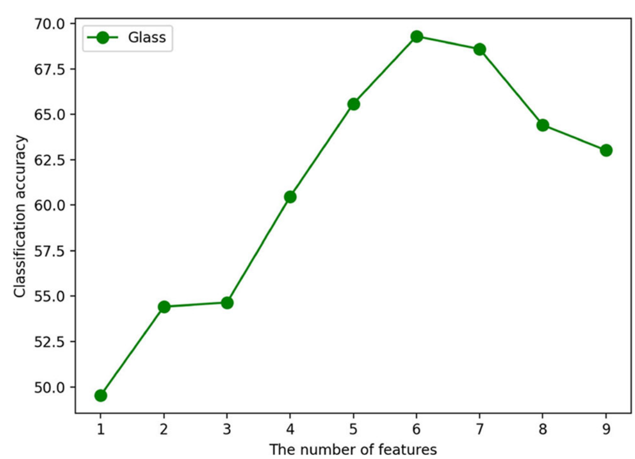


Fig. 7 Classification accuracy of randomly selecting different features for Glass data set

Table 2 Description of five multimodal UCI data sets

Data sets	Categorical	Numerical	N	C
Credit	9	6	690	2
Heart	7	6	270	2
Horse	15	8	368	2
Bands	16	19	512	2
Hepatitis	13	6	155	2

Therefore, we recommend choosing around half of the total number of features but not all features for classification tasks when using our proposed method.

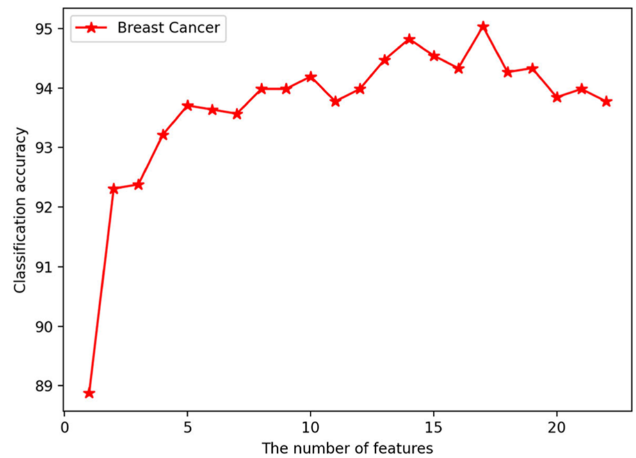


Fig. 8 Classification accuracy of randomly selecting different features for Breast Cancer data set

5.3 Multimodal UCI data sets

Some heterogeneous UCI data sets include both categorical and numerical features that can be regarded as multimodal attributes. In the following experiments, we compare the effectiveness of classical classifiers with our proposed method on five multimodal data sets that are listed in Table 2. The column $Categorical$ represents the number of categorical features and the column $Numerical$ represents the number of numerical features. The column N represents the number of samples; and the column C indicates the number of labels.

In this subsection, to handle numerical attributes, we introduce the Gaussian and linear kernel functions. As for categorical attributes, we use the match kernel function. To compare with our proposed algorithms, we employ some classical classifiers, such as KNN, Bayes, SVM (SVM_L, SVM_G) [56] and a classical neural network MLP. Certainly, we also compared the K-nearest neighbors (KNN) algorithm based on granular vector distance [57], including the K-nearest neighbor classifier based on the relative granule distance (KNGR), and the K-nearest neighbor classifier based on the absolute distance (KNGA). The experimental results are shown in Table 3. In these experiments, the SVM_L is an SVM classifier with a linear kernel, while the SVM_G means an SVM classifier with a Gaussian kernel. The MFG_{GM} is our proposed algorithm using Gaussian and match kernels, and the MFG_{LM} means using linear and match kernels.

From the result of Table 3, we can see that our proposed classifier (MFG) shows an excellent classification performance. The performance of MFG is better than traditional classifiers, the neural network classifier (MLP) and the K-nearest neighbors (KNN) algorithm based on granular vector distance (KNGR, KNGA) on Credit, Horse and Hepatitis data sets. It is only slightly inferior to the KNGR on Heart and Bands data sets.

In order to analyze the influence of the Gaussian kernel on classification performance, we choose random samples and set the parameter interval in [0.1, 0.9] with the step of

Table 3 Description and classification accuracy of five multimodal data sets (Mean%±Std%)

Algorithm	Credit	Heart	Horse	Bands	Hepatitis
KNN	83.43±2.27	71.50±4.26	65.67±3.67	72.33±2.76	84.20±3.46
Bayes	80.63±2.23	84.12±5.18	76.02±4.32	62.33±3.67	80.6±6.11
SVM_L	85.43±2.99	80.07±6.15	72.05±3.51	70.56±3.43	83.87±3.22
SVM_G	86.17±3.03	81.11±6.02	73.11±3.78	72.31±4.85	86.40±4.09
MLP	82.39±9.20	79.63±9.33	68.33±5.82	72.78±5.41	85.48±5.25
KNGR	84.70±4.46	**86.39±15.86**	66.55±8.52	**79.63±10.08**	83.50±15.89
KNGA	80.69±6.60	85.71±15.15	70.90±6.65	78.52±7.60	82.79±16.73
MFG_{G+M}	**87.54±3.09**	81.32±5.20	**76.67±4.07**	70.8±4.04	86.00±3.00
MFG_{L+M}	86.68±3.47	84.50±4.01	72.45±4.69	70.13±4.30	**88.00±3.40**

The bold entries in the table signify that the current algorithm achieved a higher accuracy rate on this dataset compared to other algorithms

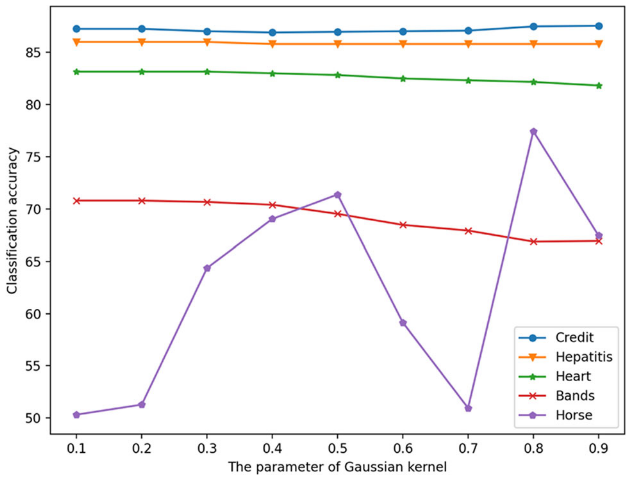


Fig. 9 Classification accuracy of MFG with Gaussian kernels on different data sets

0.1. Figure 9 illustrates classification accuracy of MFG with Gaussian kernels on different data sets. In Fig. 9, it shows that the MFG exhibits higher accuracy with the Gaussian kernel parameter in the interval [0.1, 0.5] for most data sets.

Since the MFG is an ensemble classifier, we will further discuss the performance influence on the number of local classifiers. The number of local classifiers is changed from 3 to 9 for all the test data sets. The Gaussian kernel is used to process numerical features, while the linear kernel is used to tackle numerical features. Figures 10 and 14 illustrate that the classification accuracy is the highest when the number of classifiers is 5 for Credit and Bands. As for Heart, Hepatitis and Horse, Figs. 11, 12, and 13 show that the classification accuracy is the highest when the number of classifiers is 7.

Considering the influence of local classifiers, we recommend the number of classifiers to be set $KC + 1$, where C is the number of labels and K is a natural integer. We also suggest that it is best to start K at 2. As for the training process, we recommend that the product of the number of classifiers and the number of randomly selected samples be greater than the total number of samples.

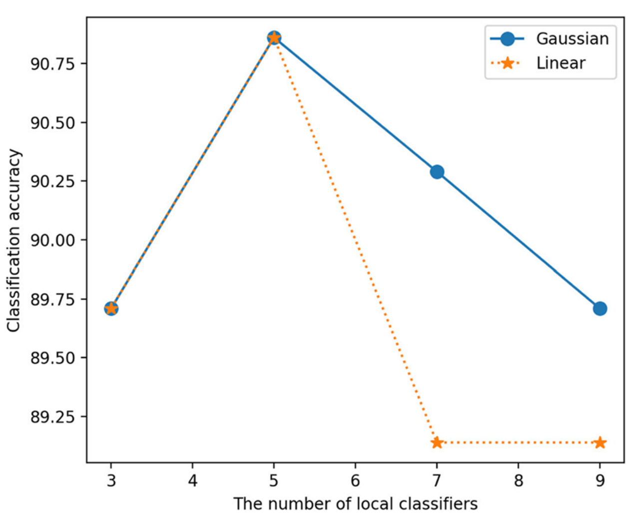


Fig. 10 Classification accuracy of MFG with the number of classifiers for Credit

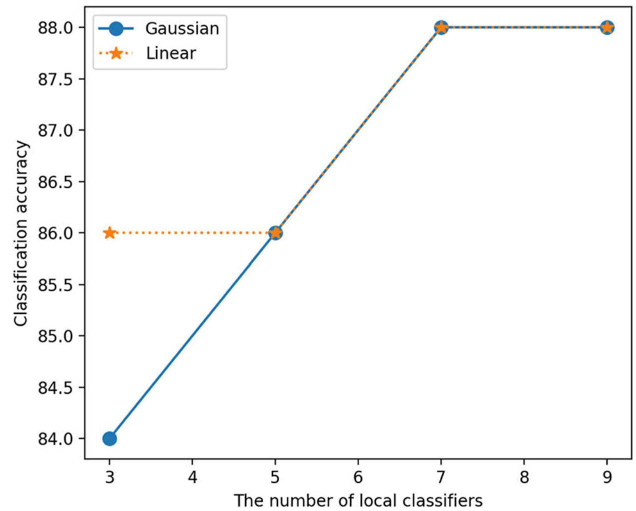


Fig. 11 Classification accuracy of MFG with the number of classifiers for Heart

6 Conclusion

The traditional classifier is based on numerical calculation, which does not involve an operation of sets. In this paper, a kind of set classifier for multimodal data, named Multimodal Fuzzy Granular (MFG) classifier, is proposed, which applies a fuzzy granular representation and a fuzzy granula-

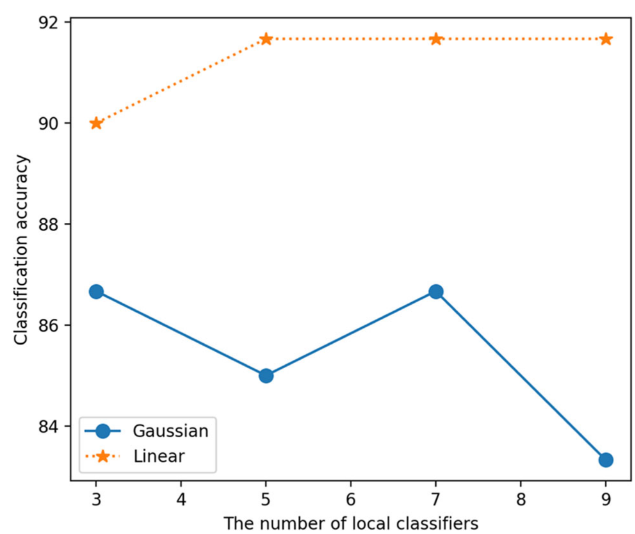


Fig. 12 Classification accuracy of MFG with the number of classifiers for Hepatitis

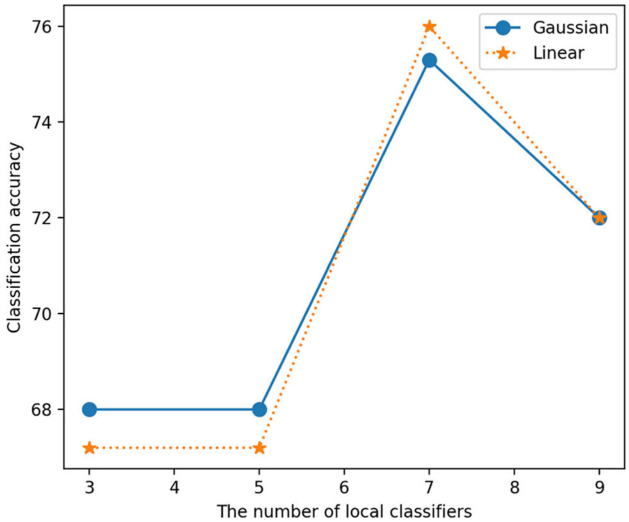


Fig. 13 Classification accuracy of MFG with the number of classifiers for Horse

tion. Firstly, the kernel function is introduced to the process of fuzzy granulation, which induces fuzzy granules that are essentially fuzzy sets. Furthermore, fuzzy granular vectors are constructed by combining fuzzy granules in a classification system. We also discuss the size measurement of fuzzy granules and granular vectors. The operators of fuzzy granules and fuzzy granular vectors are presented, and we fuse the fuzzy granules to construct local granular classifiers. Finally, a multimodal fuzzy classifier is designed by integrating local classifiers. The experimental results show that the new multimodal granular classifier has the characteristics of fast convergence and good classification performance. In the future work, the neural network method will be introduced to adjust the weights of granules. The hierarchical structure of

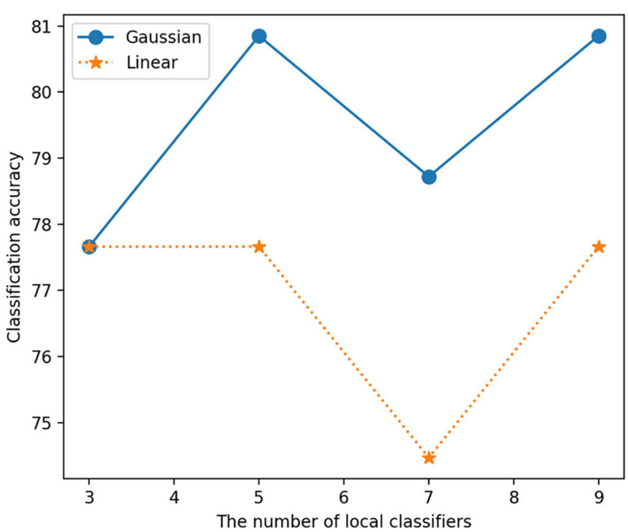


Fig. 14 Classification accuracy of MFG with the number of classifiers for Bands

neural networks will be also considered, and a fuzzy granular classifier based on deep learning will be emerged.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61976183) and the Research Cooperation Project of Industry and University in Fujian Province (No. 2020H6101).

Data Availability The datasets generated during and/or analysed during the currentstudy are available in the [UCI] repository, [https://archive.ics.uci.edu/ml/index.php].

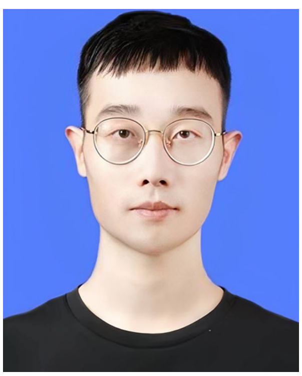
References

1. Ramachandram D, Taylor GW (2017) Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process Mag 34(6):96–108
2. Atrey PK, Hossain MA, El Saddik A, Kankanhalli MS (2010) Multimodal fusion for multimedia analysis: a survey. Multimed Syst 16:345–379
3. Shi Y, Suk HI, Gao Y, Lee SW, Shen D (2019) Leveraging coupled interaction for multimodal alzheimer's disease diagnosis. IEEE Trans. Neural Netw Learn Syst 31(1):186–200
4. Baltrušaitis T, Ahuja C, Morency LP (2018) Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence 41(2):423–443
5. Ullah H, Ullah B, Wu L, Abdalla FY, Ren G, Zhao Y (2020) Multi-modality medical images fusion based on local-features fuzzy sets and novel sum-modified-laplacian in non-subsampled shearlet transform domain. Biomed Signal Process Control 57:1–14
6. Zhou T, Ruan S, Canu S (2019) A review: Deep learning for medical image segmentation using multi-modality fusion. Array 3:1–11
7. Wang P, Cai D (2020) Multiple kernel learning by empirical target kernel. International Journal of Wavelets, Multiresolution and Information Processing 18(02):1950058
8. Zhang X, Mahoor MH (2016) Task-dependent multi-task multiple kernel learning for facial action unit detection. Pattern Recognit 51:187–196
9. Sultan R, Tamimi H, Ashhab Y (2019) Improving classification performance using genetic programming to evolve string kernels. Int Arab J Inf Technol 16(3):454–459
10. Maji S, Berg AC, Malik J (2008) Classification using intersection kernel support vector machines is efficient. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8
11. Jiang H, Shen D, Ching WK, Qiu Y (2022) A high-order norm-product regularized multiple kernel learning framework for kernel optimization. Inf Sci 606:72–91
12. Xue H, Song Y, Xu HM (2019) Multiple indefinite kernel learning for feature selection. Knowl Based Syst 191
13. Xue H, Song Y (2020) Non-convex approximation based l 0-norm multiple indefinite kernel feature selection. Appl Intell 50:192–202
14. Liao Z, Gao L, Zhou T, Fan X, Zhang Y, Wu J (2019) An oil painters recognition method based on cluster multiple kernel learning algorithm. IEEE Access 7:26842–26854
15. Yang M, Wang Z, Li Y, Zhou Y, Li D, Du W (2022) Gravitation balanced multiple kernel learning for imbalanced classification. Neural Comput Appl 34(16):13807–13823
16. Chi Z, Wang Z, Wang B, Fang Z, Zhu Z, Li D, Du W (2022) Multiple kernel subspace learning for clustering and classification. IEEE Transactions on Knowledge and Data Engineering
17. Zadeh LA (1979) Fuzzy sets and information granular. In: Fuzzy Set Theory and Applications, pp. 3–18
18. Kerk YW, Tay KM, Lim CP (2019) Monotone interval fuzzy inference systems. IEEE Trans Fuzzy Syst 27(11):2255–2264

19. Tian H, Wang X, Lu Z, Huang Y, Tian G (2017) Adaptive fuzzy logic energy management strategy based on reasonable soc reference curve for online control of plug-in hybrid electric city bus. IEEE Trans Intell Transp Syst 19(5):1607–1617
20. Xiao Z, Gong Z, Liu K (2020) Fuzzy linear systems with the two-dimension fuzzy data. J Intell Fuzzy Syst 38(4):5295–5315
21. Ngo LT, Dang TH, Pedrycz W (2018) Towards interval-valued fuzzy set-based collaborative fuzzy clustering algorithms. Pattern Recognit 81:404–416
22. Vatankhah Ghadim H, Tarafdar Hagh M, Ghassem Zadeh S (2022) Fermat-curve based fuzzy inference system for the fuzzy logic controller performance optimization in load frequency control application. Fuzzy Optim Decis Mak1–32
23. Salimi-Badr A, Ebadzadeh MM (2020) A novel self-organizing fuzzy neural network to learn and mimic habitual sequential tasks. IEEE Trans Cybern 52(1):323–332
24. Lang G, Miao D, Fujita H (2019) Three-way group conflict analysis based on pythagorean fuzzy set theory. IEEE Transactions on Fuzzy Systems 28(3):447–461
25. Singh PK (2019) Bipolar fuzzy concept learning using next neighbor and euclidean distance. Soft Computing 23(12):4503–4520
26. Zadeh LA (1996) Fuzzy logic= computing with words. IEEE Trans Fuzzy Syst 4(2):103–111
27. Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90(2):111–127
28. Zadeh LA (2001) A new direction in ai: Toward a computational theory of perceptions. AI magazine 22(1):73–73
29. Pedrycz W, Izakian H (2014) Cluster-centric fuzzy modeling. IEEE Trans Fuzzy Syst 22(6):1585–1597
30. Al-Hmouz R, Pedrycz W, Balamash A, Morfeq A (2015) Description and classification of granular time series. Soft Computing 19:1003–1017
31. Pedrycz W, Park BJ, Oh SK (2008) The design of granular classifiers: a study in the synergy of interval calculus and fuzzy sets in pattern recognition. Pattern Recognit 41(12):3720–3735
32. Roh SB, Pedrycz W, Ahn TC (2014) A design of granular fuzzy classifier. Expert Syst Appl 41(15):6786–6795
33. Yang T, Zhong X, Lang G, Qian Y, Dai J (2020) Granular matrix: A new approach for granular structure reduction and redundancy evaluation. IEEE Trans Fuzzy Syst 28(12):3133–3144
34. Singh PK (2019) Vague concept lattice reduction using granular computing and vague entropy. Math Comput Simul 165:56–73
35. Tan A, Wu WZ, Qian Y, Liang J, Chen J, Li J (2018) Intuitionistic fuzzy rough set-based granular structures and attribute subset selection. IEEE Trans Fuzzy Syst 27(3):527–539
36. Yu P, Li J, Wang H, Lin G (2019) Matrix-based approaches for updating approximations in multigranulation rough set while adding and deleting attributes. Int J Comput Intell Syst 12(2):855–872
37. Sun L, Wang L, Ding W, Qian Y, Xu J (2020) Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets. IEEE Trans Fuzzy Syst 29(1):19–33
38. Fujita H, Gaeta A, Loia V, Orciuoli F (2018) Resilience analysis of critical infrastructures: a cognitive approach based on granular computing. IEEE Trans Cybern 49(5):1835–1848
39. Fujita H, Gaeta A, Loia V, Orciuoli F (2019) Hypotheses analysis and assessment in counterterrorism activities: a method based on owa and fuzzy probabilistic rough sets. IEEE Trans Fuzzy Syst 28(5):831–845
40. Qian J, Liu C, Yue X (2019) Multigranulation sequential three-way decisions based on multiple thresholds. Int J Approx Reason 105:396–416
41. Chen Y, Miao D (2020) Granular regression with a gradient descent method. Inf Sci 537:246–260
42. Qian J, Hong C, Yu Y, Liu C, Miao D (2022) Generalized multi-granulation sequential three-way decision models for hierarchical classification. Inf Sci 616:66–87
43. Niu J, Chen D, Li J, Wang H (2022) A dynamic rule-based classification model via granular computing. Inf Sci 584:325–341
44. Shen Y, Pedrycz W, Wang X (2018) Clustering homogeneous granular data: Formation and evaluation. IEEE Trans Cybern 49(4):1391–1402
45. Chen Y, Qin N, Li W, Xu F (2019) Granule structures, distances and measures in neighborhood systems. Knowl Based Syst 165:268–281
46. Yue X, Zhou J, Yao Y, Miao D (2020) Shadowed neighborhoods based on fuzzy rough transformation for three-way classification. IEEE Trans Fuzzy Syst 28(5):978–991
47. Zadeh LA (1965) Fuzzy sets. Inf. Control 8(3):338–353
48. Pal SK, Dutta-Majumder DK (1986) Fuzzy Mathematical Approach to Pattern Recognition
49. Bede B, Bede B (2013) Fuzzy Analysis. Springer
50. Smola AJ, Schölkopf B (1998) Learning with Kernels vol. 4, 639–646. Springer Berlin Heidelberg
51. Tata S, Patel JM (2007) Estimating the selectivity of tf-idf based cosine similarity predicates. ACM Sigmod Record 36(2):7-12
52. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. Computer Ence, 1–12
53. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543
54. Spiegel MR (1975) Theory and problems of probability and statistics. ESAIM - Probab Stat 114–115
55. Lu L, Huang B, Zhang Q, Ke D, Xu Y (2012) Research on algorithm of combing lda-based discriminant classifier and mfcc feature extraction for pure acoustic listening similarity. Int J Adv Comput Technol 4(5):106–113
56. Jana DK, Bhunia P, Adhikary SD, Mishra A (2023) Analyzing of salient features and classification of wine type based on quality through various neural network and support vector machine classifiers. Results Control Optim 11:100219
57. Chen Y, Qin N, Li W, Xu F (2019) Granule structures, distances and measures in neighborhood systems. Knowl Based Syst 165:268–281

Fenggang Han received the M.E. degree in internal combustion engine from Harbin Engineering University, China.He is currently a Professor with the College of Mechanical and Automotive Engineering, Xiamen University of Technology. His primary research areas encompass the safety and lightweighting of new energy vehicles, along with intelligent chassis adaptive control systems.

Xiao Zhang received the B.S. degree in information and computing science from Shanghai Maritime University, China, in 2018. He also received the M.E. degree from the College of Computer and Information Engineering, Xiamen University of Technology, China, in 2022. His research interests include machine learning, natural language processing, and granular computing.



Linjie He received the B.S. degree in software engineering from Xiamen University of Technology, China. He is currently pursuing the M.E. degree with the College of Computer and Information Engineering, Xiamen University of Technology. His research interests include artificial intelligence, machine learning, computer vision.

Liru Kong received the B.S. degree in network engineering from Zaozhuang University, China, in 2021. She is currently pursuing the M.E. degree with the College of Computer and Information Engineering, Xiamen University of Technology. Her research interests include machine learning, artificial intelligence and granular computing.

Yumin Chen received the M.E. degree in computer application technology from Nanchang University, China, in 2005, and the Ph.D. degree in pattern recognition and intelligence system from Tongji University, China, in 2010. From November 2018 to November 2019, he was a Visiting Scholar with the Department of Mathematics and Statistics, Mississippi State University, USA. He was supported by the Education Department of Fujian Province. He is currently a Professor with the College of Computer and Information Engineering, Xiamen University of Technology. He has published over 50 articles in refereed journals and conferences. His research interests include machine learning, rough sets, and granular computing.