A Novel Fusion and Feature Selection Framework for Multisource Time-Series Data Based on Information Entropy

Xiuwei Chen, Li Lai, and Maokang Luo

Abstract—**Information technology growth brings vast time-series data. Despite richness, challenges like redundancy emphasize the need for time-series data fusion research. Rough set theory, a valuable tool for dealing with uncertainty, can identify features and reduce dimensionality, enhancing time-series data fusion. The contribution of the study lies in establishing a fusion and feature selection framework for multisource time-series data. This framework selects optimal information sources by minimizing entropy. In addition, the fusion process integrates a feature selection algorithm to eliminate redundant features, preventing a sequential increase in entropy. Crucial experiments on abundant datasets demonstrate that the proposed approach outperforms several state-of-the-art algorithms in terms of enhancing the accuracy of common classifiers. This research significantly advances the field of time-series data fusion in rough set theory, offering improved accuracy and efficiency in data processing and analysis.**

Index Terms—**Conditional entropy, feature selection, information fusion, multisource time-series data.**

I. Introduction

IN RECENT years, the rapid advancement of technology has led to the proliferation of data from various sources, including time-series data. Time-series data, which consist of a sequence of observations collected over time, are widely encountered in diverse real-life applications such as finance [1], medical [2], and anomaly detection [3]. The analysis of multisource time-series data has emerged as a crucial research area due to its potential to extract valuable insights and enhance decision-making processes. Information fusion techniques are vital in integrating data from multiple sources and extracting meaningful knowledge [4], [5]. In the field of machine learning, numerous time-series fusion methods have been proposed [6], [7]. However, these neural network methods lack interpretability. Therefore, this article combines the mathematical tools of granular computing and information theory to introduce a new fusion and feature selection framework for handling uncertainties in multisource time-series data.

Granular computing is a powerful framework that tackles uncertainty across diverse domains [8]. It operates by manipulating information at varying levels of granularity, recognizing the inherent imprecision of data. Using granules as an abstraction tool, objects sharing similar properties are grouped, enabling more effective data analysis and decision-making. This approach allows for a more nuanced understanding of complex datasets, providing valuable insights and improving the overall reliability of results. Rough set theory, pioneered by Pawlak [9] is a prominent technique within the framework of granular computing. It offers a mathematical framework specifically designed to handle imprecise and uncertain data. The central focus of rough set theory lies in approximation, aiming to extract meaningful knowledge from imprecise information. Its practicality extends across diverse fields, including knowledge discovery [10], [11], feature selection [12], [13], and uncertainty measurement [14], [15]. By leveraging the principles of rough set theory, the application of this approach enables effective analysis of complex datasets, informed decision-making, pattern identification, and enhanced performance of machine learning algorithms.

Information entropy, originating from information theory, quantifies the uncertainty in a dataset [16]. It offers a numerical measure of the information content and uncertainty inherent in the data. Within the framework of rough set theory, information entropy plays a crucial role in evaluating the importance of features or sources within the dataset. The uncertainty in multisource data can be effectively addressed through the integration of these two theories. In recent years, there has been a proliferation of fusion methods based on the integration of rough set theory and information entropy [17], [18], [19], [20], [21], [22], [23]. For fuzzy incomplete data, Xu et al. [19] proposed a fusion framework by utilizing the entropy to select the essential source with respect to each feature. Later in 2023, they established a novel incremental fusion approach for interval-valued ordered data [18]. An evaluative metric based on the fuzzy dominance entropy has been defined to select information sources. Regarding incomplete interval-valued data, Zhang et al. [20] proposed an innovative fusion approach that utilizes tolerance relations. In addition, tolerance classes are employed to complete missing values. The previously mentioned approaches primarily concentrate on selecting optimal sources without thoroughly filtering features. Consequently, the combined outcomes might still retain redundant features, implying that uncertainty is not necessarily minimized at this juncture. Thus, for multilabel systems,

Received 6 August 2024; revised 8 February 2025 and 27 February 2025; accepted 2 March 2025. *(Corresponding author: Li Lai.)*

The authors are with the School of Mathematics, Sichuan University, Chengdu 610064, China (e-mail: xiuweichen1998@163.com; laili@scu.edu.cn; makaluo@scu.edu.cn).

Digital Object Identifier 10.1109/TNNLS.2025.3548165

Qian et al. [21] introduced a fusion framework and leveraged the positive region for feature selection. Subsequently, focusing on homogeneous data, Zhang et al. [17] introduced an unsupervised fusion framework to choose sources based on maximum granulation or minimum entropy, while employing a novel mRMR framework to condense features in the fusion results. Nevertheless, none of these methodologies are designed to handle time-series data. Hence, exploring the fusion of multisource time-series decision systems (MsTsDS) is imperative.

The motivation behind this study stems from three main factors. First, the aforementioned fusion methods [17], [18], [19], [20], [21], [22], [23] do not directly accommodate time-series data fusion, prompting the need for research on fusion methods for MsTsDS. Second, these fusion techniques address each feature independently during fusion, lacking feature selection phrases. Consequently, the resulting output may still harbor redundant features, increasing the risk of local optima entrapment. Third, numerous feature selection algorithms [24], [25], [26], [27], [28] are designed for single information systems, overlooking the possibility of alternative features originating from multiple sensors or sources, which are unsuitable for multisource environments. Therefore, this study introduces a novel fusion and feature selection framework for MsTsDS based on information entropy. An innovative trend relation is proposed to extract both numerical and trend information from time-series samples. In addition, a novel conditional entropy measure is defined to assess the uncertainty of the decision label in relation to feature sets. The essential source can be selected to form the fusion results by utilizing the proposed entropy measure. Moreover, a feature selection optimization algorithm is employed to optimize the fusion process to eliminate redundant features. If a feature fails to decrease decision uncertainty, it is discarded as redundant. The primary contributions and innovations are outlined as follows.

1) For MsTsDS, the neighborhood, gradient neighborhood, and signed equivalence relation between any two samples are established. Based on the proposed three relations, a novel trend relation is defined to extract the numerical and trend information between time-series data.
2) According to the trend relation, the condition entropy is introduced to qualify the uncertainty of the decision label with respect to feature sets. A novel fusion framework is established for MsTsDS based on the proposed entropy. The essential source can be selected by minimizing the entropy. The method, denoted as TEF, can be observed in Fig. 1.
3) To eliminate redundant features, the overall impact of features is considered during the fusion process, and a selection algorithm is employed to optimize the fusion process. Features that do not contribute to reducing the overall uncertainty of the fusion results are removed. This method, denoted as GTEF, can be visually represented in Fig. 2.
4) A series of experiments demonstrates that the proposed fusion methods significantly improve the classification accuracy compared to several state-of-the-art methods in terms of the KNN, Random Forest, and ExNRule Ensemble classifiers.

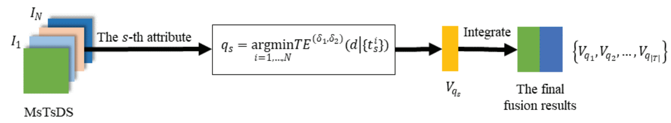

Fig. 1. Framework of TEF. The fusion process can be divided into two steps. First, the conditional entropies of all information sources for each feature are calculated. Subsequently, the information source with the minimum entropy to perform fusion can be selected, resulting in a new information table.

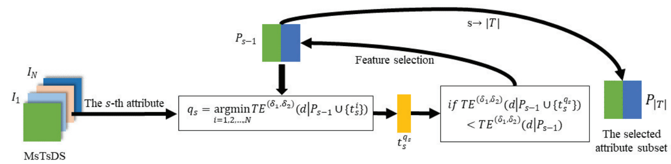

Fig. 2. Framework of GTEF. The essential source is first selected for each feature based on conditional entropy. Then, if the inclusion of the information source successfully reduces uncertainty, the feature associated with that information source is considered suitable for inclusion in the fusion result. Conversely, if the information source fails to reduce uncertainty, the corresponding feature is excluded from the fusion process.

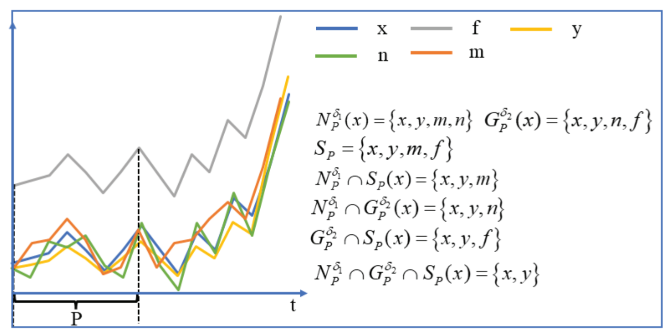

Fig. 3. Illustration of the trend relation. Take sample x as an example. The classic neighborhood relation can only differentiate far-apart samples, such as f. The single gradient neighborhood relation can only distinguish samples with significant gradient differences, such as m, while the single signed equivalence relation can only differentiate samples with opposite trends, such as n. However, the combination of these three can successfully identify samples that are close in distance and have similar trends, that is, y.

The rest of this article is organized as follows. Section II introduces the foundational concepts employed in this study, namely equivalence relation, neighborhood relation, and MsTsDS. Section III defines the neighborhood, gradient, and signed equivalence relation to extract the trend information between samples. Subsequently, the conditional entropy is proposed based on the trend relation, and a fusion framework is established by leveraging the conditional entropy. Section IV optimizes the fusion process and eliminates redundant features through a feature selection algorithm. Section V comprehensively evaluates the effectiveness of the proposed methods in enhancing the performance of three classifiers. Essential experiments are conducted on abundant public datasets to compare the proposed and state-of-the-art algorithms. Finally, Section VI encompasses the conclusion and analysis of the limitations encountered during this study. In addition, it provides an outlook on potential future research directions.

II. Preliminaries

In this section, we introduce some basic concepts and definitions of equivalence relation, neighborhood relation, and MsTsDS.

A. Classical Rough Set

Given $IS = (U, A, V, f)$ be an information system, where $U = \{x_1, \ldots, x_n\}$ is the sample set, $A = \{a_1, \ldots, a_m\}$ is the feature set, V is the domain of A, and $f : U \times A \to V$ is the information function of A, furthermore, a decision system (DS) can be denoted as $\mathrm{DS} = IS \cup \{d, V_d, f_d\}$, where d denotes the decision label of DS, V_d denotes the domain of d, and $f_d : U \times \{d\} \to V_d$ denotes the information function of d.

Definition 1: Given a DS $\mathrm{DS} = (U, A, V, f) \cup \{d, V_d, f_d\}$, for any $P \subseteq A$, the equivalence relation can be defined as [9]

$$E_P = \{(x, y) \in U \times U \,|f(x, a) = f(y, a), \forall \mathrm{a} \in \mathrm{P}\} \tag{1}$$

where $f(x, a)$ and $f(y, a)$ denote the values of x and y under a, respectively. Furthermore, the equivalence class of any $x \in U$ can be computed by $E_P(x) = \{y \in U | (x, y) \in E_P\}$. The equivalence relation based on the decision label d can be represented as $E_d = \{(x, y) \in U \times U | f_d(x, d) = f_d(y, d)\}$, where $f_d(x, d)$ and $f_d(y, d)$ denote the value of x and y under d, respectively. Furthermore, the decision partition of U can be computed as $U/d = \{E_d(x) | x \in U\}$, where $E_d(x) = \{y \in U | (x, y) \in E_d\}$ denotes the equivalence class of x with respect to the decision label d.

B. Neighborhood Rough Set

The above equivalence relation applies to categorical variables and is too strict for numerical variables. For numerical variables, the following neighborhood relation is commonly used.

Definition 2: Given a DS $\mathrm{DS} = (U, A, V, f) \cup \{d, V_d, f_d\}$, for any $P \subseteq A$, the neighborhood relation is defined as [29]

$$N_{P}^{\sigma} = \{(x, y) \in U \times U \,| d_P(x, y) \le \sigma\} \tag{2}$$

where $\sigma \in [0, 1]$ is a threshold and $d_P(x, y)$ denotes the distance between x and y. Furthermore, the neighborhood class of any $x \in U$ can be computed by $N_{P}^{\sigma}(x) = \{y \in U | (x, y) \in N_{P}^{\sigma}\}$.

C. Multisource Time-Series DS

Building upon the definition of IS, a time-series information system (TsIS) can be denoted as $\mathrm{TsIS} = (U, T, V, f)$, where $T = \{t_1, \ldots, t_m\}$ is the feature set (t_m stands for the moment m). Furthermore, a time-series DS (TsDS) can be denoted as $\mathrm{TsDS} = \mathrm{TsIS} \cup \{d, V_d, f_d\}$, where d denotes the decision label, V_d denotes the domain of d, and $f_d{:}U \times \{d\} \to V_d$ denotes the information function of d.

Moreover, a multisource time-series information system (MsTsIS) can be denoted as

$$\mathrm{MsTsIS} = \{I_i \,| I_i = (U, T, V_i, f_i), i = 1, \ldots, N\}$$

where I_i is the ith subsystem of MsTsIS. Furthermore, an MsTsDS can be defined as

$$\mathrm{MsTsDS} = \mathrm{MsTsIS} \cup \{d, V_d, f_d\} \tag{3}$$

where d denotes the decision label of DS, V_d denotes the domain of d, and $f_d{:}U \times \{d\} \to V_d$ denotes the information function of d.

Example 1: Suppose that ten patients require electrocardiogram (ECG) monitoring to assess their cardiac health status. Relying solely on data from a single ECG device may not comprehensively evaluate the patients' cardiac conditions. Therefore, to enhance the reliability of diagnostic assessments, four different ECG devices are utilized to measure the cardiac electrical signals of these patients. Each patient's ECG signals are recorded at eight time points. The domain of the decision label is $\{-1, 1\}$, where -1 and 1 denote the Ischemia and Normal, respectively. The ECG signals from the four devices are presented in Table I.

III. Information Fusion Of MsTsDS Based on Conditional Entropy

Multisource time-series data are highly prevalent in practical life, and investigating the fusion of such data holds great significance. Information entropy serves as a measure of uncertainty within the data and can be utilized to quantify the importance of information sources. This section proposes an information fusion framework for MsTsDS based on conditional entropy.

Next, the neighborhood relation is defined first to extract numerical information from time-series data.

Definition 3: Given $\mathrm{TsIS} = (U, T, V, f)$ be a TsIS, for any $P \subseteq T$, the neighborhood relation between $x \in U$ and $y \in U$ is defined as

$$N_P^{\delta} = \left\{(x, y) \in U \times U \,\middle|\, \frac{d_P(x, y)}{\max\{d_P(x, z) \,| z \in U\}} \le \delta\right\} \tag{4}$$

where $\delta \in [0, 1]$ is a threshold and $d_P(x, y) = \left(\sum_{t \in P} (f(x, t) - f(y, t))^2\right)^{(1/2)}$. Moreover, the neighborhood class of any $x \in U$ can be computed as $Nr_P^{\delta}(x) = \{y \in U | (x, y) \in Nr_P^{\delta}\}$.

Although the neighborhood relation defined above can extract numerical information from time-series data, it can only differentiate far-apart samples. Thus, in terms of trend information, the following gradient neighborhood and single signed equivalence relation are defined to distinguish samples with significant gradient differences and opposite trends.

Definition 4: Given $\mathrm{TsIS} = (U, T, V, f)$ be a TsIS, for any feature $t_s \in T$, the gradient of any $x \in U$ is defined as

$$g(x, t_s) = \begin{cases} \dfrac{1}{2}\left(f\left(x, t_{s+1}\right) - f\left(x, t_{s-1}\right)\right), & 1 < s < |T| \\ f\left(x, t_{s+1}\right) - f\left(x, t_s\right), & \text{else} \end{cases} \tag{5}$$

where $f(x, t_s)$ denotes the value of x under the feature t_s.

Definition 5: Given $\mathrm{TsIS} = (U, T, V, f)$ be a TsIS, for any $t \in T$, the gradient neighborhood relation between $x \in U$ and $y \in U$ is defined as

$$G_P^{\delta} = \left\{(x, y) \in U \times U \,\middle|\, \frac{g_P(x, y)}{\max\{g_P(x, z) \,| z \in U\}} \le \delta\right\} \tag{6}$$

TABLE I
ECG SIGNALS FROM FOUR DEVICES

U	IS1								IS2								IS3								IS4								d
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	
x_1	0.5	-0.26	-1.23	-1.84	-0.51	0.66	0.66	0.88	0.49	-0.25	-1.24	-1.84	-0.5	0.67	0.65	0.87	0.45	-0.29	-1.29	-1.84	-0.51	0.67	0.66	0.87	0.6	-0.25	-1.24	-1.84	-0.52	0.66	0.67	0.88	-1
x_2	0.15	-0.69	-1.04	0.57	1.08	0.56	-0.18	-0.46	0.15	-0.65	-1.02	0.59	1.03	0.57	-0.19	-0.46	0.16	-0.75	-1.02	0.54	1.02	0.56	-0.19	-0.47	0.14	-0.68	-1.04	0.56	1.08	0.56	-0.18	-0.47	1
x_3	0.32	-0.12	-1.06	-1.94	-0.13	0.74	0.71	0.72	0.3	-0.12	-1.03	-1.92	-0.15	0.79	0.77	0.75	0.4	-0.12	-1.13	-1.9	-0.15	0.72	0.77	0.71	0.5	-0.13	-1.14	-1.9	-0.19	0.72	0.79	0.71	-1
x_4	1.17	0.74	-0.46	-1.71	-0.72	0.44	0.26	0.2	1.14	0.71	-0.51	-1.69	-0.71	0.4	0.31	0.2	1.15	0.72	-0.52	-1.68	-0.7	0.4	0.3	0.2	1.1	0.7	-0.5	-1.69	-0.79	0.4	0.4	0.21	-1
x_5	0.65	-0.37	-1.35	-1.01	0.78	0.54	0.47	-0.15	0.7	-0.39	-1.3	-1.1	0.78	0.54	0.49	-0.15	0.71	-0.38	-1.3	-1.2	0.79	0.54	0.43	-0.15	0.7	-0.37	-1.3	-1.21	0.79	0.55	0.43	-0.14	1
x_6	0.4	-0.12	-1.96	-0.77	0.65	0.17	0.49	0.42	0.5	-0.15	-1.93	-0.79	0.69	0.17	0.52	0.45	0.4	-0.14	-1.99	-0.78	0.69	0.18	0.52	0.41	0.41	-0.15	-1.97	-0.77	0.68	0.16	0.51	0.42	1
x_7	1.21	-0.37	-1.02	-1.53	-0.93	0.57	0.63	0.94	1.25	-0.47	-1.05	-1.55	-0.94	0.56	0.64	0.99	1.2	-0.36	-1.01	-1.54	-0.94	0.59	0.64	0.95	1.21	-0.37	-1	-1.5	-0.99	0.57	0.6	0.95	-1
x_8	0.6	-0.31	-1.18	-1.87	0.26	0.52	0.6	0.35	0.8	-0.41	-1.19	-1.87	0.36	0.62	0.6	0.4	0.61	-0.32	-1.18	-1.86	0.26	0.51	0.6	0.36	0.71	-0.32	-1.2	-1.86	0.26	0.5	0.7	0.3	-1
x_9	0.33	-0.28	-1.7	-0.51	0.79	0.64	0.69	0.31	0.43	-0.29	-1.7	-0.51	0.87	0.64	0.79	0.31	0.32	-0.28	-1.71	-0.51	0.78	0.64	0.68	0.31	0.33	-0.28	-1.72	-0.51	0.79	0.64	0.65	0.3	1
x_{10}	1.67	-0.45	-1.6	-0.49	0.69	0.4	0.08	-0.01	1.77	-0.55	-1.62	-0.59	0.69	0.5	0.08	-0.09	1.66	-0.45	-1.61	-0.49	0.68	0.4	0.07	-0.01	1.67	-0.45	-1.6	-0.49	0.68	0.5	0.07	-0.05	1

where $\delta \in [0,1]$ is a threshold and $g_P(x,y) = \left(\sum_{t\in P}(g(x,t)-g(y,t))^2\right)^{(1/2)}$. Moreover, the gradient neighborhood class of any $x \in U$ can be computed as $G_P^{\delta}(x) = \{y \in U | (x,y) \in G_P^{\delta}\}$.

Definition 6: Given TsIS $= (U,T,V,f)$ be a TsIS, for any $P \subseteq T$, the signed equivalence relation between $x \in U$ and $y \in U$ is defined as

$$S_P = \left\{(x,y) \,\middle|\, \mathrm{sign}\,(g\,(x,t)) = \mathrm{sign}\,(g\,(y,t)\,,) \ \ \forall t \in P\right\} \quad (7)$$

where sign(x) denotes the sign function. Moreover, the signed equivalence class of any $x \in U$ can be computed as $S_P(x) = \{y \in U | (x,y) \in S_P\}$.

The proposed neighborhood relation $N_P^{\delta_1}$ can only differentiate far-apart samples. The single gradient neighborhood relation $G_P^{\delta_2}$ can only distinguish samples with significant gradient differences, while the single signed equivalence relation S_P can only differentiate samples with opposite trends. However, combining these three can successfully identify samples that are close in distance and have similar trends. The trend relation is defined as follows.

Definition 7: Given TsIS $= (U,T,V,f)$ be a TsIS, for any $t \in T$, the trend relation between $x \in U$ and $y \in U$ is defined as

$$T_P^{(\delta_1,\delta_2)} = N_P^{\delta_1} \cap G_P^{\delta_2} \cap S_P \quad (8)$$

where δ_1 and δ_2 denote the thresholds of the neighborhood and gradient neighborhood relations, respectively. Moreover, the trend class of any $x \in U$ can be computed as $T_P^{(\delta_1,\delta_2)}(x) = \{y \in U | (x,y) \in T_P^{(\delta_1,\delta_2)}\}$.

Fig. 3 intuitively exemplifies the benefits of capturing trend details from the time-series samples. Also, the properties of the trend relation are as follows.

Property 1: Given TsDS $=$ TsIS $\cup \{d, V_d, f_d\}$ be a TsIS, where TsIS $= (U,T,V,f)$, for any $\delta_1, \delta_2 \in [0,1]$ and $P \subseteq T$, the trend relation satisfies the following.

1) $T_P^{(\delta_1,\delta_2)}(x) = N_P^{\delta_1}(x) \cap G_P^{\delta_2}(x) \cap S_P(x)$.
2) *Reflexivity:* $(x,x) \in T_P^{(\delta_1,\delta_2)}$.
3) *Monotonicity:* $T_P^{(\delta_1,\delta_2)} \subseteq T_P^{(\delta_1^*,\delta_2^*)}$, if $\delta_1 \le \delta_1^*$ or $\delta_2 \le \delta_2^*$.

Next, to measure the information source's uncertainty, the following conditional entropy is defined by the defined trend relation.

Definition 8: Given TsDS $=$ TsIS $\cup \{d, V_d, f_d\}$ be a TsIS, where TsIS $= (U,T,V,f)$ and $U/d = \{Y_1, \ldots, Y_l\}$, for any $P \subseteq T$ and $\delta_1, \delta_2 \in [0,1]$, the conditional entropy of d with respect to P is defined as

$$TE^{(\delta_1,\delta_2)}(d\,|P) = -\sum_{x\in U}\sum_{Y\in U/d} \frac{\left|T_P^{(\delta_1,\delta_2)}(x)\cap Y\right|}{|U|} \ln \frac{\left|T_P^{(\delta_1,\delta_2)}(x)\cap Y\right|}{\left|T_P^{(\delta_1,\delta_2)}(x)\right|}. \quad (9)$$

Property 2: Given TsDS $=$ TsIS $\cup \{d, V_d, f_d\}$ be a TsIS, where TsIS $= (U,T,V,f)$, for any $\delta_1, \delta_2 \in [0,1]$ and $P, Q \subseteq T$, the conditional entropy satisfies the following.

1) $0 \le TE^{(\delta_1,\delta_2)}(d|P) \le |U| \ln |U|$.
2) $TE^{(\delta_1,\delta_2)}(d|P) \le TE^{(\delta_1,\delta_2)}(d|Q)$, if $T_P^{(\delta_1,\delta_2)}(x) \subseteq T_Q^{(\delta_1,\delta_2)}(x), \forall x \in U$.

Proof: The following conditions hold.

1) It is obvious that the function $f(x,y) = -x\ln(x/x+y)$ is increasing with respect to x and y. Any term of entropy can be rewritten as $|TE^{(\delta_1,\delta_2)}(x) \cap Y| \ln(|TE^{(\delta_1,\delta_2)}(x) \cap Y| / |TE^{(\delta_1,\delta_2)}(x) \cap Y| + |TE^{(\delta_1,\delta_2)}(x) \cap Y^C|)$. Thus, when $TE^{(\delta_1,\delta_2)}(x) = \{x\}$, $\forall x \in U$, the conditional entropy takes the minimum value 0. When $TE^{(\delta_1,\delta_2)}(x) = U$, we have $TE^{(\delta_1,\delta_2)}(d|P) = -\sum_{x\in U}\sum_{Y\in U/d}(|Y|/|U|)\ln(|Y|/|U|) \le -\sum_{x\in U}\sum_{Y\in U/d}(1/|U|)\ln(1/|U|) = |U|\ln|U|$. Thus, we can get $0 \le TE^{(\delta_1,\delta_2)}(d|P) \le |U|\ln|U|$.
2) It is obvious to demonstrate based on the proof of 1).

The conditional entropy presented in Definition 8 can be utilized to measure the significance of information sources. A smaller value indicates lower uncertainty and higher importance of the information source. Therefore, we can select the information source with the minimum entropy as the fusion result for each feature.

Definition 9: Given an MsTsDS MsTsDS $=$ MsTsIS $\cup \{d, V_d, f_d\}$, where MsTsIS $= \{I_i | I_i = (U,T,V_i,f_i), i = 1,\ldots,N\}$, let $\{t_s^1, t_s^2, \ldots, t_s^N\}$ denote the set of feature t_s of N subsystems. For any $\delta_1, \delta_2 \in [0,1]$, the most important source I_{q_s} with respect to t_s can be computed as

$$q_s = \arg\min_{i=1,2,\ldots,N} TE^{(\delta_1,\delta_2)}(d\left|\{t_s^i\}\right.). \quad (10)$$

By Definition 9, the fusion process can be divided into two steps. First, the conditional entropies of all information sources for each feature are calculated. Subsequently, the information source with the minimum entropy to perform fusion can be selected, resulting in a new information table. The corresponding algorithm of the proposed TEF method is presented in Algorithm 1. Considering the worst case scenario, where $|U/d| = |U|$, the time complexity of Algorithm 1 is $O(|T| \times N \times |U|^2)$. An intuitive framework of the proposed TEF method can be seen in Fig. 1.

Example 2 (Continue From Example 1): Assume that

Algorithm 1 TEF

Input: A MsTsDS $MsTsDS = MsTsIS \cup \{d, V_d, f_d\}$; thresholds δ_1 and δ_2
Output: A fused feature set

$P \leftarrow \emptyset$;
for $s = 1 : |T|$ **do**
  for $i = 1 : N$ **do**
    $TE^{(\delta_1,\delta_2)}(d\,|\{t_s^i\}) \leftarrow 0$;
    for each $x \in U$ **do**
      for each $Y \in U/d$ **do**
        $TE^{(\delta_1,\delta_2)}(d\,|\{t_s^i\}) \leftarrow TE^{(\delta_1,\delta_2)}(d\,|\{t_s^i\}) - \frac{|T^{(\delta_1,\delta_2)}_{\{t_s^i\}}(x)\cap Y|}{|U|} \ln \frac{|T^{(\delta_1,\delta_2)}_{\{t_s^i\}}(x)\cap Y|}{|T^{(\delta_1,\delta_2)}_{\{t_s^i\}}(x)|}$;
      end for
    end for
  end for
  compute $q_s = \arg\min_{i=1,2,\ldots,N} TE^{(\delta_1,\delta_2)}(d\,|\{t_s^i\})$;
  $P \leftarrow P \cup \{t_s^{q_s}\}$;
end for
return: P

TABLE II
CONDITIONAL ENTROPY OF D WITH RESPECT TO EACH FEATURE

T	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
I_1	2.7831	3.6354	2.7732	1.7838	1.5658	1.8004	1.6819	1.7550
I_2	3.1489	3.3564	2.5823	1.8431	1.3156	1.6792	2.0184	2.1116
I_3	3.0673	3.7414	2.7846	1.9068	1.5658	1.7045	1.7342	2.1709
I_4	3.4364	3.8219	2.6277	1.8729	1.5911	1.8954	2.1570	2.3871

$\delta_1 = \delta_2 = 0.5$. Take t_1^1 as an example. First, the trend classes of each samples can be computed as follows: $T^{(0.5,0.5)}_{\{t_1^1\}}(x_1) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_2) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_3) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_6) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_8) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_9) = \{x_1, x_2, x_3, x_5, x_6, x_8, x_9\}$; $T^{(0.5,0.5)}_{\{t_1^1\}}(x_4) = \{x_4\}$; $T^{(0.5,0.5)}_{\{t_1^1\}}(x_5) = \{x_1, x_2, x_5, x_6, x_8, x_9\}$; and $T^{(0.5,0.5)}_{\{t_1^1\}}(x_7) = T^{(0.5,0.5)}_{\{t_1^1\}}(x_{10}) = \{x_7, x_{10}\}$. Later, the entropy can be calculated by formula (9): $TE^{(0.5,0.5)}(d|\{t_1^1\}) = -6 \times (2/10) \times \ln(2/7) - (1/10)\times\ln(1/6) - 6\times(5/10)\times\ln(5/7) - (5/10)\times\ln(5/6) = 2.7831$. Similarly, the conditional entropy of each feature can be computed, as presented in Table II. Later, for each feature, the essential source with the minimum entropy can be selected to consist of a new feature set: $\{t_1^1, t_2^2, t_3^2, t_4^1, t_5^2, t_6^2, t_7^1, t_8^1\}$.

IV. FEATURE SELECTION OF MSTSDS BASED ON CONDITIONAL ENTROPY

Algorithm 1 constructs the fusion result by selecting the information source with the minimum entropy for each feature, thereby retaining all features. However, it is not always the case that all features are helpful for decision-making. In order to eliminate redundant features, a feature selection algorithm is employed to optimize the fusion process and remove redundant features. By iteratively selecting features that maximize the reduction in overall uncertainty, the selection algorithm ensures a streamlined fusion process and enhances the effectiveness of the feature selection procedure.

Definition 10: Given an MsTsDS MsTsDS = MsTsIS $\cup$ $\{d, V_d, f_d\}$, where MsTsIS $= \{I_i | I_i = (U, T, V_i, f_i), i = 1, \ldots, N\}$, let $\{t_s^1, t_s^2, \ldots, t_s^N\}$ denote the set of feature t_s of N subsystems and P_{s-1} denote the feature subset selected in the step $s-1$. For any $\delta_1, \delta_2 \in [0, 1]$, the feature subset in the step s can be updated by

$$P_s = \begin{cases} P_{s-1} \cup \{t_s^{q_s}\}, & \text{if } TE^{(\delta_1,\delta_2)}\left(d\,\middle|P_{s-1} \cup \{t_s^{q_s}\}\right) \\ & < TE^{(\delta_1,\delta_2)}(d\,|P_{s-1}) \\ P_{s-1}, & \text{else} \end{cases} \quad (11)$$

where $q_s = \arg\min_{i=1,2,\ldots,N} TE^{(\delta_1,\delta_2)}(d|P_{s-1} \cup \{t_s^i\})$.

Algorithm 2 GTEF

Input: A MsTsDS $MsTsDS = MsTsIS \cup \{d, V_d, f_d\}$; threshold δ_1,δ_2
Output: A selected feature subset

$P_0 \leftarrow \emptyset$;
for $s = 1 : |T|$ **do**
  for $i = 1 : N$ **do**
    $TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^i\}) \leftarrow 0$;
    for each $x \in U$ **do**
      for each $Y \in U/d$ **do**
        $TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^i\}) \leftarrow TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^i\}) - \frac{|T^{(\delta_1,\delta_2)}_{P_{s-1}\cup\{t_s^i\}}(x)\cap Y|}{|U|} \ln \frac{|T^{(\delta_1,\delta_2)}_{P_{s-1}\cup\{t_s^i\}}(x)\cap Y|}{|T^{(\delta_1,\delta_2)}_{P_{s-1}\cup\{t_s^i\}}(x)|}$;
      end for
    end for
  end for
  compute $q_s = \arg\min_{i=1,2,\ldots,N} TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^i\})$
  if $TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^{q_s}\}) = 0$ **then**
    break;
  else
    if $TE^{(\delta_1,\delta_2)}(d\,|P_{s-1} \cup \{t_s^{q_s}\}) < TE^{(\delta_1,\delta_2)}(d\,|P_{s-1})$ **then**
      $P_s \leftarrow P_{s-1} \cup \{t_s^{q_s}\}$;
    else
      $P_s \leftarrow P_{s-1}$;
    end if
  end if
end for
return: $P_{|T|}$

It is important to clarify that the feature subset P_0 is initialized as an empty set before commencing the first step. Consequently, the term $TE^{(\delta_1,\delta_2)}(d|P_0)$ lacks significance and can be treated as an infinitely large positive number. According to Definition 10, the essential source is first selected for each feature based on conditional entropy. Then, if the inclusion of the information source successfully reduces uncertainty, the feature associated with that information source is considered suitable for inclusion in the fusion result. Conversely, if the information source fails to reduce uncertainty, the corresponding feature is excluded from the fusion process. The relevant algorithm of the proposed GTEF method is given in Algorithm 2. Taking into account the worst case scenario, where $|U/d| = |U|$, the algorithm's time complexity can be characterized as $O(|T| \times N \times |U|^2)$. An intuitive architecture of the proposed GTEF method is presented in Fig. 2.

Example 3 (Continue from Example 1): First, the feature subset P_0 is set to be an empty set. Then, in the first step, the entropy of t_1 with respect to four sources can be calculated based on Definition 8 as follows: $TE^{(0.5,0.5)}(d|\{t_1^1\}) = 2.7831$, $TE^{(0.5,0.5)}(d|\{t_1^2\}) = 3.1489$, $TE^{(0.5,0.5)}(d|\{t_1^3\}) = 3.0673$, and $TE^{(0.5,0.5)}(d|\{t_1^4\}) = 3.4364$. The first source has the minimum entropy, so the feature $\{t_1^1\}$ is selected and put into P_0. The feature subset P_0 is updated to $P_1 = \{t_1^1\}$. Next, for the feature t_2, we can compute the following four entropy: $TE^{(0.5,0.5)}(d|P_1 \cup \{t_2^1\}) = 2.9315$, $TE^{(0.5,0.5)}(d|P_1 \cup \{t_2^2\}) = 3.1806$, $TE^{(0.5,0.5)}(d|P_1 \cup \{t_2^3\}) = 3.0556$, and $TE^{(0.5,0.5)}(d|P_1 \cup \{t_2^4\}) = 3.0420$. The first source has the minimum entropy, but $2.9315 > 2.7831$. Therefore, the feature t_2 should be removed. The feature subset P_1 is updated to $P_2 = P_1$. Continuing iteratively until all features have been traversed, the selected feature subset $P_8 = \{t_1^1, t_3^1, t_4^1, t_6^1, t_7^1, t_8^1\}$ can be obtained.

V. Experimental Analysis

This section conducts a series of experiments to validate the effectiveness and efficiency of the proposed methods. All algorithms are executed on a computer with an Intel[1] Xeon1 W-2123 at 3.60 GHz CPU and 64-GB RAM. The proposed methods are compared with ten state-of-the-art information fusion and feature selection techniques as follows.

1) *MF [30]:* $\overline{f}(x, t_s) = (1/N) \sum_{i=1}^{N} f_i(x, t_s)$, where $f_i(x, t_s)$ denotes the value of x with respect to t_s under the ith subsystem.
2) *Neighborhood Entropy Fusion (NEF and NEF-FS) [17]:* This method consists of two frameworks: NEF for fusion and NEF-FS for reduction. NEF defines a novel rough entropy and selects the optimal information source using the Sup-Inf function. After obtaining the fusion result, NEF-FS then utilizes the mRMR framework to select features. The parameter δ is set to 0.5 in the following experiments.
3) *FDEF [18]:* This method defines the fuzzy dominating conditional entropy to select the information source with the minimum conditional entropy. The threshold k is set to 1 in the subsequent experiments.
4) *TIEF [20]:* This framework introduces a novel tolerance relation and utilizes this relation to define a new conditional entropy to select the best source. The threshold α is set to 0.5 in the upcoming experiments.
5) *FIEF [19]:* This method defines a novel fuzzy conditional entropy to select the information source with the minimum entropy. The parameter L_a is assigned to 0.5 in the following experiments.
6) *SDF and SDF-FS [21]:* This method involves two frameworks: SDF for fusion and SDF-FS for reduction. SDF defines the significance of information sources to select the optimal source, while SDF-FS uses the positive region to select features after fusion.

[1]Registered trademark.

TABLE III
Details of Simulated Datasets

No.	Datasets	Size	Sources	Length	Classes
1	BME	180	3	128	3
2	Fungi	204	3	201	18
3	Chinatown	365	3	24	2
4	GunPointAgeSpan	451	3	150	2
5	SonyAIBORobotSurface1	621	3	70	2
6	CBF	930	3	128	3
7	MelbournePedestrian	3457	3	24	10
8	TwoPatterns	5000	3	128	4
9	Wafer	7164	3	152	2
10	Lightning2	121	3	637	2
11	Earthquakes	461	3	512	2
12	LargeKitchenAppliances	750	3	720	3

7) *KND-UFS [31]:* This method employs FCM clustering technique along with k-nearest neighbor rough set for feature selection. The parameter α is set from 0.5 to 1, and k is set from 0.05 to 0.5, both with increments of 0.05.
8) *Feature Selection Based on Conditional Mutual Information (KNCMI) [32]:* This method introduces a novel conditional mutual information and a feature objective evaluation function for effective feature selection. The parameter δ is configured explicitly to 0.5.

A. Performance Measurements

We conduct tests on the KNN (with k defaulting to 3), Random Forest (with estimators defaulting to 3), and ExNRule Ensemble [33], [34] (with both k and r defaulting to 3). Tenfold cross validation is adopted, and the average accuracy and variance are used as comparative metrics. Each method's average accuracy and average ranking are also included in the table. In addition, the Wilcoxon signed-rank and Nemenyi tests are employed to assess the mean accuracy. Significance levels α are set at 0.1, 0.05, 0.01, and 0.001. The proposed method significantly outperforms other algorithms if the P-value is below the significance level.

B. Performance on Simulated Datasets

To simulate the noise effects of different sensors or sources, this study employs the simulation methods proposed in literature [30] and [35] to generate MsTsDS. The datasets used in the experiments are obtained from the public database Time Series Machine Learning Website.[2] In this study, three subsystems are generated for each dataset. The details of the simulated datasets are shown in Table III. Without loss of generality, the parameters δ_1 and δ_2 are set to 0.1 in the following experiments. Tenfold cross validation is adopted, and the comparison results are presented in Tables IV–VI. The complexities of feature T in NEF-FS and KNCMI are $|T|^2$ and $|T|^3$, respectively. To enhance the efficiency for large datasets with big samples and features, we employ a stride of eight for feature subsampling. Results are annotated with underlines for clarity.

[2]https://www.timeseriesclassification.com

TABLE IV

COMPARISON OF ACCURACY WITH SIMULATED DATASETS BASED ON THE KNN. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets	GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
Fungi	12.2±4.5	14.7±6.2	8.3±6.1	8.8±5.6	8.8±5.6	9.8±5.7	8.8±5.6	8.3±6.1	11.3±5.5	4.9±3.8	11.2±5.8	**16.2±9.0**	8.8±5.6	9.3±5.0
BME	**66.7±14.3**	49.4±12.5	32.8±10.1	50.6±12.8	48.9±11.9	48.9±11.3	48.3±11.9	34.4±11.3	49.4±9.1	62.2±10.8	52.8±15.0	64.4±12.2	48.9±11.9	49.4±12.5
SonyAIBORobotSurface1	53.6±6.3	51.0±6.0	50.3±5.2	51.7±6.1	50.7±5.3	50.9±6.0	50.7±5.8	49.6±3.7	**55.1±4.7**	53.1±5.3	49.9±5.4	53.0±6.4	51.0±5.5	51.4±5.8
Wafer	**89.4±0.7**	**89.4±0.7**	86.9±0.7	87.0±0.6	87.0±0.7	87.1±0.7	87.0±0.7	86.9±0.8	87.3±0.9	87.0±0.7	86.8±0.9	87.2±1.2	87.0±0.7	87.0±0.7
TwoPatterns	**27.3±2.0**	26.5±1.7	26.1±2.3	26.2±2.4	26.2±1.5	26.4±2.7	26.3±2.3	25.8±2.3	26.7±2.4	24.4±1.5	25.2±1.5	27.2±1.6	26.5±1.6	26.5±1.6
CBF	33.0±5.4	32.9±3.0	32.7±4.3	31.6±3.4	31.8±3.8	31.1±3.6	31.7±3.3	32.8±5.5	**37.4±5.4**	33.0±4.1	32.5±4.1	32.7±3.9	31.9±3.6	32.8±5.5
MelbournePedestrian	17.6±2.2	**17.7±2.2**	14.2±1.4	14.3±1.5	14.4±1.4	14.2±1.4	14.3±1.5	14.2±1.4	15.6±1.5	16.5±1.2	14.1±1.7	15.2±1.4	14.3±1.4	14.4±1.4
Chinatown	**70.2±8.6**	67.7±10.7	61.4±5.7	60.9±5.5	61.4±5.7	61.4±5.7	61.1±5.7	61.4±5.7	62.5±7.8	64.7±11.3	66.0±8.7	64.2±7.3	61.4±5.7	61.4±5.7
GunPointAgeSpan	**57.6±5.5**	53.2±6.4	49.9±4.1	49.7±4.3	50.3±4.2	50.1±4.4	49.9±4.1	49.9±4.1	**57.6±5.7**	51.4±6.4	54.5±6.5	52.3±7.0	49.9±4.1	49.9±4.1
Lightning2	57.0±13.4	62.1±12.4	57.9±10.9	62.8±11.9	57.9±11.3	60.3±9.7	62.0±8.4	58.7±10.3	54.4±13.3	63.7±16.7	**66.9±10.0**	61.2±7.9	62.1±13.3	63.7±11.8
Earthquakes	**79.9±6.0**	**79.9±6.0**	72.5±5.7	74.0±5.1	74.0±4.5	73.8±5.2	74.2±5.2	71.8±4.7	76.8±7.4	74.9±5.2	77.5±4.7	78.8±6.0	73.5±5.4	73.5±5.4
LargeKitchenAppliances	**38.8±6.2**	37.9±4.4	34.7±4.9	37.6±2.7	36.7±3.4	36.9±4.0	37.2±3.6	34.8±4.6	38.3±8.0	34.5±6.9	38.7±4.8	37.9±4.9	36.7±3.2	36.8±3.8
Average	**50.3**	48.5	44.0	46.3	45.7	45.9	46.0	44.1	47.7	47.5	48.0	49.2	46.0	46.3
Average ranking	**12.4**	11.2	3.4	6.2	5.6	6.0	5.5	3.5	10.6	8.3	7.9	10.8	6.0	7.6
Nemenyi Pvalues	GTEF		**<0.001**	**<0.05**	**<0.01**	**<0.05**	**<0.01**	**<0.001**	1.00	0.47	0.32	1.00	**<0.05**	0.21
	TEF		**<0.001**	0.16	**<0.1**	0.13	**<0.05**	**<0.001**	1.00	0.91	0.81	1.00	0.13	0.69
Wilcoxon Pvalues	GTEF		**<0.001**	**<0.01**	**<0.001**	**<0.01**	**<0.01**	**<0.01**	**<0.05**	**<0.05**	**<0.05**	**<0.1**	**<0.01**	**<0.01**
	TEF		**<0.001**	**<0.05**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.32	**<0.1**	**0.19**	0.29	**<0.01**	**<0.05**

TABLE V

COMPARISON OF ACCURACY WITH SIMULATED DATASETS BASED ON THE RANDOM FOREST. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets	GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
Fungi	16.7±10.2	**27.9±11.4**	12.3±6.7	14.2±7.7	16.7±8.1	11.3±6.3	14.2±7.7	13.2±6.1	13.6±6.6	7.8±6.6	16.2±6.3	7.9±5.5	14.7±7.0	18.6±8.5
BME	**72.8±10.7**	68.9±14.9	62.2±9.9	64.4±13.7	58.3±13.4	64.4±12.2	65.6±12.6	65.0±13.4	51.7±9.0	69.4±13.9	62.2±10.2	60.6±14.8	63.3±10.3	65.0±13.4
SonyAIBORobotSurface1	56.2±4.5	**58.0±7.9**	54.3±5.5	52.5±6.3	55.7±5.5	55.7±3.0	54.1±6.7	52.5±4.6	56.0±4.8	53.1±4.3	51.4±3.7	54.6±6.0	55.7±6.3	56.4±3.8
Wafer	**89.4±0.7**	**89.4±0.7**	89.3±0.7	89.3±0.7	89.3±0.7	89.3±0.7	89.3±0.7	89.3±0.7	**89.4±0.7**	**89.4±0.7**	89.3±0.7	89.3±0.7	89.3±0.7	**89.4±0.7**
TwoPatterns	**26.9±1.6**	26.3±1.5	25.1±2.5	23.8±1.7	25.3±2.0	24.3±1.3	25.5±1.7	23.7±1.9	25.9±1.7	25.3±1.6	25.7±1.2	25.4±1.7	25.2±2.0	25.4±1.8
CBF	36.1±6.3	**37.1±4.2**	34.5±4.0	33.1±5.7	30.8±5.8	32.3±4.1	30.5±5.5	34.8±2.4	32.9±4.2	30.9±4.1	35.8±6.8	31.4±3.7	32.7±6.5	34.8±2.4
MelbournePedestrian	**26.1±2.9**	25.9±2.9	23.1±2.5	23.6±2.4	23.6±2.4	23.3±2.3	23.3±2.3	23.1±2.5	21.9±2.0	15.7±2.4	21.0±3.5	23.1±2.5	23.1±2.5	23.6±2.4
Chinatown	**71.3±8.5**	70.8±8.1	66.1±8.4	65.5±7.6	68.0±9.6	67.2±10.4	66.1±9.3	66.1±8.4	68.5±9.8	67.4±7.5	66.6±6.8	67.7±8.3	66.1±8.4	67.7±9.9
GunPointAgeSpan	**61.9±4.4**	60.3±5.0	57.4±4.6	57.7±5.4	55.4±6.1	53.2±5.6	57.2±5.6	57.4±5.2	55.5±8.2	58.3±5.1	55.4±6.8	55.4±6.0	56.6±4.8	57.4±5.2
Lightning2	60.3±12.4	62.8±6.6	58.8±11.2	56.0±15.6	61.2±16.0	57.9±15.9	60.4±10.8	61.2±15.2	55.3±15.2	59.6±9.2	57.9±9.2	52.8±6.9	61.2±15.2	**63.7±12.9**
Earthquakes	**80.3±4.6**	79.9±6.0	78.5±5.2	79.0±5.7	78.8±6.2	78.6±6.8	79.0±6.1	79.2±5.6	78.5±5.2	78.8±6.5	78.6±6.0	79.2±5.8	79.2±5.6	79.2±5.6
LargeKitchenAppliances	**46.0±4.7**	45.5±5.4	37.3±4.3	40.7±4.1	42.3±4.9	39.7±3.9	43.2±4.1	39.7±6.2	37.3±5.0	38.5±6.1	44.1±3.7	37.6±5.7	40.4±5.6	42.3±3.4
Averge	53.7	**54.4**	49.9	50.0	50.5	49.8	50.7	50.4	48.9	49.5	50.4	48.8	50.6	52.0
Average ranking	12.9	**13.1**	5.0	6.2	7.3	5.2	7.2	6.5	6.2	6.4	6.2	5.2	7.0	10.8
Nemenyi Pvalues	GTEF		**<0.001**	**<0.01**	**<0.1**	**<0.001**	**<0.1**	**<0.05**	**<0.01**	**<0.05**	**<0.01**	**<0.001**	**<0.05**	1.00
	TEF		**<0.001**	**<0.01**	**<0.05**	**<0.001**	**<0.05**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.001**	**<0.05**	0.99
Wilcoxon Pvalues	GTEF		**<0.001**	**<0.001**	**<0.01**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.05**
	TEF		**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.01**

TABLE VI

COMPARISON OF ACCURACY WITH SIMULATED DATASETS BASED ON THE EXNRULE ENSEMBLE. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets	GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
Fungi	14.2±6.5	**21.1±11.2**	11.3±5.6	11.8±8.1	9.2±4.5	13.7±7.6	12.8±8.9	12.2±5.8	8.8±6.9	9.8±5.3	17.6±7.7	12.2±5.8	14.3±5.7	15.6±7.0
BME	**72.8±11.2**	60.0±14.9	30.0±12.7	55.6±12.9	52.8±12.0	44.4±11.1	53.9±11.7	45.6±19.7	39.4±8.0	62.8±10.0	59.4±13.6	65.0±16.9	46.7±15.2	56.1±9.4
SonyAIBORobotSurface1	57.2±5.8	**57.3±5.4**	53.5±3.9	51.7±7.0	54.1±2.8	53.0±4.8	51.2±5.5	52.7±5.5	51.7±5.6	54.3±8.4	53.0±8.6	49.3±8.2	53.0±3.7	53.2±6.2
Wafer	**89.4±0.7**	**89.4±0.7**	89.1±0.7	89.0±0.7	89.1±0.8	89.2±0.8	89.1±0.8	89.1±0.7	89.2±0.7	87.6±0.7	89.0±0.7	89.1±0.8	89.1±0.8	89.1±0.7
TwoPatterns	26.8±1.5	**26.9±1.7**	24.9±1.9	26.9±1.1	26.1±1.9	26.3±2.1	26.5±1.5	25.9±1.9	25.7±1.5	24.6±1.3	25.5±1.0	25.7±1.7	25.6±0.9	26.5±2.2
CBF	35.1±4.8	**36.0±7.2**	28.5±4.1	30.0±4.4	32.5±4.2	36.1±4.3	32.7±3.3	30.1±4.6	34.1±5.9	31.4±5.6	35.1±3.6	34.7±2.8	32.3±4.4	33.8±4.3
MelbournePedestrian	**20.2±1.9**	**20.3±2.3**	14.1±1.4	14.6±1.9	14.3±1.9	16.1±2.2	15.5±1.9	13.7±1.3	13.7±1.4	17.1±1.5	13.0±1.2	14.3±2.1	14.3±1.6	16.3±2.6
Chinatown	**72.1±8.2**	**72.1±8.2**	64.2±8.7	65.0±8.8	62.7±9.3	63.4±7.7	64.2±6.3	64.7±8.8	61.4±6.9	63.9±6.7	66.0±9.8	65.3±7.1	64.5±5.0	67.7±6.0
GunPointAgeSpan	**59.9±5.4**	59.0±6.7	53.2±5.5	54.7±5.8	52.3±6.6	52.1±6.0	49.0±5.1	51.2±4.9	51.9±5.5	53.4±8.4	50.1±6.2	51.7±9.6	53.0±5.8	54.1±5.4
Lightning2	62.7±12.2	**70.2±7.8**	59.5±10.2	61.2±10.5	60.4±10.1	61.0±8.7	56.2±14.1	57.0±11.1	56.2±12.3	60.3±19.6	60.2±14.5	61.1±10.7	56.2±12.5	57.1±5.9
Earthquakes	80.1±6.0	**80.5±5.8**	79.0±5.9	79.9±6.5	79.0±6.3	79.4±5.6	78.3±4.1	79.4±5.6	79.4±6.6	77.5±6.6	79.2±5.3	77.7±5.2	79.9±6.0	79.9±6.0
LargeKitchenAppliances	41.9±6.0	**42.4±5.7**	35.6±3.5	39.2±4.9	38.0±4.7	40.4±5.6	40.1±7.4	37.9±3.7	35.5±6.1	33.7±6.6	38.1±4.5	37.9±4.3	37.2±5.8	42.0±5.9
Average	52.7	**52.9**	45.2	48.3	47.5	47.9	47.5	46.6	45.6	48.0	48.9	48.7	47.2	49.3
Average ranking	12.7	**13.5**	4.8	8.0	6.3	8.5	6.0	5.3	4.4	5.8	6.8	6.8	6.4	9.9
Nemenyi Pvalues	GTEF		**<0.001**	0.23	**<0.05**	0.42	**<0.01**	**<0.01**	**<0.001**	**<0.01**	**<0.05**	**<0.05**	**<0.05**	0.93
	TEF		**<0.001**	**<0.1**	**<0.01**	0.17	**<0.01**	**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.01**	**<0.01**	0.70
Wilcoxon Pvalues	GTEF		**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.001**	**<0.001**	**<0.01**
	TEF		**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.01**	**<0.001**	**<0.01**	**<0.001**	**<0.001**

Based on these findings, in the BME, Wafer, TwoPatterns, MelbournePedestrian, Chinatown, Earthquakes, and LargeKitchenAppliances datasets, the proposed methods exhibit superior performance over the remaining ten algorithms and the raw data when applied to the KNN classifier. On average, there is a 6.3% enhancement in KNN accuracy. Similarly, concerning the Random Forest classifier, the proposed methodologies excel over competing algorithms in datasets such as Fungi, BME, SonyAIBORobotSurface1, TwoPatterns, CBF, MelbournePedestrian, Chinatown, GunPointAgeSpan, Earthquakes, and LargeKitchenAppliances. In this scenario, there is an average improvement of 5.7% in

TABLE VII
DETAILS OF REAL DATASETS

No.	Datasets	Size	Sources(Dimensions)	Length	Classes
1	BasicMotions	80	6	100	4
2	RacketSports	303	6	30	4
3	ArticularyWordRecognition	575	9	144	25
4	LSST	4925	6	36	14
5	PenDigits	10992	2	8	10
6	NATOPS	360	24	51	6
7	Cricket	180	6	1197	12
8	DuckDuckGeese	100	1345	270	5
9	EMOPain	1143	30	180	3

Random Forest accuracy. Moreover, in the case of the ExN-Rule Ensemble, the proposed techniques outshine alternative algorithms across all 12 datasets, showcasing an average accuracy boost of 7.7% in the ExNRule Ensemble classifier. In addition, the Wilcoxon signed-rank and Nemenyi tests results show that the GETF and TEF methods significantly outperform most algorithms, such as mean fusion (MF), fuzzy dominated entropy fusion (FDEF), fuzzy incomplete entropy fusion (FIEF), significance degree fusion (SDF), K-nearest unsupervised feature selection (KND-UFS), and KNCMI.

C. Performance on Real Datasets

In general, the channel dimension of multivariate time series can be regarded as distinct sources of information [6]. Therefore, to further demonstrate the practicality of our methods, we conduct tests on nine real multivariate datasets. The details of the real datasets are shown in Table VII. Consistent with the simulated data, refer to Tables VIII–X for the comparative results.

Based on the findings, in the ArticularyWordRecognition, PenDigits, NATOPS, and EMOPain datasets, both the GTEF and TEF methods exhibit superior performance over the remaining ten algorithms and the raw data when applied to the KNN classifier. On average, there is a 0.5% enhancement in KNN accuracy. Similarly, concerning the Random Forest classifier, the proposed methodologies excel over competing algorithms in all datasets. In this scenario, there is an average improvement of 21.0% in Random Forest accuracy. Moreover, in the case of the ExNRule Ensemble, the proposed techniques outshine alternative algorithms across eight datasets, showcasing an average accuracy boost of 9.1% in the ExNRule Ensemble classifier. Furthermore, the Wilcoxon signed-rank and Nemenyi test results show that the GETF and TEF methods significantly outperform most algorithms, such as KND-UFS, SDF-FS, and NEF-FS.

To position the proposed methods within the current research landscape, we compare TEF and GTEF against several contemporary machine learning and deep learning techniques [36], [37], [38], [39], [40], [41], [42]. These baseline models are executed five times with various random seeds on the original training and testing sets. The batch size is set to 32, the learning rate is 0.001, and the number of epochs is 30. The average accuracy is then calculated as the final result. The comparison results of accuracy can be seen in Table XI. The results show that the proposed TEF+Random Forest approach outperforms the other seven techniques, and

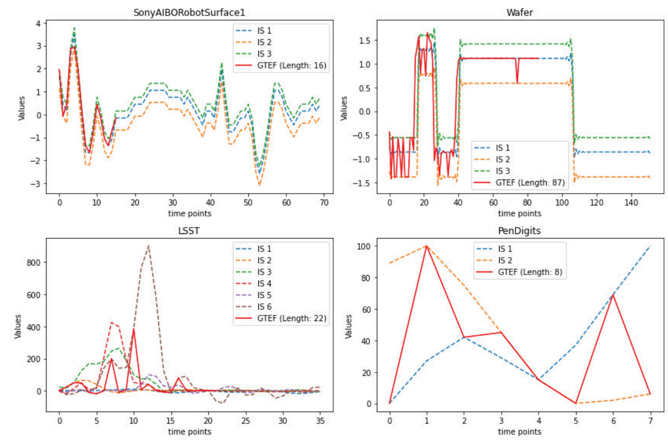

Fig. 4. Illustration of the selected subset of features. The dashed lines represent multisource data, while the red solid lines denote the outcomes of GTEF.

the GTEF+KNN/Random Forest/ExNRule Ensemble performs better than the LSTM.

D. Efficiency Analysis

In order to intricately compare the computational efficiency of the proposed methods, the time complexities of all algorithms, along with the average runtimes of each algorithm on all datasets, are presented in Table XII. The results indicate that the runtime of TEF is lower than that of NEF-FS, while the runtime of GTEF is lower than NEF-FS, KNCMI, NEF, SDF-FS, and SDF. In overall computational efficiency, the ranking is as follows: MF > tolerance incomplete entropy fusion (TIEF) > FDEF> FIEF > KND-UFS > **GTEF** > SDF > SDF-FS > NEF > KNCMI > **TEF** > NEF-FS.

Furthermore, in terms of selection efficiency, the GTEF method is compared with state-of-the-art algorithms based on the number of selected features, as shown in Table XIII. Results indicate that the number of features selected by GTEF is relatively small across the majority of datasets. Moreover, it can be found that in some situations, the number of features obtained by GTEF is much smaller than that obtained by other methods. However, the selection efficiency has increased, which may lead to the loss of certain important features, resulting in decreased accuracy, such as BasicMotions and DuckDuckGeese. Fig. 4 illustrates the performance of the selected feature subsets on two simulated and real datasets. It is evident that the GTEF method can capture trend information and eliminate redundant features.

E. Robustness Analysis

To demonstrate the model's robustness, the proposed models' performance is tested on nine simulated datasets under different noise levels, with noise variance ranging from 0.1 to 1 in increments of 0.1. Without loss of generality, Fig. 5 only shows the performance based on KNN. From the results, it can be observed that the accuracy of the GTEF and TEF methods remains relatively stable across different levels of noise. This

TABLE VIII

COMPARISON OF ACCURACY WITH REAL DATASETS BASED ON THE KNN. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets		GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
BasicMotions		42.5±12.7	63.7±16.2	61.3±15.3	61.3±17.2	57.5±17.0	60.0±15.6	60.0±22.2	57.5±17.9	**72.5±14.6**	27.5±15.6	45.0±21.8	70.0±13.9	57.5±17.9	60.0±20.8
RacketSports		65.0±6.9	75.6±4.8	65.4±7.3	76.2±6.7	72.9±8.4	73.6±8.4	**79.2±5.7**	70.9±8.8	67.6±11.0	35.0±6.8	74.9±7.8	67.0±8.9	70.9±8.8	71.6±7.4
ArticularyWordRecognition		39.7±7.5	**95.8±2.4**	64.0±6.4	91.7±4.9	76.3±4.9	88.3±3.4	92.7±4.4	94.6±3.0	62.9±7.5	11.8±3.3	74.1±6.0	61.0±6.3	48.9±7.0	79.0±2.2
LSST		42.5±1.5	44.5±2.5	40.8±2.1	42.3±1.3	42.6±1.6	39.6±1.7	41.1±1.2	**45.3±2.8**	39.6±2.0	24.1±1.2	37.2±2.2	39.8±1.9	40.2±1.7	41.6±1.1
PenDigits		**97.5±0.5**	96.2±0.4	95.6±0.7	94.9±0.6	94.2±0.8	97.5±0.5	94.9±0.6	91.5±0.5	95.6±0.7	22.9±3.7	74.1±0.8	95.6±0.7	91.5±0.5	91.5±0.5
NATOPS		17.5±8.2	**75.0±6.1**	40.6±4.2	64.7±8.6	64.7±7.9	68.1±6.6	54.7±8.8	60.6±7.7	40.0±6.1	18.3±7.8	50.0±8.5	41.7±8.4	52.8±6.0	73.6±7.8
Cricket		73.3±6.9	86.1±5.1	72.8±8.4	90.0±3.3	43.3±17.2	81.1±8.3	**92.2±5.1**	57.2±10.8	60.0±14.0	35.0±13.2	62.2±16.4	68.9±12.0	61.1±9.3	75.6±6.7
DuckDuckGeese		26.0±10.2	38.0±12.5	28.0±18.3	**48.0±17.8**	39.0±7.0	41.0±14.5	45.0±18.6	24.0±16.2	22.0±12.5	21.0±13.0	32.0±13.3	32.0±18.9	26.0±12.0	36.0±10.2
EMOPain		**88.4±2.3**	85.4±1.8	83.7±3.5	86.5±2.5	85.5±2.9	78.6±3.5	85.6±2.9	85.0±2.8	84.2±3.4	73.6±2.8	74.8±3.3	84.4±2.8	76.2±4.0	82.8±3.5
Average		54.7	**73.4**	61.4	72.8	64.0	69.8	71.7	65.2	60.5	29.9	58.3	62.3	58.3	68.0
Average ranking		6.56	**12.11**	6.72	11.39	8.28	9.33	10.72	7.39	5.94	1.11	5.17	7.06	4.89	8.33
Nemenyi P-values	GTEF			1.00	0.44	1.00	0.98	0.69	1.00	1.00	0.24	1.00	1.00	1.00	1.00
	TEF			0.26	1.00	0.80	0.98	1.00	0.48	**<0.1**	**<0.001**	**<0.05**	0.36	**<0.05**	0.82
Wilcoxon P-values	GTEF			0.82	0.98	0.90	0.97	0.98	0.79	0.67	**<0.01**	0.63	0.88	0.69	0.96
	TEF			**<0.01**	0.29	**<0.01**	**<0.05**	0.46	**<0.01**	**<0.05**	**<0.01**	**<0.01**	**<0.05**	**<0.01**	**<0.01**

TABLE IX

COMPARISON OF ACCURACY WITH REAL DATASETS BASED ON THE RANDOM FOREST. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets		GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
BasicMotions		52.5±20.8	**82.5±10.0**	61.3±16.2	68.8±20.3	70.0±17.9	53.8±11.2	72.5±20.0	63.7±14.2	47.5±18.4	41.2±16.8	55.0±12.7	66.2±15.9	63.7±14.2	67.5±11.5
RacketSports		64.7±10.5	**75.5±5.9**	53.8±6.8	62.0±7.4	62.7±8.6	61.6±8.4	62.0±6.5	54.7±8.2	46.9±8.5	31.0±4.2	51.2±6.0	50.2±8.7	54.7±8.2	56.8±9.1
ArticularyWordRecognition		38.9±5.3	**89.9±2.9**	27.0±7.3	36.7±7.7	28.2±7.2	32.2±4.8	35.5±8.1	35.2±9.4	25.8±8.4	12.5±3.6	33.1±5.9	24.0±6.8	21.2±3.3	30.1±5.8
LSST		**46.9±1.8**	46.3±1.9	38.0±2.0	39.7±1.5	38.5±2.2	39.4±2.0	39.2±2.0	40.4±2.1	38.2±2.1	38.7±2.2	38.6±1.5	38.1±2.1	38.2±1.8	40.0±1.6
PenDigits		**95.9±0.4**	94.3±0.9	61.4±2.1	65.8±2.6	67.6±1.6	64.8±2.1	65.8±2.6	60.1±1.4	59.3±3.9	30.5±1.8	54.9±1.8	62.4±2.5	60.1±1.4	60.1±1.4
NATOPS		24.7±3.4	**76.9±6.6**	36.9±12.0	54.2±6.9	58.9±3.5	54.2±6.6	48.3±7.4	57.2±7.9	35.3±6.0	14.4±4.1	48.9±6.9	35.8±7.1	45.6±6.8	59.7±9.2
Cricket		78.3±7.6	**90.6±7.0**	32.8±10.1	48.3±10.6	32.8±10.7	43.9±11.2	46.7±11.2	35.0±13.6	30.0±12.2	37.8±18.1	46.1±12.9	37.2±10.3	37.8±15.9	46.1±10.0
DuckDuckGeese		34.0±11.1	**42.0±21.4**	23.0±13.5	30.0±12.6	23.0±12.7	24.0±12.0	28.0±14.7	22.0±15.4	18.0±9.8	11.0±11.4	25.0±11.2	29.0±14.5	23.0±12.7	37.0±14.9
EMOPain		**84.2±2.5**	80.9±2.2	81.8±3.3	84.1±2.4	79.0±2.9	77.5±3.3	83.7±2.4	77.9±3.7	80.6±3.1	77.2±3.6	76.1±3.7	81.0±2.6	77.1±3.3	80.9±3.2
Average		57.8	**75.4**	46.2	54.4	51.2	50.2	53.5	49.6	42.4	32.7	47.7	47.1	46.8	53.1
Average ranking		10.89	**13.17**	5.28	11.06	8.06	7.50	10.22	7.11	3.06	2.50	5.83	6.00	4.89	9.44
Nemenyi P-values	GTEF			0.20	1.00	0.98	0.91	1.00	0.82	**<0.01**	**<0.01**	0.36	0.42	0.12	1.00
	TEF			**<0.01**	1.00	0.35	0.19	0.97	0.11	**<0.001**	**<0.001**	**<0.05**	**<0.05**	**<0.01**	0.83
Wilcoxon P-values	GTEF			**<0.1**	0.15	0.18	**<0.05**	0.15	0.13	**<0.05**	**<0.01**	**<0.05**	**<0.1**	0.13	0.21
	TEF			**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**

TABLE X

COMPARISON OF ACCURACY WITH REAL DATASETS BASED ON THE EXNRULE ENSEMBLE. UNDERLINED RESULTS INDICATE FEATURES DOWNSAMPLED WITH A STRIDE OF 8. P-VALUE < α INDICATES THAT THE PROPOSED METHOD SIGNIFICANTLY OUTPERFORMS THE CONTRAST ALGORITHM

Datasets		GTEF	TEF	MF	TIEF	FDEF	NEF	FIEF	SDF	KND-UFS	SDF-FS	NEF-FS	KNCMI	Mean of sources	Max of sources
BasicMotions		48.8±17.2	**81.3±15.1**	60.0±14.6	66.2±14.8	66.2±9.8	62.5±15.8	75.0±14.8	66.2±14.8	67.5±13.9	32.5±11.5	58.8±12.6	78.8±11.2	60.0±14.6	73.8±14.2
RacketSports		64.3±11.3	**74.9±6.5**	54.5±6.6	61.0±7.7	62.3±11.7	53.4±10.7	60.7±6.8	56.1±5.9	48.8±5.5	31.0±7.8	53.1±6.2	54.8±7.7	55.8±7.4	66.0±9.1
ArticularyWordRecognition		24.9±6.1	**90.8±3.9**	51.5±9.6	84.5±3.8	61.9±4.5	69.4±4.7	82.4±4.6	80.0±4.0	48.7±5.6	10.4±3.3	47.1±7.1	39.6±6.7	39.3±5.3	71.8±3.5
LSST		44.3±2.0	**45.6±2.5**	36.2±2.4	34.8±2.1	34.7±2.3	33.5±2.5	34.0±1.3	36.4±2.0	34.4±2.3	24.2±2.3	30.7±2.9	35.0±1.8	34.4±1.9	36.5±2.4
PenDigits		**87.9±1.3**	84.0±1.7	79.1±1.6	69.7±2.4	75.5±2.2	69.5±2.8	70.6±2.0	72.5±2.0	79.7±2.0	25.9±1.5	46.2±3.1	79.8±2.4	72.8±1.7	72.8±1.7
NATOPS		24.2±6.9	**73.9±4.8**	36.4±6.9	58.1±8.4	58.3±9.5	56.9±5.7	48.6±4.5	60.8±8.9	36.4±7.8	16.7±5.3	49.7±6.3	34.2±6.6	46.7±6.8	71.4±8.4
Cricket		66.7±11.1	**83.9±7.2**	62.8±13.4	78.9±9.6	39.4±17.1	65.6±12.1	73.9±9.0	49.4±12.8	46.1±17.6	29.4±11.1	48.3±15.7	57.2±16.1	55.6±15.1	63.9±7.1
DuckDuckGeese		32.0±16.0	38.0±13.3	23.0±16.2	35.0±20.6	30.0±17.3	27.0±12.7	**40.0±19.5**	21.0±8.3	14.0±8.0	20.0±10.0	29.0±8.3	28.0±13.3	25.0±12.0	38.0±19.9
EMOPain		84.7±2.0	**86.5±2.0**	83.1±2.7	86.9±2.3	82.3±3.7	77.5±3.5	85.5±2.7	81.3±3.2	82.0±3.6	73.3±2.8	77.0±2.9	81.8±2.5	77.1±3.5	82.5±2.6
Average		53.1	**73.2**	54.1	63.9	56.7	57.3	63.4	58.2	50.8	29.3	48.9	54.4	51.9	64.1
Average ranking		8.56	**13.61**	7.00	10.11	8.11	6.00	9.67	7.67	5.56	1.11	4.11	7.44	5.28	10.78
Nemenyi P-values	GTEF			1.00	1.00	1.00	0.99	1.00	1.00	0.96	**<0.05**	0.59	1.00	0.93	1.00
	TEF			**<0.1**	0.89	0.23	**<0.01**	0.77	0.13	**<0.01**	**<0.001**	**<0.001**	**<0.1**	**<0.01**	0.98
Wilcoxon P-values	GTEF			0.59	0.85	0.54	0.54	0.88	0.59	0.41	**<0.01**	0.33	0.59	0.50	0.79
	TEF			**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**

TABLE XI

COMPARISON OF ACCURACY WITH MACHINE LEARNING AND DEEP LEARNING METHODS

Datasets	GTEF+KNN	TEF+KNN	GTEF+RF	TEF+RF	GTEF+ExNRule	TEF+ExNRule	LSTM [36]	PatchTST [37]	Autoformer [38]	Crossformer [39]	DLinear [40]	LightTS [41]	TimeSeriesForest [42]
BasicMotions	42.5	63.7	52.5	82.5	48.8	81.3	42	61	54	73	72.5	57	98.5
RacketSports	65	75.6	64.7	75.5	64.3	74.9	25.4	73.2	78.6	78.9	66.8	70.5	74.5
ArticularyWordRecognition	39.7	95.8	38.9	89.9	24.9	90.8	6.6	82.5	58.5	96.9	96.9	61.6	87.7
LSST	42.5	44.5	46.9	46.3	44.3	45.6	31.5	51.2	41.4	35.8	31.7	36.6	43
PenDigits	97.5	96.2	95.9	94.3	87.9	84	82.8	95.6	96.4	91.8	88.6	95.8	89.5
NATOPS	17.5	75	24.7	76.9	24.2	73.9	18	63.9	79.2	84.8	89.0	83.7	80.4
Cricket	73.3	86.1	78.3	90.6	66.7	83.9	7.2	93.3	24.7	75.3	90.3	88.9	83.9
DuckDuckGeese	26	38	34	42	32	38	24	21.6	29.2	29.6	62.8	51.6	38.8
EMOPain	88.4	85.4	84.2	80.9	84.7	86.5	69.4	81.2	70.5	78.5	78.3	79.7	78.2
Average	54.7	73.4	57.8	**75.4**	53.1	73.2	34.1	69.3	59.2	71.6	75.2	69.5	74.9

indicates that these two methods are robust and can adapt to noise interference in various scenarios.

F. Parameter Sensitivity Analyses

To demonstrate the performance of the models under different parameters, the models' accuracy is tested on several simulated and real datasets. The parameters δ_1 and δ_2 are varied from 0.05 to 0.5 in increments of 0.05. Without loss of generality, Fig. 6 only shows the results under the KNN. For the majority of datasets, the impact of different parameters on the models' performance is relatively small.

TABLE XII

COMPLEXITY AND RUNTIME COMPARISON OF THE PROPOSED ALGORITHM WITH OTHER COMPARATIVE ALGORITHMS

Algorithms	Time complexity	Mean runtime(s)
GTEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	1395.70
TEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	4411.00
MF	$O\left(\lvert T\rvert \times \lvert U\rvert \times \lvert N\rvert\right)$	0.01
TIEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	732.90
FDEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	918.40
NEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	1926.00
NEF-FS	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2 + \lvert T\rvert^2\right)$	13753.20
FIEF	$O\left(\lvert T\rvert \times N \times \lvert U\rvert^2\right)$	1207.90
SDF	$O\left(N \times \lvert T\rvert \times \lvert U\rvert\right)$	1632.70
SDF-FS	$O\left(N \times \lvert T\rvert \times \lvert U\rvert + \lvert U\rvert \times \lvert T\rvert^2\right)$	1922.90
KND-UFS	$O\left(\lvert T\rvert \times \lvert U\rvert^2\right)$	1363.50
KNCMI	$O\left(\lvert T\rvert \times \lvert U\rvert^2 + \lvert T\rvert^3\right)$	2585.00

TABLE XIII

COMPARISON OF SELECTED FEATURES NUMBER BASED ON GTEF WITH OTHER COMPARATIVE ALGORITHMS

Datasets	Raw	GTEF	KND-UFS	SDF-FS	NEF-FS	KNCMI
Fungi	201	10	58	1	100	176
BME	128	6	6	1	64	16
SonyAIBORobotSurface1	70	16	10	1	35	6
Wafer	152	87	32	1	9	11
TwoPatterns	128	36	49	1	8	15
CBF	128	8	14	1	8	9
MelbournePedestrian	24	23	11	1	12	13
Chinatown	24	16	21	1	12	11
GunPointAgeSpan	150	50	2	1	75	6
Lightning2	637	69	12	1	40	8
Earthquakes	512	31	41	1	32	61
LargeKitchenAppliances	720	622	86	1	45	64
BasicMotions	100	6	8	1	50	7
RacketSports	30	8	19	1	15	30
ArticularyWordRecognition	144	5	86	1	9	18
LSST	36	22	33	1	18	36
PenDigits	8	8	8	1	4	8
NATOPS	51	9	36	1	25	27
Cricket	1197	4	86	1	75	143
DuckDuckGeese	270	6	17	1	135	182
EMOPain	200	8	17	1	12	20
Average	233.81	50.00	31.05	1.00	37.29	41.29

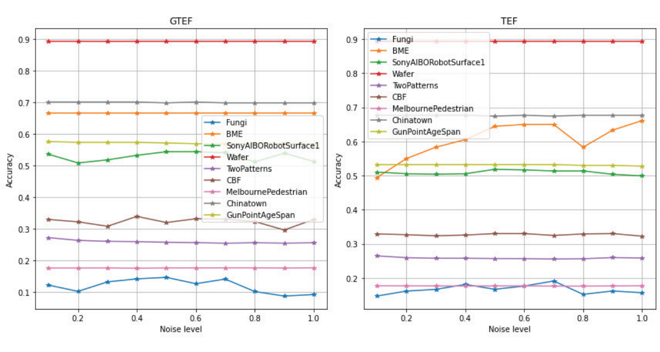

Fig. 5. Performance of the proposed methods based on KNN under different noise levels. The noise variance is ranged from 0.1 to 1 in increments of 0.1.

This suggests that our methods are not highly sensitive to parameter selection. Based on the experiments in this article, setting the parameters δ_1 and δ_2 to 0.1 is appropriate. It is generally recommended that these parameters should not be excessively large or small because overly large values may result in larger trend classes, leading to model underfitting. In contrast, excessively small values can create tiny trend classes, also causing underfitting [43]. Therefore, to achieve better performance in practical applications, it is advisable to utilize grid search or Bayesian optimization for fine-tuning parameter values, with recommended adjustments within the range of [0.05, 0.5].

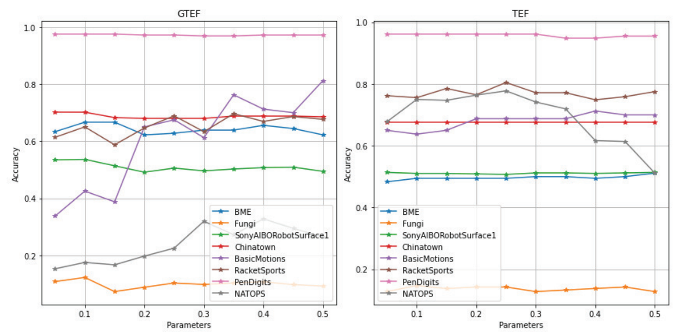

Fig. 6. Performance of the proposed methods based on KNN under different parameters. The parameters δ_1 and δ_2 are varied from 0.05 to 0.5 in increments of 0.05.

VI. CONCLUSION

The fusion of multisource time-series data is of great importance due to its widespread occurrence in real-life scenarios. This study introduces a fusion and selection framework for integrating multisource time-series data, leveraging the concept of information entropy. The framework identifies the most important data sources by minimizing information entropy. Moreover, a feature selection algorithm is employed to optimize the fusion process, ensuring that the conditional entropy of each fusion result does not surpass the previous step. It effectively selects the most important sources while considering the overall impact of features and eliminating redundancy. The experimental results demonstrate the proposed algorithms' superiority in improving classification accuracy, outperforming the most existing state-of-the-art fusion and feature selection methods. However, the methods proposed in this article have two notable limitations: the time complexity of $O(|T| \times N \times |U|^2)$ may restrict applicability to large datasets, prompting future exploration of efficient data structures, such as tree structures, or sampling techniques to mitigate this computational burden. In addition, the static nature of the framework does not accommodate dynamic data, necessitating research into incremental fusion approach for real-time streaming environments.

REFERENCES

[1] L. Bai, L. Cui, Z. Zhang, L. Xu, Y. Wang, and E. R. Hancock, "Entropic dynamic time warping kernels for co-evolving financial time series analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1808–1822, Apr. 2023.

[2] Y. Huang, G. G. Yen, and V. S. Tseng, "Snippet policy network v2: Knee-guided neuroevolution for multi-lead ECG early classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2167–2181, Feb. 2024.

[3] Z. Zhong, Z. Yu, Z. Fan, C. L. P. Chen, and K. Yang, "Adaptive memory broad learning system for unsupervised time series anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 26, 2024, doi: 10.1109/TNNLS.2024.3415621.

[4] P. Zhang et al., "Multi-source information fusion based on rough set theory: A review," *Inf. Fusion*, vol. 68, pp. 85–117, Apr. 2021.

[5] W. Wei and J. Liang, "Information fusion in rough set theory: An overview," *Inf. Fusion*, vol. 48, pp. 107–118, Aug. 2019.

[6] T. Zhan, Y. He, Y. Deng, Z. Li, W. Du, and Q. Wen, "Time evidence fusion network: Multi-source view in long-term time series forecasting," 2024, *arXiv:2405.06419*.

[7] T. Lei, J. Li, and K. Yang, "Time and frequency-domain feature fusion network for multivariate time series classification," *Expert Syst. Appl.*, vol. 252, Oct. 2024, Art. no. 124155.

[8] J.-T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.

[9] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.

[10] L. Chen and Y. Deng, "GDTRSET: A generalized decision-theoretic rough sets based on evidence theory," *Artif. Intell. Rev.*, vol. 56, no. S3, pp. 3341–3362, Dec. 2023.

[11] B. Sang, Y. Guo, D. Shi, and W. Xu, "Decision-theoretic rough set model of multi-source decision systems," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 11, pp. 1941–1954, Nov. 2018.

[12] S. An, E. Zhao, C. Wang, G. Guo, S. Zhao, and P. Li, "Relative fuzzy rough approximations for feature selection and classification," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2200–2210, Apr. 2023.

[13] S. Xia et al., "An efficient and accurate rough set for feature selection, classification, and knowledge representation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 7724–7735, Aug. 2023.

[14] X. Wang and X. Zhang, "Linear-combined rough vague sets and their three-way decision modeling and uncertainty measurement optimization," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 11, pp. 3827–3850, Nov. 2023.

[15] J. Wang, J. Tang, S. Tong, and X. Li, "Uncertainty measure based on rough set in information systems," in *Proc. 5th Int. Conf. Inf. Technol. Comput. Commun.*, New York, NY, USA, Jun. 2023, pp. 68–71.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Oct. 1948.

[17] P. Zhang et al., "A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data," *Inf. Fusion*, vol. 80, pp. 87–103, Apr. 2022.

[18] W. Xu, Y. Pan, X. Chen, W. Ding, and Y. Qian, "A novel dynamic fusion approach using information entropy for interval-valued ordered datasets," *IEEE Trans. Big Data*, vol. 9, no. 3, pp. 845–859, Mar. 2023.

[19] W. Xu, M. Li, and X. Wang, "Information fusion based on information entropy in fuzzy multi-source incomplete information system," *Int. J. Fuzzy Syst.*, vol. 19, no. 4, pp. 1200–1216, Aug. 2017.

[20] X. Zhang, X. Chen, W. Xu, and W. Ding, "Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes," *Inf. Sci.*, vol. 608, pp. 1–27, Aug. 2022.

[21] W. Qian, S. Yu, J. Yang, Y. Wang, and J. Zhang, "Multi-label feature selection based on information entropy fusion in multi-source decision system," *Evol. Intell.*, vol. 13, no. 2, pp. 255–268, Jun. 2020.

[22] M. Li and X. Zhang, "Information fusion in a multi-source incomplete information system based on information entropy," *Entropy*, vol. 19, no. 11, p. 570, Nov. 2017.

[23] W. Xu, K. Cai, and D. D. Wang, "A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets," *Int. J. Approx. Reasoning*, vol. 164, Jan. 2024, Art. no. 109081.

[24] W. Li, H. Zhou, W. Xu, X.-Z. Wang, and W. Pedrycz, "Interval dominance-based feature selection for interval-valued ordered data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6898–6912, Oct. 2023.

[25] W. Xu, M. Huang, Z. Jiang, and Y. Qian, "Graph-based unsupervised feature selection for interval-valued information system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 9, pp. 12576–12589, Sep. 2024, doi: 10.1109/TNNLS.2023.3263684.

[26] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, and Z. Yi, "Large-scale metaheuristic feature selection based on BPSO assisted rough hypercuboid approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10889–10903, Dec. 2023.

[27] P. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, and X. Yang, "Heterogeneous feature selection based on neighborhood combination entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 3514–3527, Mar. 2024.

[28] N. N. Thuy and S. Wongthanavasu, "A novel feature selection method for high-dimensional mixed decision tables," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3024–3037, Jul. 2022.

[29] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, Sep. 2008.

[30] X. Chen and W. Xu, "Double-quantitative multigranulation rough fuzzy set based on logical operations in multi-source decision systems," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 4, pp. 1021–1048, Apr. 2022.

[31] W. Xu, Y. Zhang, and Y. Qian, "A novel unsupervised feature selection for high-dimensional data based on FCM and *k*-nearest neighbor rough sets," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 30, 2024, doi: 10.1109/TNNLS.2024.3460796.

[32] W. Xu, Z. Yuan, and Z. Liu, "Feature selection for unbalanced distribution hybrid data based on *k*-nearest neighborhood rough set," *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 229–243, Jan. 2024.

[33] A. Ali, M. Hamraz, N. Gul, D. M. Khan, S. Aldahmani, and Z. Khan, "A *k* nearest neighbour ensemble via extended neighbourhood rule and feature subsets," *Pattern Recognit.*, vol. 142, Oct. 2023, Art. no. 109641.

[34] A. Ali, Z. Khan, D. Muhammad Khan, and S. Aldahmani, "An optimal random projection *k* nearest neighbors ensemble via extended neighborhood rule for binary classification," *IEEE Access*, vol. 12, pp. 61401–61409, 2024.

[35] B. Sang, L. Yang, H. Chen, W. Xu, Y. Guo, and Z. Yuan, "Generalized multi-granulation double-quantitative decision-theoretic rough set of multi-source information system," *Int. J. Approx. Reasoning*, vol. 115, pp. 157–179, Dec. 2019.

[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[37] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," 2022, *arXiv:2211.14730*.

[38] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. NIPS*, vol. 34, Dec. 2021, pp. 22419–22430.

[39] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *Proc. 11th Int. Conf. Learn. Represent.*, 2023, pp. 1–11.

[40] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 9, pp. 11121–11128.

[41] T. Zhang et al., "Less is more: Fast multivariate time series forecasting with light sampling-oriented MLP structures," 2022, *arXiv:2207.01186*.

[42] H. Deng, G. Runger, E. Tuv, and M. Vladimir, "A time series forest for classification and feature extraction," *Inf. Sci.*, vol. 239, pp. 142–153, Aug. 2013.

[43] X. Chen and M. Luo, "Incremental information fusion in the presence of object variations for incomplete interval-valued data based on information entropy," *Inf. Sci.*, vol. 667, May 2024, Art. no. 120479.

Xiuwei Chen received the B.S. degree in economic statistics from Nanjing Audit University, Nanjing, China, in 2020, and the M.S. degree in applied statistics from Southwest University, Chongqing, China, in 2022. He is currently pursuing the Ph.D. degree with the College of Mathematics, Sichuan University, Chengdu, China.

His current research interests include information fusion, feature selection, and granular computing.

Li Lai received the B.S. degree in statistics from Nankai University, Tianjin, China, in 2006, and the M.Sc. and Ph.D. degrees in mathematics from Sichuan University, Chengdu, China, in 2009 and 2014, respectively.

She is currently working as an Associate Professor with the College of Mathematics, Sichuan University. Her current research interests include uncertainty processing and artificial intelligence.

Maokang Luo received the M.Sc. and Ph.D. degrees in mathematics from Sichuan University, Chengdu, China, in 1984 and 1992, respectively.

He is currently working as the Director of the National Center for Applied Mathematics in Sichuan, Sichuan Key Laboratory of Nonlinear Uncertain Engineering System Control, and Sichuan Key Laboratory of Information and Mathematics Technology, Chengdu. His research interests include uncertainty theory and signal processing. He has published more than 100 papers on these fields in international journals.