Information Sciences 642 (2023) 119107

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Multi-channel EEG signals classification via CNN and multi-head self-attention on evidence theory

Lang Zhang [a], Fuyuan Xiao [a,*], Zehong Cao [b]

[a] School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China
[b] STEM, Mawson Lakes Campus, University of South Australia, Adelaide, SA 5095, Australia

ARTICLE INFO

Keywords:
Multi-source information fusion
Evidence theory
Convolutional neural networks (CNN)
Multi-head self-attention mechanisms
Uncertainty
Electroencephalography (EEG)
Classification

ABSTRACT

Electroencephalography (EEG) provides valuable physiological information to identify human activities. However, it can be difficult to analyze the EEG data in human patterns identification, because both subjective and objective factors can easily affect sensitivity. In this study, a novel multi-head self-attention convolutional neural networks (CNN) framework based on Dempster-Shafer (D-S) evidence theory, called ETNN, is proposed to classify the EEG signal. The ETNN model considers the multi-type networks and fuses multi-output with D-S evidence theory, which can handle the EEG data more reasonably. In particular, a classification algorithm for EEG signals is derived with information fusion. Finally, an application for event-related potential signal classification and sensitivity analysis is used to demonstrate the effectiveness of the proposed ETNN model compared with existing classification techniques.

1. Introduction

The recognition of brain wave signals is a useful method for studying the workings of the human mind. Therefore, EEG-based classification tasks can contribute to various fields [1]. However, there is a challenge when classifying EEG patterns because they can be easily influenced by both subjective and objective factors in raw data. As a result, the uncontrollable factors mentioned above may lead to uncertainty issues in multisource information fusion [2,3]. However, uncertain multi-source information fusion can be applied in many fields, such as uncertainty expressing [4], database retrieval [5], decision-making [6–8], image scene classification [9], assessment [10,11], and so on [12,13]. Hence, several well-known theories have been proposed to manage uncertain multi-source information, such as Z-number [14]. With multisource information fusion, it can help to classify the EEG signals more accurately. Several methods have been proposed to process EEG signals for feature extraction, such as inherent fuzzy entropy [15].

In this study, it is considered that multi-type convolutional neural networks [16] with multi-head self-attention mechanisms play significant parts in EEG signal classification. Deep neural networks can extract useful characteristics from EEG data by tweaking the parameters until the loss is as low as possible. Nonetheless, these neural networks did not completely consider location information, indicating that certain important neurons in the networks are inefficient at extracting information. In this case, multi-head self-attention mechanisms allow the model to focus on information from distinct representation subspaces at different places in neural networks, allowing for effective feature extraction. In practice, neural networks with multi-head self-attention processes are commonly utilized [17]. However, there is still a limitation when facing sensitive data. A single neural network may hardly obtain

* Corresponding author.
E-mail addresses: xiaofuyuan@cqu.edu.cn, doctorxiaofy@hotmail.com (F. Xiao).

https://doi.org/10.1016/j.ins.2023.119107
Received 13 May 2022; Received in revised form 26 April 2023; Accepted 2 May 2023

Available online 18 May 2023
0020-0255/

high classification accuracy. To address this issue, a large number of methods have been proposed based on anti-interference training or increasing the complexity of networks. Bagherzadeh et al. [18] proposed an enhanced artificial neural network to calculate sensitivity analytically rather than numerically. Bui et al. [19] coupled a whale optimization algorithm with a neural network to optimize computational parameters. While these efforts have led to some improvement in model performance, they have focused solely on using a single neural network to produce a single classification. To obtain more classification information, it may be beneficial to explore the use of multi-type neural networks.

With multi-type neural networks that employ multi-head self-attention mechanisms, increasing the amount of classification information can lead to increased model uncertainty. So, Managing the uncertainty associated with EEG-based classification information is a critical task. In this context, D-S evidence theory is taken into consideration as it offers a flexible approach to dealing with uncertainty problems [20–22]. Multi-source information can be flexibly fused using the D-S evidence theory, resulting in a more plausible categorization result. Particularly, the basic probability assignment (BPA) demonstrates the uncertainty of EEG-based classification information. However, when dealing with highly conflicting bodies of evidence, applying Dempster's rule of combination may produce unexpected results. Various pieces of study are offered in this case to deal with the conflicting evidence, such as divergence measure [23,24], distance measure [25], credal belief redistribution [26], uncertainty quantification [27,28], and so on [29,30]. Here, the distances between the classification information and the correct label are considered as a means of processing the original evidence in order to improve the accuracy of EEG-based classification tasks and manage conflicting evidence.

In this paper, a novel multi-head self-attention convolutional neural networks framework based on evidence theory, called ETNN is proposed. In this case, our proposed ETNN model first combines the multi-head self-attention mechanisms with evidence theory, which uses multiple neural networks to improve the performance of single neural networks, and uncertainty arising from multiple neural networks is addressed using evidence theory. Specifically, an EEG classification information fusion algorithm is derived. Furthermore, an application demonstrates the usefulness of the ETNN model in dealing with ERP signal categorization challenges based on a long-term attentive driving experiment. The main contribution of this study as follows:

- The proposed ETNN model is the first study to integrate convolutional neural networks with evidence theory. In particular, multi-head self-attention mechanisms are deployed to extract key features of EEG signals.
- A new algorithm based on evidence theory is used to combine the outputs of multi-type neural networks for EEG-based classification tasks.
- An application in ERP signal classification is put forward. ETNN model illustrates the excellent performance in valuation indexes over a single network or classical machine learning methods.

The organization of this paper is as follows. Section 2 delivers a review of the preliminaries of this work. An EEG signal classification framework is constructed in Section 3. Specifically, an algorithm for outputs information fusion is derived in this part. In Section 4, an application for ERP signal classification is put forward to verify the superiority of the ETNN model by comparing several well-known single networks and classical machine learning methods. Besides, components setting and other EEG datasets are discussed. In addition, sensitivity analysis is carried out to illustrate the robustness of the ETNN model. Section 5 concludes the summary of our study.

2. Preliminaries

This section briefly introduces the construction of CNN and multi-head self-attention mechanisms. Next, the D-S evidence theory is explained.

2.1. Convolutional neural network

A Convolutional neural network (CNN) is a multi-layer feed-forward neural network that can extra feature information automatically from datasets [16]. The main architecture of CNN consists of multiple hidden layers that are usually composed of convolutional and pooling layers. Kernels in convolutional layers are able to extract sophisticated and effective features from the raw data. Pooling layers can be used to lower the dimensionality of the feature after a convolutional layer to minimize over-fitting and reduce computational cost. The progress of CNN in dimensional data analysis may be effective for extracting EEG features for classification tasks.

2.2. Multi-head self-attention mechanisms

A fully connected network in deep learning may be limited in its ability to handle variable-length input sequences. For varying input lengths, the magnitude of the connection weight varies as well. As a result, the self-attention process can be utilized to dynamically produce weights for distinct neural network connections [17].

Definition 1 *(Multi-head self-attention mechanisms)*. Multi-head self-attention mechanisms usually adopt Query-Key-Value (Q-K-V) model. Let W_i^Q, W_i^K, W_i^V, W^O be parameter matrixes, d_k be the dimension of V. Then the multi-head attention can be defined as follows:

$$MultiHead(Q,K,V) = Concat(\ head_1, \dots, head_i, \dots head\ _h)\, W^O, \tag{1}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right), \tag{2}$$

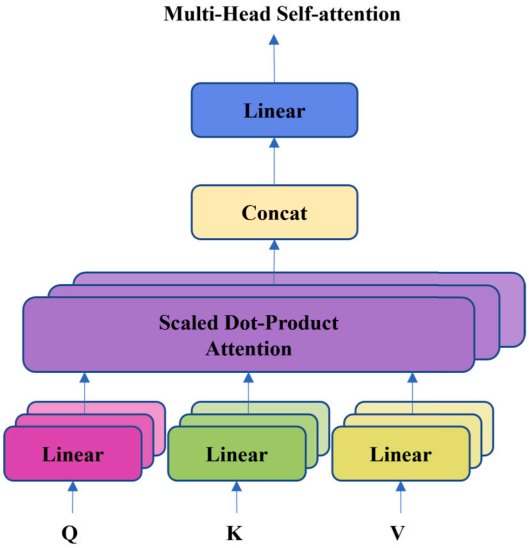


Fig. 1. The multi-head self-attention mechanisms.

$$Attention(Q,K,V) = Softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V, \tag{3}$$

$$Softmax(X)_{ij} = \frac{\exp\left(X_{ij}\right)}{\sum_j \exp\left(X_{ij}\right)}. \tag{4}$$

If Q, K and V are the same dimensions, the mechanisms are multi-head self-attention. The structure of multi-head self-attention mechanisms is shown in Fig. 1.

2.3. D-S evidence theory

D-S evidence theory [20,21] shows good performance in uncertain information fusion [31–33]. Specifically, singleton sets and multi-element sets of evidence can be expressed by BPA to illustrate the uncertainty [34]. Hence, D-S evidence theory contributes to data fusion [35], and so on [36–38]. When all probability values are known, evidence theory degenerates into probability theory.

Definition 2 *(Framework of discernment).* Let Θ be a set that consists of r mutually exclusive and collectively exhaustive events,

$$\Theta = \left\{e_1, e_2, \dots, e_i, \dots, e_r\right\}, \tag{5}$$

which indicates the framework of discernment. Here, Θ represents the set of all the hypotheses. Each hypothesis e_i represents a specific scenario of the variable. In evidence theory, the power set 2^{Θ} is used to describe uncertainty which can be defined as follows:

$$2^{\Theta} = \left\{\emptyset, \left\{e_1\right\}, \dots, \left\{e_r\right\}, \left\{e_1, e_2\right\}, \dots, \left\{e_1, e_2, \dots, e_h\right\}, \dots, \Theta\right\}, \tag{6}$$

where $\emptyset$ indicates the empty set. So, there will be 2^r propositions, which cover all the possibilities of a problem. In addition, the power set has expandability meaning in evidence theory [39].

Definition 3 *(Mass function).* In order to represent the probability assignment of events in evidence theory, m as a mass function, also known as BPA, is a mapping from 2^{Θ} to $[0,1]$ which is defined as:

$$m : 2^{\Theta} \rightarrow [0,1]. \tag{7}$$

Because events must arise from propositions in the framework of discernment, an empty set is not the cause. It abides the rule of

$$\sum_{E \in 2^{\Theta}} m(E) = 1 \quad \text{and} \quad m(\emptyset) = 0. \tag{8}$$

If $m(E) > 0$, E is a focal element. The mass function has been explored thoroughly to measure uncertainty, including information volume [40], overlapping degree [41], and so on [42].

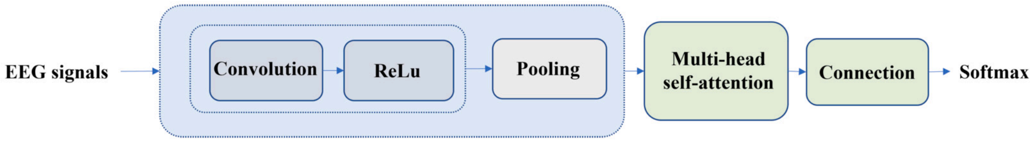


Fig. 2. Convolutional neural network based on multi-head self-attention.

Definition 4 *(Dempster's rule of combination).* Let m_1 and m_2 be two BPAs. Dempster's combination rule follows $m = m_1 \oplus m_2$:

$$m(A) = \begin{cases} \frac{1}{1-k} \sum_{P \cap Q = A} m_1(P) m_2(Q), & A \neq \emptyset, \\ 0, & A = \emptyset, \end{cases} \tag{9}$$

and

$$k = \sum_{P \cap Q = \emptyset} m_1(P) m_2(Q), \tag{10}$$

where P and Q are focal elements. Besides, k is regarded as a coefficient to represent the degree of conflict between m_1 and m_2. Several studies have been derived to illustrate the conflict more clearly. Specifically, the conflict between complex basic belief assignments in complex evidence theory can be measured based on a complex conflict coefficient [43].

3. A novel EEG signal classification framework based on evidence theory

This section starts by constructing a basic convolutional neural network based on multi-head self-attention mechanisms. Then, a novel EEG signal classification framework based on evidence theory, named ETNN, is derived.

3.1. Convolutional neural network based on multi-head self-attention

The basic architecture of CNN is based on multi-head self-attention shown in Fig. 2. The input is a set of EEG signals and the output is classification labels.

The structure can be divided into four parts. Firstly, there is a convolution layer to extract the local feature of the EEG signal. A convolution layer consists of a set of input feature maps, a convolution kernel, and a set of output feature maps. The active function, ReLu, can be used to increase the degree of model non-linearity:

$$ReLu: \ f(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{11}$$

where x is the value of the input feature map. A set of output feature map comes after a convolution layer. Secondly, a pooling layer contributes to feature extraction and reduces the connections in the network effectively. Maximum pooling and average pooling are both fundamental pooling functions:

$$y^d_{m,n} = \begin{cases} \max \ x_i, \\ \frac{1}{\left|R^d_{m,n}\right|} \sum_{i \in R^d_{m,n}} x_i, \end{cases} \tag{12}$$

where $|\cdot|$ represents the absolute value. $R^d_{m,n}$ is the pooling area and x_i is an active value in the area with $i \in R^d_{m,n}$. Thirdly, a multi-head self-attention mechanism follows closely behind. The results obtained by the pooling layer can be processed with the support of With Eqs. (1)-(4). The scaled dot-product attention can avoid reaching a region of the minimal gradient, and multi-head can improve the degree of parallelization.

Finally, there is a connection layer that reassembles the previous local features into complete global information by the weight matrix. And the final classification information is obtained with Eq. (4). Specifically, cross-entropy loss function L can be defined as:

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^{M} y_{ic} \log\left(p_{ic}\right), \tag{13}$$

where M represents the number of categories, y_{ic} represents the sign function (0 or 1), and p_{ic} represents the predicted probability that the observed sample i belongs to category c.

In this work, PyTorch (https://pytorch.org) was used to build the model architecture and Table 1 shows the specific parameters. Here, 70% of samples are randomly selected from the EEG signal data set as training samples, and the remaining 30% of samples are taken as testing samples. Especially, momentum algorithm is an optimization of SGD, and its momentum term can accelerate SGD in the relevant direction, suppress oscillations, and thus accelerate convergence.

Table 1
Hyper parameter settings for neural networks.

Hyper-parameters	Chosen value
Training samples	70%
Test samples	30%
Batch size	64
Optimizer	Stochastic Gradient Descent (SGD)
Momentum	0.9
Weight decay	0.02
Number of epochs	70
Learning rate	0.005
Dropout rate	0.2
Activation function	Relu, Softmax
Loss	Cross entropy

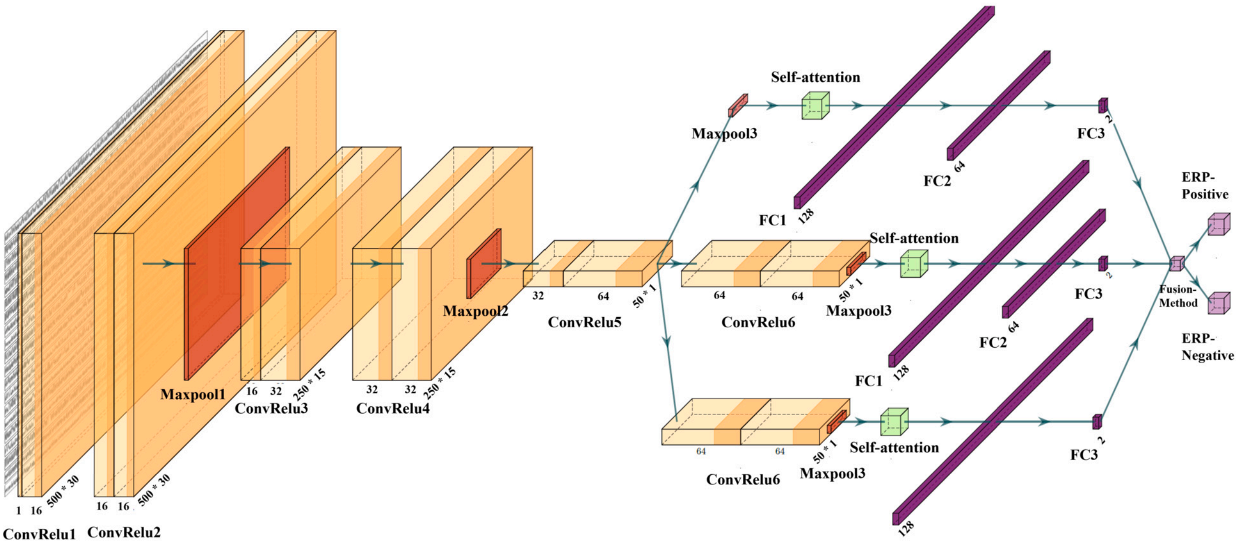


Fig. 3. A specific ETNN model for EEG signal classification with three types of neural network.

3.2. Multi-type neural network based on evidence theory

Most classification issues can be addressed with a convolutional neural network based on multi-head self-attention. However, it is well known that EEG signals are easily manipulated, which results in uncertainty outputs when a neural network is turned on. In this example, the evidence theory is employed to fuse multi-source information in order to eliminate the uncertainty created by a single neural network. Then, a novel EEG signal classification model, named ETNN, is proposed in this section.

The ETNN model could include a variety of multi-head self-attention neural networks and an information fusion model. In this case, the structure of neural networks must be constructed to meet the needs of individual applications, and the applications must also determine the number of networks. Furthermore, the experimental section will investigate the effect of network numbers on algorithm performance. In this section, based on the requirements of EEG signals, three specific neural networks are discussed. They are regarded as network A, network B and network C. In the beginning, there is the same basic part, which consists of five convolutional layers and two pooling layers. As for the first neural network, one pooling layer, the multi-head self-attention mechanisms, and three connection layers come after the basic part. As for the second neural network, there are one convolutional layer, one pooling layers, the multi-head self-attention mechanisms and three connection layers after the basic part. As for the third neural network, there are one convolutional layer, one pooling layer, the multi-head self-attention mechanisms, and two connection layers after the basic part. To show the structure more intuitively, the framework of this specific ETNN model is put forward in Fig. 3.

Suppose that there is EEG signal classification information after the ith neural network and it can be regarded as a BPA m_r, which contains the probabilities of $\mathcal{N}$ signal categories with the frame of discernment $\Theta = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{\mathcal{N}}\}$. Then the m_r can be defined as:

$$m_r : \; m_r(\{\mathcal{S}_1\}), m_r(\{\mathcal{S}_2\}), \dots, m_r(\{\mathcal{S}_i\}), \dots, m_r(\{\mathcal{S}_1, \mathcal{S}_2\}), \dots, m_r(\Theta)$$

Let a real label of one signal be built as a BPA $\mathcal{L}_r$. If the signal belongs to class C, the $\mathcal{L}_r$ can be defined as:

$$\mathcal{L}_r : \; \mathcal{L}_r(\{\mathcal{S}_1\}), \mathcal{L}_r(\{\mathcal{S}_2\}), \dots, \mathcal{L}_r(\{\mathcal{S}_C\}) = 1, \dots, \mathcal{L}_r(\{\mathcal{S}_1, \mathcal{S}_2\}), \dots, \mathcal{L}_r(\Theta)$$

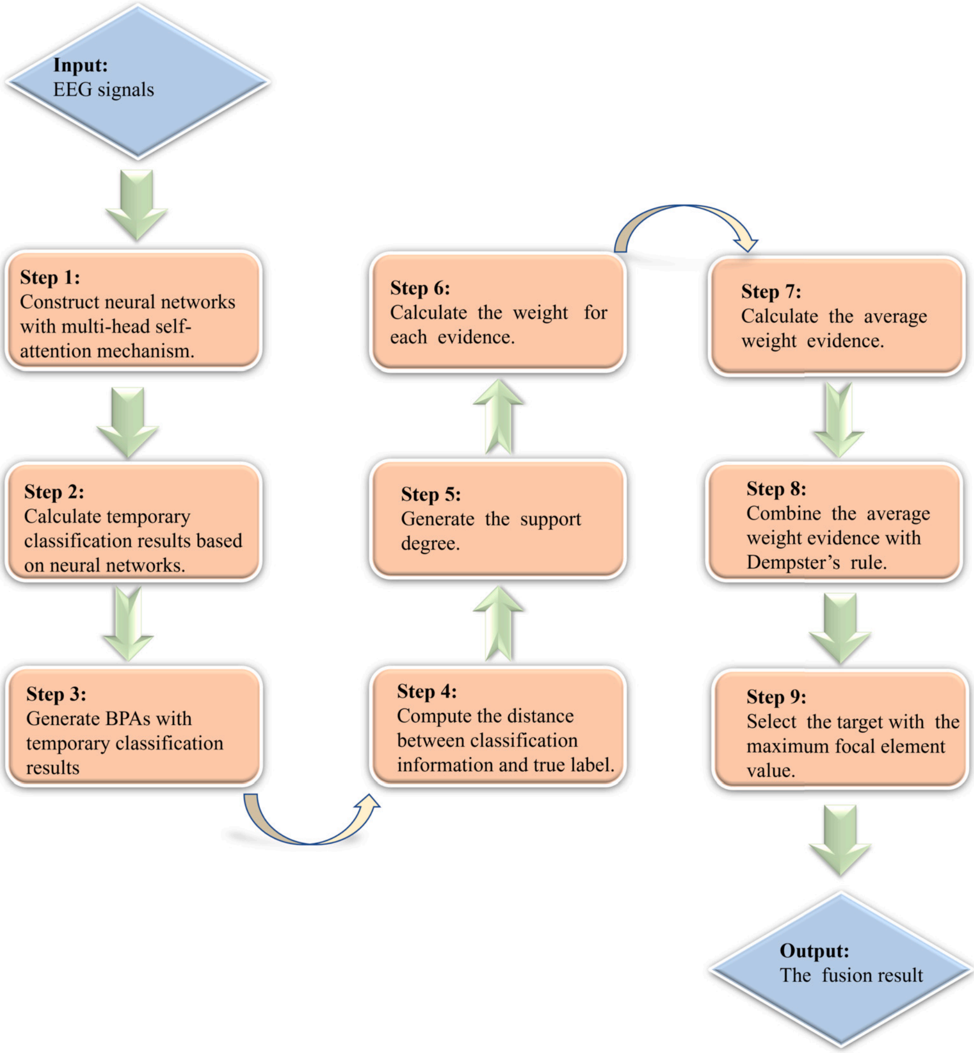


Fig. 4. Flowchart of the ETNN model for multisource information fusion.

After each neural network, several types of classification results can be got. Then, the distance between m_i and $\mathcal{L}_r$ can be calculated. Next, the support degree of each BPA is obtained based on the inverse proportional function. After that, the weight of each BPA is received based on the support degree. Finally, the weighted evidence is accessible and fused with Dempster's rule. Here, Fig. 4 is put forward to show the process of model implementation.

Step 1. Suppose that there are n types of neural network. Then, after all neural networks, n BPAs as $m_1, m_2, \ldots, m_i, \ldots, m_n$ can be generated as follows:

$$
\begin{aligned}
m_1 &: m_1(\{S_1\}), m_1(\{S_2\}), \ldots, m_1(\{S_i\}), \ldots, m_1(\{S_1, S_2\}), \ldots, m_1(\Theta),\\
m_2 &: m_2(\{S_1\}), m_2(\{S_2\}), \ldots, m_2(\{S_i\}), \ldots, m_2(\{S_1, S_2\}), \ldots, m_2(\Theta),\\
&\vdots\\
m_i &: m_i(\{S_1\}), m_i(\{S_2\}), \ldots, m_i(\{S_i\}), \ldots, m_i(\{S_1, S_2\}), \ldots, m_i(\Theta),\\
&\vdots\\
m_n &: m_n(\{S_1\}), m_n(\{S_2\}), \ldots, m_n(\{S_i\}), \ldots, m_n(\{S_1, S_2\}), \ldots, m_n(\Theta).
\end{aligned}
\tag{14}
$$

Each different neural network can produce a different classification information.

Step 2. Then, in order to measure the discrepancy between classification information and true label, the distance $Dis(m_i, \mathcal{L}_r)$ between m_i and $\mathcal{L}_r$ can be calculate as:

Algorithm 1: Information fusion algorithm for EEG signal classification.

Input: A set of EEG signal classification information;
Output: Fusion result $\mathcal{F}(\bar{m})$
for $i = 1; i \leq n$ **do**
  Calculate temporary classification results $m = \{m_1, \ldots, m_i, \ldots, m_n\}$ based on ETNN model;
end
for $i = 1; i \leq n$ **do**
  for $j = 1; j \leq 2^n$ **do**
    Construct focal element relation matrix;
  end
  Compute the distance $Dis(m_i, \mathcal{L}_r)$ between m_i and $\mathcal{L}_r$ using Eq. (15);
end
for $i = 1; i \leq n$ **do**
  Generate the support degree $sup(m_i)$ of m_i using Eq. (16);
end
for $i = 1; i \leq n$ **do**
  Calculate the weight $\omega(m_i)$ of m_i using Eq. (17);
end
for $i = 1; i \leq n$ **do**
  Construct the average weighted evidence $\bar{m}(E)$ using Eq. (18);
end
for $i = 1; i \leq n - 1$ **do**
  Obtain the fusion result $\mathcal{F}(\bar{m})$ using Eq. (19);
end
for $i = 1; i \leq n$ **do**
  Select the target with the maximum focal element value;
end

$$Dis(m_i, \mathcal{L}_r) = \sqrt{\frac{1}{2}\left(\vec{m}_i - \vec{\mathcal{L}}_r\right)^{\mathrm{T}} \underline{\underline{D}} \left(\vec{m}_i - \vec{\mathcal{L}}_r\right)}. \tag{15}$$

Here, $\underline{\underline{D}}$ is an $2^N \times 2^N$ matrix with elements as:

$$D(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

where $A, B \in 2^\Theta$, which means that A and B are focal elements of discernment Θ. Besides, $|\cdot|$ represents the cardinality of focal elements.

Step 3. According to the discrepancy measurement above, the more distant the evidence, the less support it should have. So, the support degree of each BPA $sup(m_i)$ is obtained based on inverse proportional function:

$$sup(m_i) = \frac{1}{Dis(m_i, \mathcal{L}_r)}. \tag{16}$$

In this case, the more representative neural networks can be found.

Step 4. Based on different support degrees, the greater the support, the more important the evidence. Hence, the weight $\omega(m_i)$ for each evidence can be calculated as:

$$\omega(m_i) = \frac{sup(m_i)}{\sum_{i=1}^{n} sup(m_i)}. \tag{17}$$

Step 5. To put all the evidence together reasonably, the average weighted evidence $\bar{m}(E)$ can be calculated as:

$$\bar{m}(E) = \sum_{i=1}^{n} \omega(m_i) \cdot m_i(E), \quad E \subseteq \Theta. \tag{18}$$

Step 6. Here, the final combination classification information $\mathcal{F}(m)$ can be obtained by fusing weighted evidence $n - 1$ times with Dempster's rule of combination:

$$\mathcal{F}(\bar{m}) = (\bar{m}(E) \oplus \bar{m}(E) \oplus \ldots \bar{m}(E))_{n-1}. \tag{19}$$

By means of fusing all the evidence, the uncertainty produced by neural networks can be efficiently addressed.

Step 7. According to the combination results in $\mathcal{F}(\bar{m})$, select the target with the maximum focal element value, which means the most likely outcome.

Here, Algorithm 1 shows the pseudo-code of information fusion algorithm for EEG signal classification.

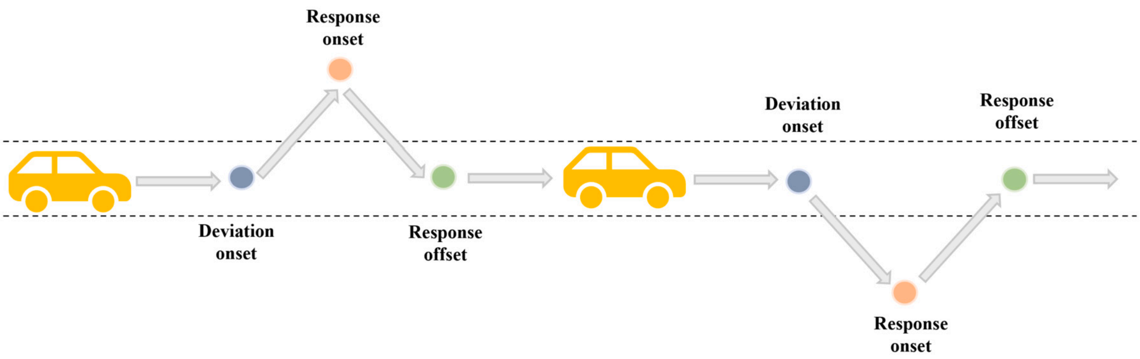


Fig. 5. Lane-departure events.

4. Application

In this section, an application based on a long-term attentive driving experiment is put forward to demonstrate the effectiveness of the ETNN model for ERP signal classification, which may reflect the alertness of drivers.

4.1. Material of multi-channel electroencephalography signal

Electroencephalography is a well-performed method to monitor human brain electrophysiology while they are involved in a real-world task with natural movements.

In this study, 30 subjects participated in a sustained-attention driving experiment [44]. And a VR environment with a dynamic simulator is provided to simulate real driving environments. Each driving task lasts for 90 minutes, which contains several random lane-departure events. Lane-departure causes the car to deviate from its original route toward the right or left sides (deviation onset). Then, participants are suggested to steer the wheel (response onset) to move the car back to the original route (response offset). Fig. 5 shows a lane-departure event.

Multi-channel EEG signal can be obtained by using wired EEG cap with 32 Ag/AgCl electrodes that include 30 EEG electrodes and 2 reference electrodes. It contributes to many fields, especially in detecting event-related potential (ERP) which illustrates the changes of potential when there is stimulus caused by movement events. In this case, deviation onset can be regarded as ERP-negative and response onset can be regarded as ERP-positive.

This experiment is performed in accordance with recommendations in the Guide for the Committee of Laboratory Care and Use of the National Chiao Tung University, Taiwan. And the study is approved by the Institutional Review Board of the Veterans General Hospital, Taipei.

Then, in this section, the pre-processing data-set is used, which can be downloaded from the publicly accessible repository of figshare. There are two main files are used in this study, including event file and data file. The type of events are classified as deviation onset (mark: 251 or 252), response onset (mark: 253) or response offset (mark: 254). Note that subjects can only response after the deviation onset. In this case, there are ERP-negative signal before the time of deviation onset and ERP-positive signal before the time of response onset. A data file contains 30-channel EEG data signal in a whole task with 500 recordings in one second.

4.2. Data processing

Because it records the whole task within 90 minutes, a large amount of data will be produced. It is suggested to select characteristic data to train and test on our model. So, the data is processed by following steps:

Step 1. Feature data extraction:

The ERP-negative and ERP-positive signal are suggested to be implied into this study. So, the EEG signal data before event 251 and event 252 are taken into account to be ERP-negative signal, as there is no movement of subjects. The EEG signal data before event 253 are take into consider to be ERP-positive signal, as there will be changes in ERP signal before subjects steer the wheel. The number of both signal are the same.

Step 2. Dimension unified:

In order to fit the model, data dimensions should be unified. Suppose that one can respond to an emergency event in less than one second while concentrating. So, the EEG data in the first second before the event is chosen. As 500 pieces of 30 channels of EEG data can be recorded in one second, the shape of each sample is 500×30.

Step 3. Outlier handling:

Many psychological experts consider that a person can physiologically make a response to an event in at least 0.1 seconds. However, the reaction time of a person who is fatigued will be greatly prolonged, about 1 second. In this case, some response onset events whose elapsed time between the previous event is less than 0.1 second or more than 1 second should be regarded as outliers. Among the data that be filtered out, these outliers should be eliminated to avoid a negative impact on model results.

Step 4. Tagging:

Each sample should be labeled for training or testing. The ERP-negative signal are labeled as:

$$\mathcal{L}_r : \mathcal{L}_r(positive) = 0, \mathcal{L}_r(negative) = 1, \mathcal{L}_r(positive, negative) = 0.$$

And ERP-positive signal are labeled as:

$$\mathcal{L}_r : \mathcal{L}_r(positive) = 1, \mathcal{L}_r(negative) = 0, \mathcal{L}_r(positive, negative) = 0.$$

Step 5. Data partitioning:
Here, 70% samples are randomly selected as training sets and the remaining 30% samples are taken as test sets.

4.3. Implementation of the proposed method

Here, data will be processed by using the ETNN model with three specific types of neural networks mentioned above.

Step 1. As for a specific ERP-positive signal sample, based on Eq. (14), suppose that three types classification information are got after three neural networks as follows:

$$m_1 : m_1(positive) = 0.75, m_1(negative) = 0.25,$$

$$m_2 : m_2(positive) = 0.83, m_2(negative) = 0.17,$$

$$m_3 : m_3(positive) = 0.44, m_3(negative) = 0.56.$$

Step 2. Based on Eq. (15), the distance $Dis(m_i)$ between m_i and $\mathcal{L}_r$ can be calculate as:

$$Dis(m_1, \mathcal{L}_r) = 0.3535,$$

$$Dis(m_2, \mathcal{L}_r) = 0.2304,$$

$$Dis(m_3, \mathcal{L}_r) = 0.7918.$$

Step 3. Based on Eq. (16), the support degree $sup(m_i)$ can be calculated as:

$$sup(m_1) = 2.8289,$$

$$sup(m_2) = 4.3402,$$

$$sup(m_3) = 1.2629.$$

Step 4. Based on Eq. (17), the weight $\omega(m_i)$ for each evidence can be calculated as:

$$\omega(m_1) = 0.3355,$$

$$\omega(m_2) = 0.5147,$$

$$\omega(m_3) = 0.1498.$$

Step 5. Based on Eq. (18), the average weighted evidence $\bar{m}(E)$ can be calculated as:

$$\bar{m}(E) : \bar{m}(positive) = 0.7447, \bar{m}(negative) = 0.2553.$$

Step 6. The combination classification information $\mathcal{F}(\bar{m})$ can be obtained according to Eq. (19) with Dempster's rule of combination:

$$\mathcal{F}(\bar{m})(positive) = 0.9613,$$

$$\mathcal{F}(\bar{m})(negative) = 0.0387.$$

Step 7. According to the combination results in $\mathcal{F}(\bar{m})$, note that this signal are selected as a ERP-positive signal.

4.4. Comparison

In this section, three representative single neural networks (such as network A, network B and network C), recent well-known neural networks for EEG signal classification (including EEG-Inception [45] and FBCNet [46]) and several classical machine learning classification methods (including K-Nearest Neighbor (KNN), Naive Bayes and Support Vector Machine (SVM)), are used to make comparisons in this application.

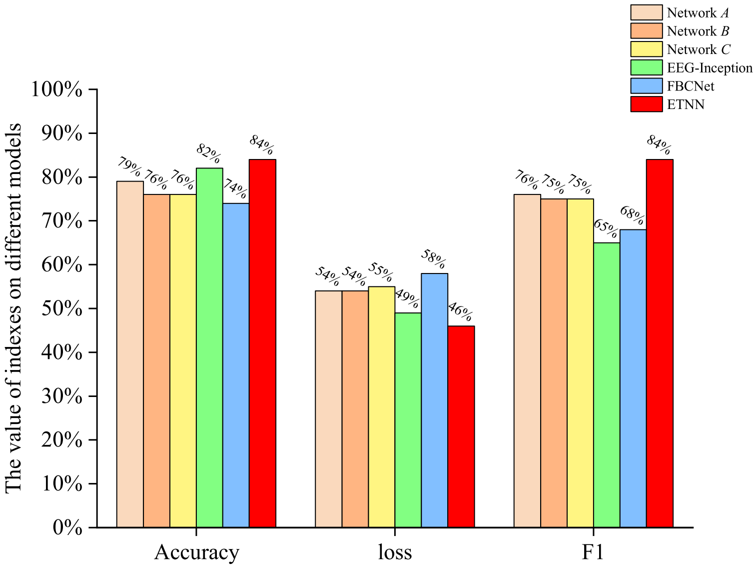


Fig. 6. The valuation indexes on different methods.

Table 2
Compared with classical machine learning methods.

Methods	Accuracy
KNN	58.73%
Naive Bayes	52.90%
SVM	54.59%
ETNN	84.05%

The cross-entropy in Eq. (13) is used to calculate the loss of ETNN model. The accuracy of test data and the F1 evaluation index are used to evaluate the model. Here, the F1 evaluation index can be defined as:

$$\mathrm{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{20}$$

From Fig. 6, note that ETNN model reaches the highest accuracy value as 0.84, which means that ETNN model performs well in ERP signal classification. Also, there is the highest F1 score in ETNN model, which show that ETNN model has the excellent results in precision and recall. Loss value represents the distance between predicted value and true value. So, the lower it is, the better the model works. There is little difference in valuation indexes among the three exemplary single neural networks. As a result, a single neural network may not be able to categorize ERP signals efficiently. According to recent well-known methodologies, EEG-Inception outperforms three representative single neural networks, although FBCNet is slightly inferior. It should be noted that the ETNN model outperforms representative neural networks and recent well-known approaches.

Several classical machine learning methods are involved to compare the performance of our proposed method. As is shown in Table 2, our proposed ETNN model shows superiority such as the highest accuracy. As for KNN, φ neighboring points were taken, where $10 \leq \varphi \leq 15$, and average classification accuracy is 0.5873, yet Naive Bayes and SVM even received worse accuracy.

4.5. The influence of components setting

In this section, several components in the ETNN model, such as the self-attention module, evidence theory module, and neural networks, are taken into account to show their influence. Moreover, the ETNN model is applied to several other EEG datasets with different distributions, which illustrates that the ETNN model is widely applicable in EEG-based classification tasks.

4.5.1. The influence of self-attention and evidence theory module

To demonstrate the effectiveness of the self-attention and evidence theory modules in the ETNN model, we analyze their impact on the classification accuracy, as shown in Table 3. The results indicate that the absence of these modules leads to the lowest accuracy of 0.74 and the highest loss of 0.55. After separately testing the self-attention and evidence theory modules, we observe a slight improvement in accuracy with the self-attention module, while the evidence theory module significantly enhances the performance of the neural network. Furthermore, the combination of both modules in the ETNN model leads to further improvements in a EEG-based classification task.

Table 3
The influence of self-attention and evidence theory modules.

Neural networks	Accuracy	loss	F1
Single neural network	0.74	0.55	0.75
Multi-type neural networks + self-attention module	0.76	0.54	0.75
Multi-type neural networks + evidence theory module	0.80	0.51	0.79
ETNN model	0.84	0.46	0.84

Table 4
Three specific combination of neural networks.

Neural networks	Accuracy
Network A + Network B + Network C	0.84
Network A + Network B + EEG-Inception	0.86
Network A + Network B + FBCNet	0.82

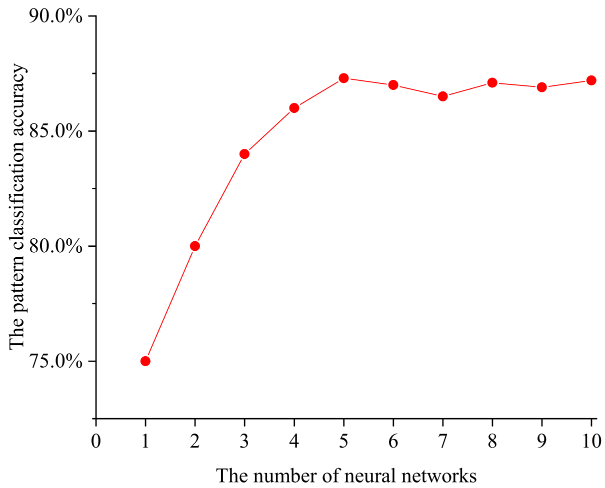


Fig. 7. The influence of the number of networks.

Table 5
The classification accuracy on different datasets based on ETNN model.

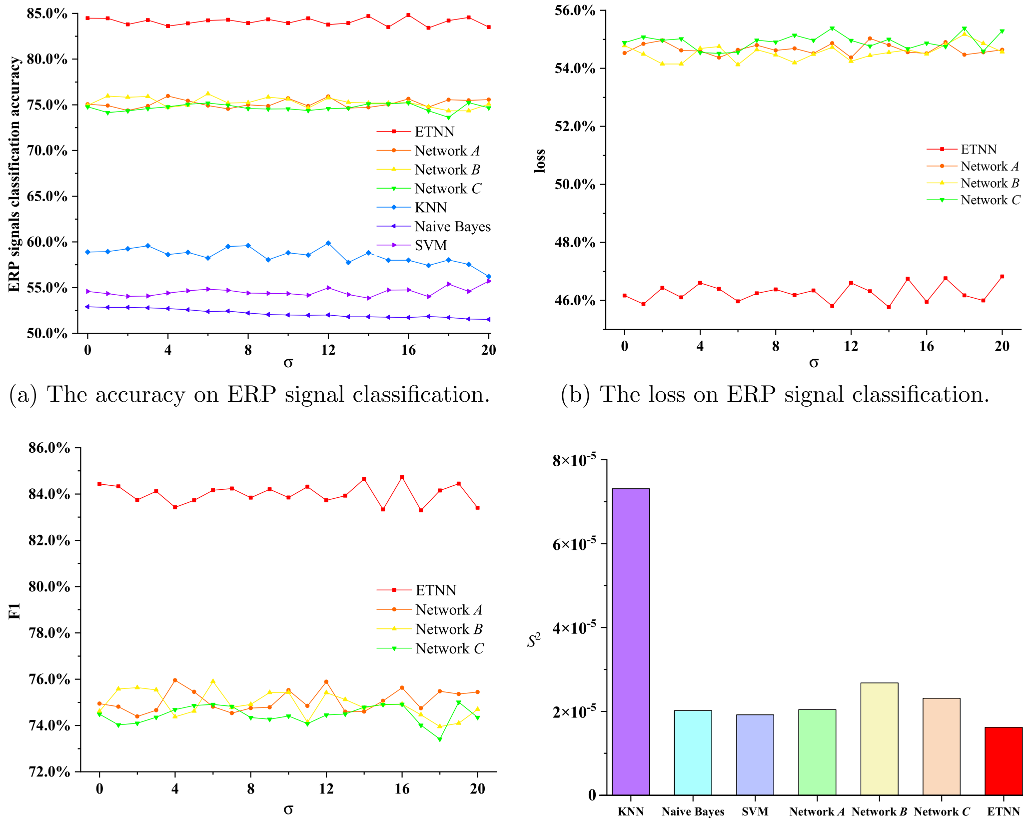
Datasets	Subjects	Classes	Channels	Time points	Accuracy
BCIC-IV-2A Data	9	4	22	1000	0.84
OpenMBI Data	54	2	20	1000	0.79
Stroke Data: A	37	2	27	1000	0.84
Stroke Data: B	34	2	27	1000	0.85

4.5.2. The influence of neural networks

To show the influence of neural networks, the qualities of networks and the number of neural networks are discussed in this section.

When considering the qualities of neural networks, the number of neural networks is fixed as three. Here, three specific combinations of neural networks are used in Table 4. It can be found that the quality of local neural networks can affect the performance of the overall model. When there are sub-networks with higher accuracy in the network, the accuracy can be improved; on the contrary, if the network quality is not good, the model will be negatively affected.

According to the information presented, networks of the same type as network A are used when considering the number of neural networks to reduce the impact of network quality. As shown in Fig. 7, the performance of the classification model improves with an increase in the number of neural networks. When the number is less than 5, there is a significant increase in accuracy, indicating that the evidence theory module can effectively handle information uncertainty and improve accuracy at these points. However, when the number of neural networks exceeds 5, the accuracy tends to stabilize because the performance of the neural network itself limits the performance of the ETNN model. In this case, the sub-networks should be improved to enhance the overall performance.



(a) The accuracy on ERP signal classification. (b) The loss on ERP signal classification.

(c) The F1 on ERP signal classification. (d) Variance of accuracy on different methods.

Fig. 8. Valuation indexes for noise sensitive analysis.

4.5.3. Application in other EEG datasets

To demonstrate the applicability of the ETNN model, additional four datasets are used, including BCIC-IV-2A Data [47], OpenMBI Data [48], Stroke Data: A [49] and Stroke Data: B [50]. These four datasets have different distributions, which means that the differences between subjects are taken into account. According to Table 5, the ETNN model performed well in the four datasets mentioned. The ETNN model can achieve a classification accuracy of about 85% on three of the datasets.

4.6. Sensitivity analysis

An experiment is carried out with the EPR-based classification task for sensitivity analysis. This experiment tends to evaluate the robustness of the ETNN model when involving noise samples.

At the beginning of experiment, data need to be noised. Original data x_i are added Gaussian noise as $X = x_i + \omega$. X is noised data and ω fits a Gaussian probability distribution:

$$f(\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\omega-\mu)^2}{2\sigma^2}\right), \tag{21}$$

where $f(\cdot)$ represents the probability density function. Besides, μ is the mean and σ is the standard deviation. In this study, μ is fixed as 0 and σ changes from 1 to 20.

Fig. 8 shows the performance of different methods based on ERP signal classification. Specifically, Fig. 8(a) demonstrates the classification accuracy. Note that the proposed neural networks reach the highest accuracy than classical machine learning, (e.g., KNN, Naive Bayes and SVM), in both low and high noise situations. In addition, it is obvious that fluctuation occurs in all methods, while $\sigma > 10$. To show the difference clearly, the variance S^2 is carried out. In Fig. 8(d), the proposed ETNN model has a lower variance which illustrates the strong robustness.

Figs. 8(b) and (c) show the loss value and F1 score, respectively. Both of them demonstrate that the ETNN model shows superiority over a single neural network in low and high noise situations. Thus, a multi-type neural network based on evidence theory performs well in ERP signal classification issues.

From the sensitive analysis above, we found that the ETNN model can achieve good stability and high accuracy in a variety of noise situations.

4.7. The discussion of computational complexity

Though the ETNN model performs admirably in EEG signal categorization, there are several limitations that must be addressed. Neural networks are known to have complicated architectures. In order to extract the characteristic of an EEG signal, the ETNN model's neural networks frequently contain complicated multi-layer structures including self-attention modules and evidence theory. Furthermore, because the ETNN model includes multi-type neural networks, more than one network will need to be trained, significantly increasing training time. Furthermore, the evidence theory module adds to the time commitment. As a result, ETNN has a higher computational cost than other single neural network models, including classical CNN, which indicates that the calculation process will be slowed. Nonetheless, quantum technology can be used in future work to alleviate the problem of high time consumption.

5. Conclusions

A novel model, named ETNN, was proposed in this paper and the structure of this model was briefly explained. It initially shed new light on the consideration of multi-head self-attention neural networks and multi-source information fusion methods based on evidence theory, which fills the gap that a single neural network may not be able to classify EEG data well. Specifically, multi-type neural networks were able to select more features of sensitive data and the multi-source information fusion method could further improve the model performance by combining the classification outputs of neural networks. According to an application in EEG-based classification task, the superiority and effectiveness were demonstrated that the ETNN model showed high accuracy and F1 score and lower loss value in ERP signal detecting, which meant that the proposed method performs well in the field of selecting features from EEG data. In addition, ETNN is better than a single neural network and classical machine learning method, including KNN, Naive Bayes and SVM. Moreover, the ETNN model also performed well in noised data, which shows strong robustness. In summary, the ETNN model provided a new and effective approach for EEG-based classification tasks. In the future study, to reduce the time consumption in neural networks, we expect to deploy feature extraction methods to be processed in advance and apply quantum techniques to the model.

CRediT authorship contribution statement

Lang Zhang: Validation, Writing-Original Draft. **Fuyuan Xiao:** Methodology, Modification. **Zehong Cao:** Modification.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors greatly appreciate the reviewers' suggestions and editor's encouragement. This research is supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), Natural Science Foundation of Chongqing, China (No. 2022NSCQ-MSX2993), and Chongqing Overseas Scholars Innovation Program (No. cx2022024). In addition, this study is partially supported by Dr. Cao's Australian Research Council (ARC) DECRA Fellowship DE220100265.

References

[1] C. Zhu, F. Xiao, Z. Cao, A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis, Inf. Sci. 605 (2022) 225–243.
[2] Z. Wang, D. Hou, C. Gao, J. Huang, Q. Xuan, A rapid source localization method in the early stage of large-scale network propagation, in: Proceedings of the ACM Web Conference, WWW-22, 2022, p. 1372.
[3] D. Meng, S. Yang, C. He, H. Wang, Z. Lv, Y. Guo, P. Nie, Multidisciplinary design optimization of engineering systems under uncertainty: a review, Int. J. Struct. Integr. 13 (4) (2022) 565–593.
[4] F. Xiao, W. Pedrycz, Negation of the quantum mass function for multisource quantum information fusion with its application to pattern classification, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2) (2022) 2054–2070, https://doi.org/10.1109/TPAMI.2022.3167045.
[5] R.R. Yager, N. Alajlan, Y. Bazi, Uncertain database retrieval with measure-based belief function attribute values, Inf. Sci. 501 (2019) 761–770.
[6] Y. Che, Y. Deng, Y.-H. Yuan, Maximum-entropy-based decision-making trial and evaluation laboratory and its application in emergency management, J. Organ. End-User Comput. 34 (7) (2022) 1–16.
[7] Y.-J. Zhou, M. Zhou, X.-B. Liu, B.-Y. Cheng, E. Herrera-Viedma, Consensus reaching mechanism with parallel dynamic feedback strategy for large-scale group decision making under social network analysis, Comput. Ind. Eng. 174 (2022) 108818.
[8] P. Liu, Y. Li, P. Wang, Consistency threshold- and score function-based multi-attribute decision-making with Q-rung orthopair fuzzy preference relations, Inf. Sci. 618 (2022) 356–378.

[9] W. Miao, J. Geng, W. Jiang, Multi-granularity decoupling network with pseudo-label selection for remote sensing image scene classification, IEEE Trans. Geosci. Remote Sens. 61 (2023) 5603813, https://doi.org/10.1109/TGRS.2023.3244565.

[10] M. Zhou, Y.-Q. Zheng, Y.-W. Chen, B.-Y. Cheng, E. Herrera-Viedma, J. Wu, A large-scale group consensus reaching approach considering self-confidence with two-tuple linguistic trust/distrust relationship and its application in life cycle sustainability assessment, Inf. Fusion 94 (2023) 181–199.

[11] H. Fujita, A. Gaeta, V. Loia, F. Orciuoli, Hypotheses analysis and assessment in counter-terrorism activities: a method based on OWA and fuzzy probabilistic rough sets, IEEE Trans. Fuzzy Syst. 28 (2020) 831–845.

[12] R. Tao, Z. Liu, R. Cai, K.H. Cheong, A dynamic group MCDM model with intuitionistic fuzzy set: perspective of alternative queuing method, Inf. Sci. 555 (2021) 85–103.

[13] Z. Wang, Z. Song, C. Shen, S. Hu, Emergence of punishment in social dilemma with environmental feedback anonymous submission, in: Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI-23, 2023.

[14] R. Zhu, Q. Liu, C. Huang, B. Kang, Z-ACM: an approximate calculation method of Z-numbers for large data sets based on kernel density estimation and its application in decision-making, Inf. Sci. 610 (2022) 440–471.

[15] Z. Cao, C.-T. Lin, K.-L. Lai, L.-W. Ko, J.-T. King, K.-K. Liao, J.-L. Fuh, S.-J. Wang, Extraction of SSVEPs-based inherent fuzzy entropy using a wearable headband EEG in migraine patients, IEEE Trans. Fuzzy Syst. 28 (1) (2019) 14–27.

[16] L. Yann, B. Yoshua, H. Geoffrey, Deep learning, Nature 521 (7553) (2015) 436–444.

[17] D. Wang, Z. Zhang, Y. Jiang, Z. Mao, D. Wang, H. Lin, D. Xu, DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism, Nucleic Acids Res. 49 (8) (2021) e46.

[18] S.A. Bagherzadeh, A. D'Orazio, A. Karimipour, M. Goodarzi, Q.-V. Bach, A novel sensitivity analysis model of EANN for F-MWCNTs–FE_3O_4/EG nanofluid thermal conductivity: outputs predicted analytically instead of numerically to more accuracy and less costs, Phys. A, Stat. Mech. Appl. 521 (2019) 406–415.

[19] D.T. Bui, M.M. Abdullahi, S. Ghareh, H. Moayedi, H. Nguyen, Fine-tuning of neural computing using whale optimization algorithm for predicting compressive strength of concrete, Eng. Comput. 37 (1) (2021) 701–712.

[20] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, in: Classic Works of the Dempster-Shafer Theory of Belief Functions, Springer, 2008, pp. 57–72.

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

[22] F. Xiao, Generalized quantum evidence theory, Appl. Intell. (2022), https://doi.org/10.1007/s10489-022-04181-0.

[23] W. Fan, F. Xiao, A complex Jensen-Shannon divergence in complex evidence theory with its application in multi-source information fusion, Eng. Appl. Artif. Intell. 116 (2022) 105362, https://doi.org/10.1016/j.engappai.2022.105362.

[24] F. Xiao, J. Wen, W. Pedrycz, Generalized divergence-based decision making method with an application to pattern classification, IEEE Trans. Knowl. Data Eng. (2022), https://doi.org/10.1109/TKDE.2022.3177896.

[25] D. Han, J. Dezert, Y. Yang, Belief interval-based distance measures in the theory of belief functions, IEEE Trans. Syst. Man Cybern. Syst. 48 (6) (2018) 833–850.

[26] Z. Liu, Y. Liu, J. Dezert, F. Cuzzolin, Evidence combination based on credal belief redistribution for pattern classification, IEEE Trans. Fuzzy Syst. 28 (4) (2020) 618–631.

[27] X. Xu, D. Zhang, Y. Bai, L. Chang, J. Li, Evidence reasoning rule-based classifier with uncertainty quantification, Inf. Sci. 516 (2020) 192–204.

[28] Q. Zhou, É. Bossé, Y. Deng, Modeling belief propensity degree: measures of evenness and diversity of belief functions, IEEE Trans. Syst. Man Cybern. Syst. 53 (5) (2023) 2851–2862, https://doi.org/10.1109/TSMC.2022.3219996.

[29] L. Xiong, X. Su, H. Qian, Conflicting evidence combination from the perspective of networks, Inf. Sci. 580 (2021) 408–418.

[30] F. Xiao, GEJS: a generalized evidential divergence measure for multisource information fusion, IEEE Trans. Syst. Man Cybern. Syst. 53 (4) (2022) 2246–2258, https://doi.org/10.1109/TSMC.2022.3211498.

[31] S. Zhang, F. Xiao, A TFN-based uncertainty modeling method in complex evidence theory for decision making, Inf. Sci. 619 (2022) 193–207, https://doi.org/10.1016/j.ins.2022.11.014.

[32] C. Yang, F. Xiao, An exponential negation of complex basic belief assignment in complex evidence theory, Inf. Sci. 622 (2022) 1228–1251, https://doi.org/10.1016/j.ins.2022.11.160.

[33] J. Wang, Z. Zhou, C. Hu, S. Tang, W. He, T. Long, A fusion approach based on evidential reasoning rule considering the reliability of digital quantities, Inf. Sci. 612 (2022) 107–131.

[34] Y. Deng, Uncertainty measure in evidence theory, Sci. China Inf. Sci. 63 (11) (2020) 210201.

[35] C. Fu, Q. Zhan, W. Liu, Evidential reasoning based ensemble classifier for uncertain imbalanced data, Inf. Sci. 578 (2021) 378–400.

[36] Z. Hua, L. Fei, H. Xue, Consensus reaching with dynamic expert credibility under Dempster-Shafer theory, Inf. Sci. 610 (2022) 847–867.

[37] R. Fang, H. Liao, A. Mardani, How to aggregate uncertain and incomplete cognitive evaluation information in lung cancer treatment plan selection? A method based on Dempster-Shafer theory, Inf. Sci. 603 (2022) 222–243.

[38] Q. Zhou, G. Tian, Y. Deng BF-QC, Belief functions on quantum circuits, Expert Syst. Appl. 223 (2023) 119885, https://doi.org/10.1016/j.eswa.2023.119885.

[39] Y. Deng, Random permutation set, Int. J. Comput. Commun. Control 17 (1) (2022) 4542.

[40] Y. Deng, Information volume of mass function, Int. J. Comput. Commun. Control 15 (6) (2020) 3983.

[41] X. Deng, Y. Cui, An improved belief structure satisfaction to uncertain target values by considering the overlapping degree between events, Inf. Sci. 580 (2021) 398–407.

[42] C. Qiang, Y. Deng, K.H. Cheong, Information fractal dimension of mass function, Fractals 30 (2022) 2250110, https://doi.org/10.1142/S0218348X22501109.

[43] F. Xiao, Z. Cao, C.-T. Lin, A complex weighted discounting multisource information fusion with its application in pattern classification, IEEE Trans. Knowl. Data Eng. (2022), https://doi.org/10.1109/TKDE.2022.3206871.

[44] Z. Cao, C.-H. Chuang, J.-K. King, C.-T. Lin, Multi-channel EEG recordings during a sustained-attention driving task, Sci. Data 6 (2019), https://doi.org/10.1038/s41597-019-0027-4.

[45] C. Zhang, Y.-K. Kim, A. Eskandarian, EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification, J. Neural Eng. 18 (4) (2021) 046014.

[46] R. Mane, E. Chew, K. Chua, K.K. Ang, N. Robinson, A.P. Vinod, S.-W. Lee, C. Guan, Fbcnet: a multi-view convolutional neural network for brain-computer interface, arXiv preprint, arXiv:2104.01233, 2021.

[47] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, et al., Review of the bci competition IV, Front. Neurosci. (2012) 55.

[48] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, S.-W. Lee, EEG dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy, GigaScience 8 (5) (2019) giz002.

[49] K.K. Ang, C. Guan, K.S. Phua, C. Wang, L. Zhao, W.P. Teo, C. Chen, Y.S. Ng, E. Chew, Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation, Arch. Phys. Med. Rehabil. 96 (3) (2015) S79–S87.

[50] K.K. Ang, C. Guan, K.S. Phua, C. Wang, L. Zhou, K.Y. Tang, G.J. Ephraim Joseph, C.W.K. Kuah, K.S.G. Chua, Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke, Front. Neuroeng. 7 (2014) 30.