Article

A Fine-Grained Approach for EEG-Based Emotion Recognition Using Clustering and Hybrid Deep Neural Networks

Liumei Zhang [1,*] , Bowen Xia [1], Yichuan Wang [2,3] , Wei Zhang [1] and Yu Han [2]

1 School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China; 21211060749@stumail.xsyu.edu.cn (B.X.); 21211060739@stumail.xsyu.edu.cn (W.Z.)
2 School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; chuan@xaut.edu.cn (Y.W.); 20212060708@stumail.xsyu.edu.cn (Y.H.)
3 Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an 710048, China
* Correspondence: zhangliumei@xsyu.edu.cn; Tel.: +86-180-9233-0186

Abstract: Emotion recognition, as an important part of human-computer interaction, is of great research significance and has already played a role in the fields of artificial intelligence, healthcare, and distance education. In recent times, there has been a growing trend in using deep learning techniques for EEG emotion recognition. These methods have shown higher accuracy in recognizing emotions when compared with traditional machine learning methods. However, most of the current EEG emotion recognition performs multi-category single-label prediction, and is a binary classification problem based on the dimensional model. This simplifies the fact that human emotions are mixed and complex. In order to adapt to real-world applications, fine-grained emotion recognition is necessary. We propose a new method for building emotion classification labels using linguistic resource and density-based spatial clustering of applications with noise (DBSCAN). Additionally, we integrate the frequency domain and spatial features of emotional EEG signals and feed these features into a serial network that combines a convolutional neural network (CNN) and a long short-term memory (LSTM) recurrent neural network (RNN) for EEG emotion feature learning and classification. We conduct emotion classification experiments on the DEAP dataset, and the results show that our method has an average emotion classification accuracy of 92.98% per subject, validating the effectiveness of the improvements we have made to our emotion classification method. Our method for emotion classification holds potential for future use in the domain of affective computing, such as mental health care, education, social media, and so on. By constructing an automatic emotion analysis system using our method to enable the machine to understand the emotional implications conveyed by the subjects' EEG signals, it can provide healthcare professionals with valuable information for effective treatment outcomes.

Keywords: brain-computer interface; CNN; DBSCAN; emotion recognition; EEG

1. Introduction

Emotion, as a cognitive domain in psychology, represents an individual's attitude toward an objective object and the corresponding neural behavioral responses within the human brain. This phenomenon is both perceptible and measurable. Initially, research on emotions primarily focused on psychology, neurology, and physiology. However, the concept of affective computing marked a pivotal shift, initiating research and development in emotion recognition within the field of artificial intelligence. In recent years, an increasing amount of research has focused on enhancing brain-computer interfaces (BCIs) by utilizing information about the user's emotional state from EEG, which is referred to as affective brain-computer interfaces [1]. This approach aims to provide machines with the capability to perceive, understand, and regulate emotions. Affective BCIs have applications in various fields, including human-computer interaction [2], mental health care, and driving fatigue

detection [3]. To establish emotional interaction between humans and computers, the key issue is how to identify human emotional states from EEG signals.

Representation of emotions is usually divided into two perspectives: the discrete perspective and the dimensional perspective. The discrete perspective analyzes emotions in such a way that each particular emotion (e.g., fear, sad, happy, etc.) is mapped to its unique environmental, physiological, and behavioral parameters [4]. From a dimensional perspective, human emotions are represented in relation to several basic dimensions. The two-dimensional representation model of emotion was first proposed by Russell [5], with valence indicating the pleasantness of the emotion, ranging from negative to positive, and arousal indicating the intensity of the emotion. Figure 1 depicts emotional valence arousal two-dimensional space model.
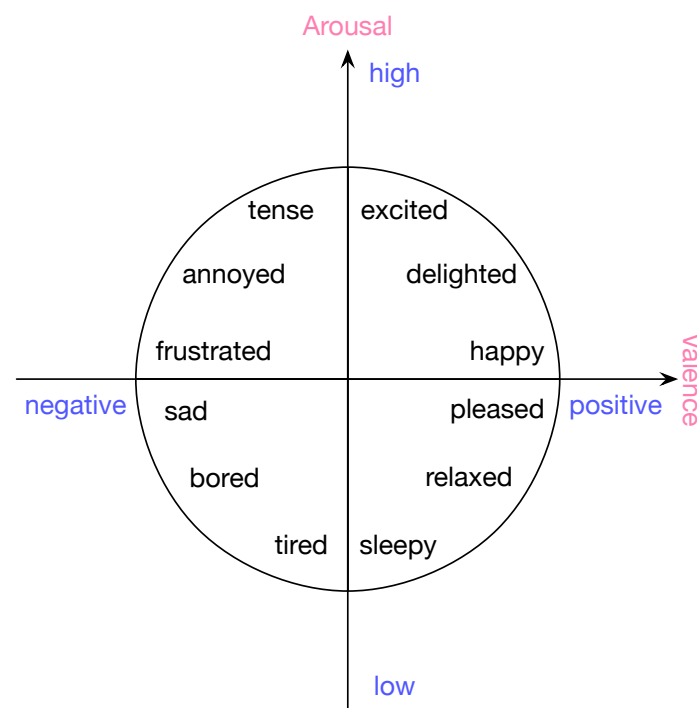
Figure 1. Emotional valence arousal two-dimensional space model.

Currently, the use of machine learning to recognize emotional states through EEG has been extensively studied. Theoretical methods based on machine learning can generally be effectively used for emotion classification. However, these methods often rely on manually extracted features as inputs, which is not only time-consuming and labor-intensive but also susceptible to information loss. In addition, EEG signals are susceptible to noise interference during data acquisition, have low signal-to-noise ratios, and exhibit asymmetry and instability over time. These characteristics pose a challenge for traditional machine learning methods that rely relatively on manual feature extraction and prior knowledge. As a result, deep learning algorithms have been applied to EEG emotion recognition, allowing for the automatic learning of high-level features from the data. These methods not only have high recognition accuracy, but also propose new solutions in feature extraction. In this paper, we first extract frequency domain features from emotional EEG signals within frequency bands that have strong ties to emotions. These features are then organized into a three-dimensional structure, accounting for the spatial relationships between electrodes. CNN is employed to learn frequency information and the correlation between electrodes from three-dimensional input structure, while LSTM is used to learn temporal information from CNN outputs, aiming to improve the accuracy of emotion recognition based on EEG data.

Commonly open datasets for affective computing based on EEG include DEAP [6], SEED [7], DREAMER, and so on. DEAP and SEED are the two most widely used EEG-emotion databases. Currently, Most of the studies using the DEAP dataset are conducted

with binary classification based on the valence-arousal two-dimensional spatial model. This simplification overlooks the complexity of human emotions. To address this issue, we propose a new method for building emotion classification labels. Our method aims to perform more fine-grained and practical emotion recognition and analysis to achieve more accurate emotional computation intelligence. First, we use the linguistic resource WordNet-Affect [8] to select labels resembling emotional descriptions. Then, we use the extended Affective Norms for English Words database [9] to embed these labels into an affective arousal space. Finally, introduce how to use DBSCAN clustering to map labels from EEG trials to specific emotions.

In summary, The main contributions of this paper are as follows:

(1) We propose a new method for building emotion classification labels. In contrast to other research works, instead of only using EEG signal data and two emotion labels for binary classification, our method use linguistic resource and DBSCAN clustering can achieve finer-grained emotion recognition, and the method can recognize six categories of emotions.

(2) We organize the three-dimensional feature structure of emotional EEG signals and combine the proposed new method of building emotion classification labels with the hybrid deep learning model CNN and LSTM. The advantage of the model is that it can learn frequency band, spatial and temporal features in the three-dimensional feature structure. Experimental results show that we achieve state-of-the-art performance on the DEAP dataset.

The paper is structured as follows: In Section 2, we provide an overview of related studies. Section 3 presents our proposed framework for emotion classification and details our research approach. Section 4 is an experimental description, and finally, in Section 5, we summarize the paper with our conclusions.

2. Related Work

Research on emotion recognition using EEG signals involves several key steps, including EEG signal acquisition, preprocessing, emotion related feature extraction, and emotion recognition modeling. Among these steps, the latter two re considered critical technical stages in this research field. We provide a review of the related work in these areas, and additionally, we review the basis for emotion classification when constructing recognition systems in emotional quantification models.

2.1. Emotion Related Feature Extraction

The feature extraction stage aims to extract valuable information from EEG signals for further analysis. Typically, EEG feature extraction techniques are based on the frequency domain, time domain, and time-frequency domain. The time-domain features are the most intuitive and easy to obtain, mainly including event related potential (ERP), signal statistics, higher order crossings (HOC), fractal dimension (FD), and so on. Time-domain feature computation directly utilizes the preprocessed time-series data, which is considered high-performing, however, it lacks the ability to capture frequency information in the signal. Consequently, researchers have incorporated frequency domain analysis. By utilizing the fast Fourier transform (FFT), features can be extracted from the frequency domain, a process that decomposes the original EEG signal into multiple frequency bands associated with emotional states. For example, Pleasant emotions are associated with an increase in the frontal midline (Fm) theta frequency band power in the brain [10].During experiences of anger, there were specific enhancements in power within the beta-2, beta-3, and gamma frequency bands in the frontopolar (Fp1, Fp2) electrodes on both sides of the brain [11]. After the original signal is transformed into the frequency domain, EEG features are then computed from it. One of the widely used features is obtained from the power spectrum of the EEG signals [12]. Zheng et al. conduct a systematic evaluation of six popular features and electrode combinations, including power spectral density (PSD), differential entropy (DE), differential asymmetry (DASM), rational asymmetry (RASM),

asymmetry (ASM), and differential cumulative (DCAU), and find that methods utilizing the DE feature outperform others [13]. The scope of the Fourier transform encompasses the entire time domain, making it unable to determine the corresponding moments in time for various frequency components of non-stationary signals. Therefore, the concept of time-frequency domain was introduced. The typical approach involves dividing the signal into multiple time windows, each containing approximately stationary sub-signals, and then transforming them into the frequency domain to obtain a set of frequency-domain features. This can be accomplished using techniques like short-time Fourier transform (STFT) and discrete wavelet transform (DWT) [14].

2.2. Emotion Recognition Modeling

Similarly, various classification methods have been proposed for the classification of emotional EEG data. These methods include machine learning and ensemble learning, which have demonstrated good classification performance in emotion recognition, such as Bayesian [15], SVM [16], Random forest [17], etc. With the continuous development of high-performance computing technologies, deep learning has achieved significant breakthroughs in fields such as computer vision and multimedia learning, this trend has also extended to the field of EEG emotion recognition. Compared to traditional methods, deep learning demonstrates more powerful automatic feature learning and generalization capabilities. Deep belief network (DBN), CNN and RNN are the most commonly used deep learning techniques for emotion recognition tasks, followed by multilayer perceptron neural network (MLPNN) [18]. CNN stands out among various network algorithms due to its outstanding capability in extracting features through convolutional kernels, it possesses a significant advantage in efficiently handling the generation of relevant graph features from EEG data. Russo [19] use topographic maps (TOPO-FM) and holograms (HOLO-FM) based on EEG signal features as representations, employing a CNN to mine feature information, and then fed the fused features into a SVM classifier. Through experiments on four public datasets, the research results show that the proposed method can improve the accuracy of emotion recognition. Chen et al. [20] present a convolutional neural network approach for learning and classifying EEG emotion features. Their approach is based on time and frequency domain features and includes models like CVCNN, GSCNN, and GSLTCNN. The best-performing method achieved accuracy rates of 84.02% for valence and 88.51% for arousal dimensions. Another more effective approach is to utilize RNN to effectively capture the temporal correlations and dynamic changes in EEG time-series signals. Sharma et al. [21] use nonlinear higher-order statistics and LSTM to automatically classify emotionally labeled EEG signals. The accuracy is estimated using a 10-fold cross-validation approach, the accuracy for the four-class emotion classification task is 82.01%, which is an improvement of 6.8% compared to the traditional SVM classifier. Li et al. [22] employ BILSTM to learn temporal features and incorporate an attention mechanism into the network framework. There is also research combining CNN and RNN to create a C-RNN model for emotion recognition, achieving an average accuracy of 84.44% on the SEED database [23].

2.3. Emotion Classification Basis

Most studies typically focus on designing features and optimizing algorithms for EEG signals to achieve accurate recognition but often overlook the complexity of emotions. This trend is particularly common in experiments conducted using the publicly available DEAP dataset, where researchers primarily concentrate on the valence and arousal dimensions and uses threshold division to conduct binary classification experiments. Similarly, experiments on the SEED dataset tend to categorize emotions simply as positive, neutral, and negative. A few studies involve assigning specific emotion category labels to samples and conducting fine-grained emotion classification. Hasan et al. [24] propose an 8-class emotion classification method, although they conduct a finer division of the emotional space, they do not explain the correspondence between the 8 ranges obtained after thresh-

old division of valence and arousal and specific emotions. Moise et al. [25] utilize the correspondence between the VAD model provided by Russell and Mehrabian and the discrete emotion model. They perform interval division on the dimensions of valence, arousal, and dominance, allowing for the classification of six basic emotions. To overcome this limitation, we expand the recognition task into the field of multi-label learning and introduce a six-category technique capable of recognizing six different emotions, including joy, liking, fear, sadness, dislike, and despair.

3. Method

3.1. EEG Emotion Recognition Pipeline

The general flow of our work is shown in Figure 2. The project is divided into four modules. First, the raw signal is collected from the standard 10–20 system [26], then preprocessed, and the EEG trials are segmented without overlapping, with each segment assigned the label of the original trial. The second module is build emotion classification labels, we cluster the raw trial labels using DBSCAN and combine them with the linguistic resource WordNet-Affect to map the sentiment dimension model to the discrete model. The third module is feature extraction, this module involves performing band decomposition on each segment, extracting DE features from these segments, and organizing these features into three-dimensional structures. The features are then fed into the final module, which is the classification stage. It includes CNN and RNN structures with LSTM units to further learn the spatial and temporal information and utilizes a softmax classifier for emotion recognition. Finally, the output consists of classified discrete emotions.
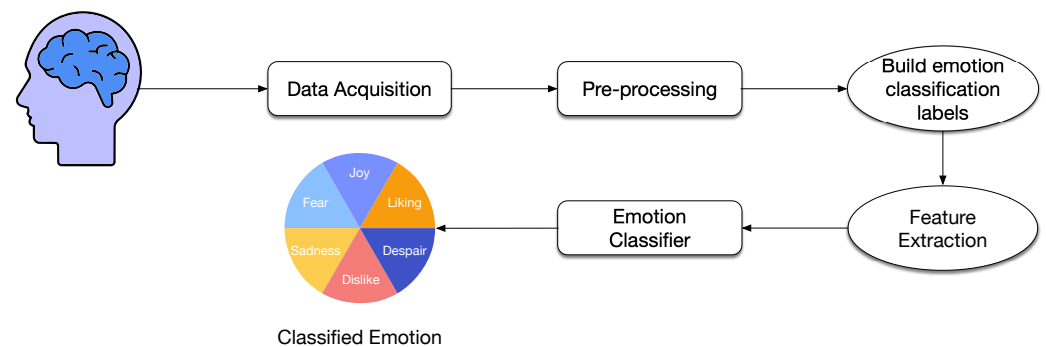

Figure 2. Workflow of the proposed method for emotion recognition.

3.2. Pre-Processing

The raw signal is downsampled to a sampling rate of 128 Hz. Since eye movements are the main source of noise, eye artifacts are removed using blind source separation algorithms. A bandpass filter is applied to remove noise smaller than 4 Hz or larger than 45 Hz. EEG data is averaged to the common reference. Additionally, the data is segmented into 60-s trials, with the first three seconds as a pre-trial baseline signal that needs to be removed. Furthermore, for the purpose of increasing the amount of training data, we segment the EEG signal into 2-s time windows without overlapping.

3.3. Build Emotion Classification Labels

In a two-dimensional emotion recognition system, human emotions are recognized using only primary data and two emotion labels (valence and arousal). By setting 5 as the threshold for emotion classes, label values are categorized into two groups: one class includes values from 1 to 4.99, and the other class includes values from 5 to 8.99. The result of the classification is four composite emotion states: HAHV, HALV, LALV, and LAHV, as shown in Figure 3, with multiple real-life emotions are present in each state. Therefore, this binary classification does not accurately recognize real-life emotions because it only uses four emotion states.
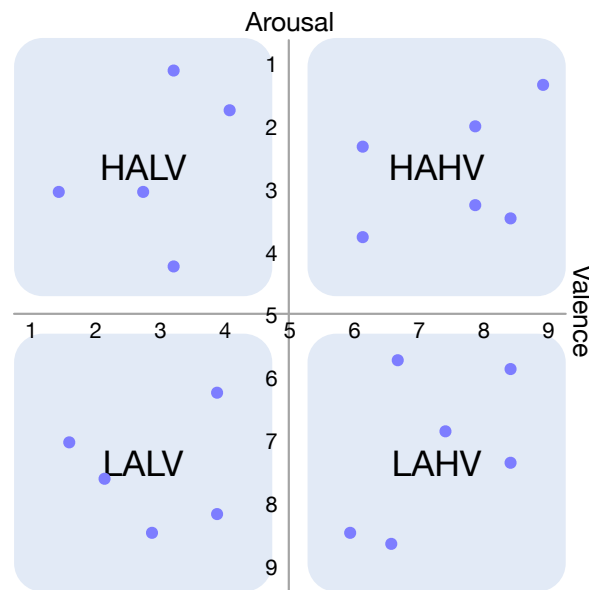
Figure 3. The distribution on the arousal-valence plane for the four compound emotional states.

In order to map EEG trials label to specific emotion words, we first utilize the linguistic resource WordNet-Affect to select labels that resemble emotional descriptions. WordNet-Affect is developed from WordNet by selecting and tagging a subset of synonyms representing the meaning of emotions. Strapparava and Valitutti [8] assign one or more emotion labels (a-labels), which help to express the meaning of an emotion accurately, to multiple WordNet synonym sets. For example, affective concepts representing emotional states are personalized by synonyms tagged with the a-label EMOTION. These labels are hierarchically organized, and within the synonyms tagged with EMOTION, 32 groups of emotion words representing emotional states are further specified. Table 1 provides the emotion categories within EMOTION.

Table 1. Categories in a-label EMOTION.

EMOTION(a-label)	Category
positive	joy, love, affection, liking, enthusiasm, gratitude, self-pride, levity, calmness, fearlessness, positive-expectation, positive-fear, positive-hope
negative	negative-fear, sadness, general-dislike, ingratitude, shame, compassion, humility, despair, anxiety, daze
neutral	apathy, neutral-unconcern
ambiguous	thing, gravity, surprise, ambiguous-agitation, ambiguous-fear, pensiveness, ambiguous-expectation

Then, we utilize the dataset published by Warriner et al. [9]. This dataset contains nearly 14,000 English words with their embedding in valence-arousal plane. We use it to embed the previously selected labels into the valence-arousal plane. However, we noticed that there is a certain degree of overlap in the distribution of these 32 groups of emotional words on the valence-arousal plane, which may be due to the high correlation between some emotion categories. Therefore, we combined the distribution of emotional words and the six basic emotions identified by Paul Ekman [27], namely anger, disgust, fear, joy, sadness, and surprise. From these 32 groups of words, we select six groups of emotion words that best matched the basic emotions and are more clearly distributed in the data. These words are "joy", "liking", "fear", "sadness", "disliking", and "despair", "fear" and "dislike" correspond to the original categories "negative-fear" and "general-dislike", respectively.

We project the combination of the two dimensional representations of valence and arousal onto the affective concepts representing emotional states by the following steps:

Step 1: Calculate the cluster centers of the six groups of emotion words we selected representing emotional states. These six groups of emotionwords naturally form clusters. By applying the concept of determining cluster centers from the K-means algorithm, we find the cluster centers for all associated points by taking the average coordinates of all data points within each of the six clusters.

Step 2: Cluster the labels of EEG trials and calculate the corresponding cluster centers. Due to the complex distribution of label data and an uncertain number of clusters, the DBSCAN algorithm is effective in discovering clusters with irregular shapes and can adapt to complex data distributions. Therefore, we have chosen the DBSCAN algorithm for label clustering.

Step 3: Calculate the similarity between the cluster center of each EEG trials label cluster and the center of the cluster of emotion words representing emotional states. This way, we can determine which emotion the EEG trials label most closely matches. Therefore, for a given point in the EEG trials label data, we can always find the emotion point that is closest to it.

3.4. Feature Extraction and Emotion Classifier

As shown in Figure 4, the pre-processed EEG data needs to be operated with band decomposition and the calculation of frequency domain feature, these features are then organized into a 3D structure and fed into the deep learning model.
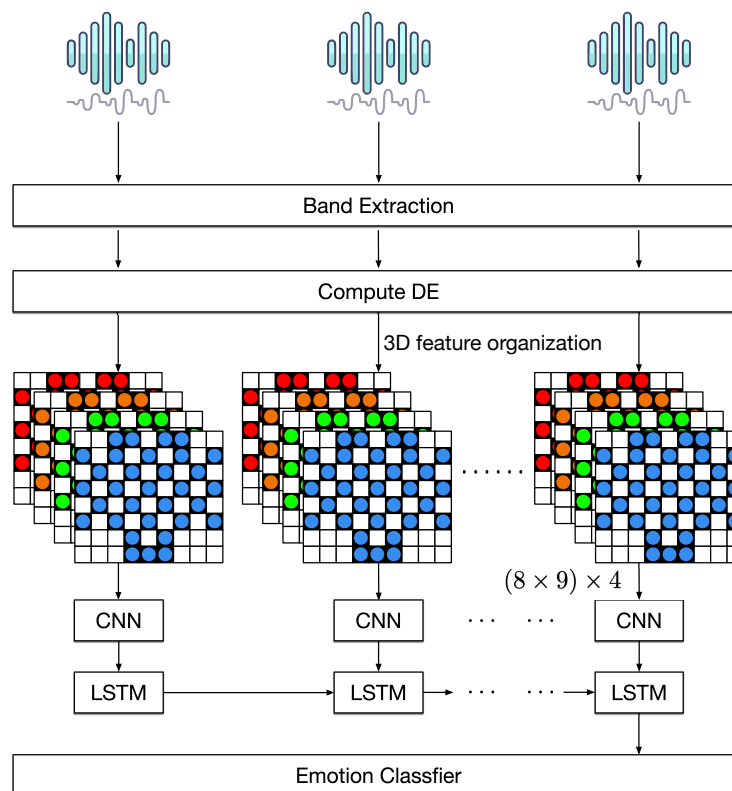
Figure 4. Framework of feature extraction and deep neural network for emotion classification.

3.4.1. Feature Organization

For each EEG segment, we decompose it into four frequency bands including θ (4–8 Hz), α (8–14 Hz), β (14–31 Hz) and γ (31–45 Hz), by employing a Butterworth filter. And the DE features are extracted from each frequency band with a 0.5 s window.

Differential entropy is a generalized form of Shannon's information entropy over continuous variables, which is defined as:

$$DE = -\int_a^b p(x)\log(p(x))dx \tag{1}$$

$p(x)$ denotes the probability density function of continuous information, $[a, b]$ denotes the interval over which the information takes values. For a specific length of a segment that approximately obeys the Gauss distribution $N(\mu, \sigma_i^2)$ of EEG, its differential entropy is:

$$DE = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}}\right) dx = \frac{1}{2}\log\left(2\pi e\sigma_i^2\right) \tag{2}$$

In order to maintain the information concealed in the electrode layout, we organize the DE features into 2D maps. The International 10–20 System has 62 electrodes, and we convert the system planes into compact 2D maps (height of 8, width of 9). Since the data we use are obtained from 32 electrode acquisitions, the unused electrode positions in the 2D maps will be populated with zeros. Finally, these planes for all bands are stacked into a 3D EEG cube.

3.4.2. Continuous Convolutional Netural Network

The core of the convolutional neural network is the convolutional layer, which through forward propagation enables different convolutional kernels to operate with the input feature maps, thus outputting different feature maps. The convolution formula is as follows:

$$Y_n = \sum_{i=1}^{M}\left[\left(W_n^i * x_i\right) + b_n\right] \tag{3}$$

x_i denotes the input feature map, W_n^i is the nth convolution kernel on channel i, b_n is the bias and Y_n is the nth output feature map, M denotes the constant of the total number of channels.

The main role of the pooling layer is downsampling, which further reduces the number of parameters by removing some unimportant samples. Since the size of the 2D graph organized by DE features is very small, it is best to retain all the information. Therefore, we do not add pooling layer after all the convolutional layers. In convolutional neural networks, each layer uses an activation function to nonlinearly represent the output of the previous layer, and commonly used activation functions include TanHyperbolic function, Relu function, sigmoid function, and softmax function. For each convolutional layer, we use the Relu function [28] for each convolutional layer due to its simplicity in implementation, acceleration of computation and convergence, absence of saturation issues, and significant reduction in gradient dissipation.

In this work, we use a CNN model consisting of four consecutive convolutional layers, a maximum pooling layer, and a fully connected layer. Specifically, the first convolutional layer has 64 feature maps and the kernel size is set to 5 × 5. The feature maps are doubled in the next two convolutional layers, so that there are 128 and 256 feature maps in the second and third layers, respectively, with a kernel size of 4 × 4. In order to fuse the feature maps of the previous convolutional layers, a convolutional layer with 64 feature maps with a kernel size of 1 × 1 is added. After 4 consecutive convolutional layers, a maximum pooling layer with a size of 2 × 2 and a step size of 2 is added. Finally, the final features are converted to one-dimensional vectors and connected to a fully connected layer.

3.4.3. LSTM Recurrent Neural Networks

In the DEAP experiment, the stimulus intensity varies during a one-minute duration. It remains uncertain which specific moment during this period significantly influences the final evaluation of the subjects, therefore, we need to model the contextual information

of the temporal signals. LSTM can be a good solution to this problem due to its recursive structure in time.

LSTM is an improved RNN combined with the "gate" mechanism [29], the specific working mechanism is as follows:

The first gate is the forgetting gate, which determines how much of the previous moment's cell state C_{t-1} is saved to the current moment's cell state C_t. The hidden state h_{t-1} from the previous moment and the input x_t from the current moment are concatenated into a new feature vector, which is multiplied by the weight parameter W_f and fed into the sigmoid activation function. The vector f_t is used as a decision vector to determine how much of the cell state C_{t-1} from the previous moment has been added to the cell state C_t. The computation of f_t is shown as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4}$$

The second gate is the input gate, which determines how much of the input x_t of the cell at the current moment is saved to the cell state C_t at the current moment. tanh activation function determines the candidate information $\bar{C}_t$ at the current moment, and the decision vector I_t determines how much of the $\bar{C}_t$ has been added to the cell state C_t by performing an elementwise multiplication operation with the candidate information $\bar{C}_t$. The calculations of I_t and $\bar{C}_t$ are shown as follows, respectively:

$$I_t = \sigma(W_I \cdot [h_{t-1}, x_t] + b_I) \tag{5}$$

$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

The cell state C_t at the current moment is calculated as shown as follows:

$$C_t = C_{t-1} \times f_t + \bar{C}_t \times I_t \tag{7}$$

The last gate is the output gate, which determines how much of the current moment's cell state C_t is input to the cell's hidden state h_t. The decision vectors o_t and h_t are computed as follows, respectively:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = \tanh(C_t) \times o_t \tag{9}$$

The W and b terms are the weight parameters and bias terms corresponding to the activation function respectively.

4. Experiment

4.1. Experiment Setup and Datasets

We implement CNN and LSTM using Keras and train them on NVIDIA Tesla P100-PCIE-12GB GPUs. DBSCAN is implemented using scikit-learn. We use the DEAP dataset as a source of brain signals. In the data collection experiments, physiological signals, such as EEG, ECG, and EMG, induced by 32 subjects watching 40 music videos of about 1-min duration with different emotional tendencies are detected and recorded. Their EEG signals are collected using a 32-channel Biosemi ActiveTwo device, following the International 10–20 system. After 20 video trials, subjects are allowed to take a short break. Each trial last 63 s, comprising 60 s of video and a 3-s pretrial baseline. After each video trial, subjects use the Self-assessment manikins [30] to self-assess their level of arousal, valence, dominance, and familiarity using a continuous rating scale ranging from 1 to 9, determining whether the corresponding emotion is correctly triggered. Rating values from low to high indicate the indicators from negative to positive and from weak to strong respectively.

4.2. EEG Label Processing Based on DBSCAN

Since the distribution of tags has no obvious clusters and is scattered, as shown in Figure 5, predefining the number of clusters becomes challenging. DBSCAN can automatically cluster the features and effectively recognize clusters with arbitrary shapes while filtering out anomalies. Additionally, DBSCAN has relatively low time complexity. Therefore, we use DBSCAN to cluster the labels of EEG trials. The main idea of DBSCAN is to identify clusters based on the density of data points. For each observation in the clustering process, the Eps neighborhood of a given radius must contain at least MinPts observations. Clusters are then formed by transitively expanding core points. Points that have insufficient neighboring points to reach the MinPts threshold are considered non-core and are still assigned to a cluster. Data points that are neither core objects nor border points are not assigned to any cluster.
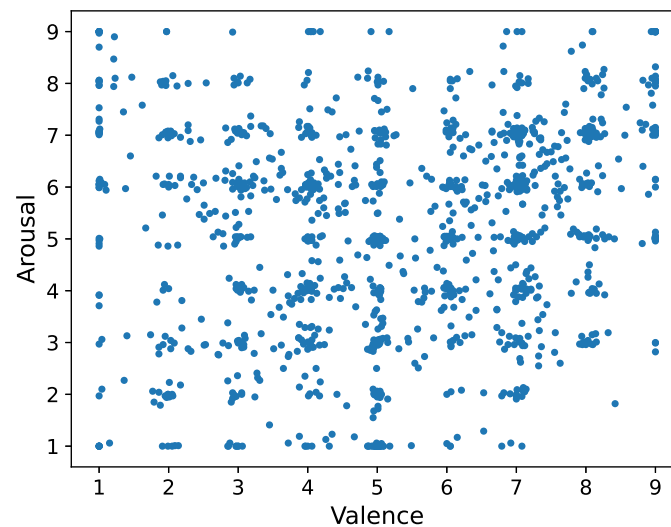
Figure 5. Scatter plot of EEG label.

The parameters Eps and MinPts have a significant impact on the accuracy of emotion mapping. Typically, the values for Eps and MinPts are determined based on the sorted k-dist map [31]. Setting k equal to MinPts, for two-dimensional data, the clustering results do not change significantly when k > 4. We set k to 2, 3, 4, and 5. Figure 6 shows the sorted k-dist map.
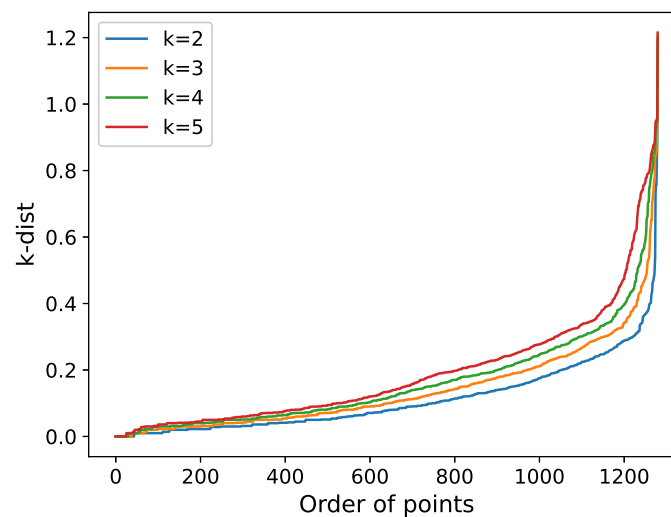

Figure 6. Sorted k-dist graph of EEG label.

The determination of the Eps value is based on the location of the "valley" in the sorted k-dist map. The "valley" refers to a specific threshold at which the curve of the plot changes drastically, and the value of the vertical coordinate of this threshold is the Eps. However, it is a challenging task to determine the appropriate "valley" location precisely by visual inspection. Therefore, we must rely on reliable Eps selection methods, and for this purpose, we adopted the algorithm described by Guang Yu et al. in [32] to determine the "valley point". The identified "valley points" are shown in Table 2.

Table 2. The Eps values corresponding to different k values.

k	Eps
2	0.32
3	0.33
4	0.34
5	0.4

The results of the DBSCAN clustering visualization are shown in Figure 7, In the visualization, core samples, which are denoted by large dots, and non-core samples, indicated by small dots, are color-assigned to their respective clusters. The filtered-out noise points are represented in black. The number of clusters generated through clustering diminishes with the growth of the MinPts parameter.
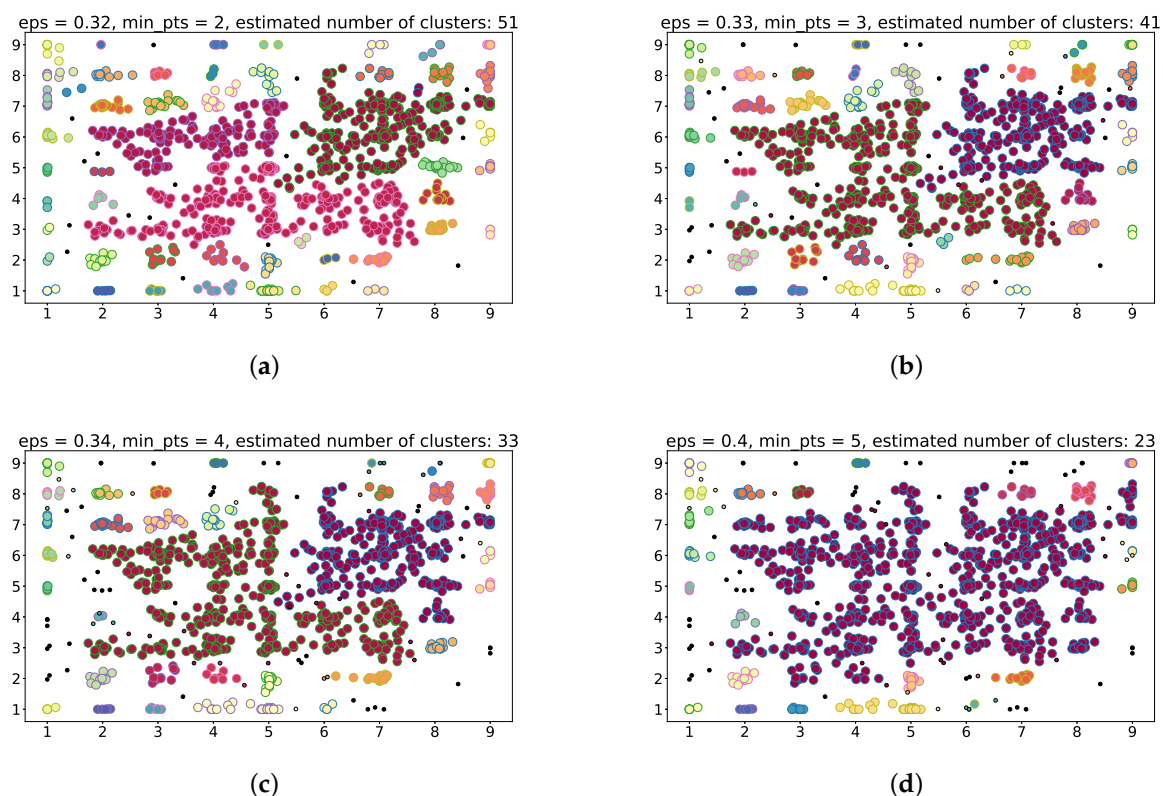

Figure 7. The EEG label clustering results under determined parameters, the use of different colors is intended to highlight the distinction among various clusters. (**a**) Eps = 0.32, MinPts = 2. (**b**) Eps = 0.33, MinPts = 3. (**c**) Eps = 0.34, MinPts = 4. (**d**) Eps = 0.4, MinPts = 5.

4.3. Emotion Classification Performance

We conduct 5-fold cross-validation using the data of each subject, representing their mean classification accuracy and standard deviation as subject results, and then presenting these results combined as the average for the entire study population (32 subjects). The outcomes of the emotion classification model applied to the DEAP dataset are presented

in Figure 8. Overall, the average accuracy for the 32 subjects is 92.98% with a standard deviation of 1.77%, and 25 subjects demonstrate classification accuracies higher than 90% in the experiment, with the exception of 2, 5, 11, 17, 19, 22, and 24. Among these, 14 subjects achieve accuracy levels above 95%. However, it should be noted that the accuracy of subject 22 is 70.75%, which is significantly lower than the others. discrepancy may be attributed to that subjects' failure to provide accurate subjective feedback after the experiment.
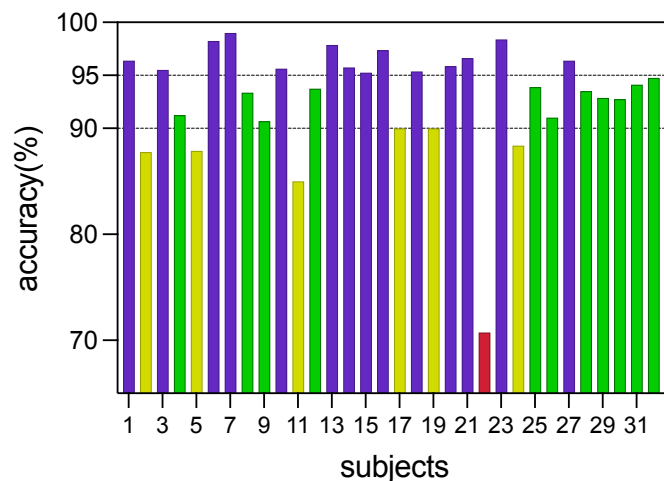
Figure 8. Overall performance of the emotion classifier.

We also compare our model with different approaches that have applied the DEAP dataset [17,19–21], which are described in Section 2. These related methods also involve extracting features from time and space, combined with CNN or LSTM. As shown in Table 3, the comparison demonstrates the efficiency of our model for a finer-grained emotion classification task.

Table 3. Comparison with different approaches on DEAP dataset.

Authors	Classification Methods	Emotion Classification Basis	Acc (%)
Pane et al. [17]	RF	happy, sad, angry, relaxed	75.6
Sharma et al. [21]	LSTM	HAHV, HALV, LAHV, LALV 2class/Arousal, 2class/Valence	4 classes: 82.01 Arousal: 85.21 Valence: 84.16
Russo [19]	CNN+SVM	2class/Arousal, 2class/Valence	Arousal: 77.7 Valence: 76.6
Chen et al. [20]	CVCNN	2class/Arousal, 2class/Valence	Arousal: 88.51 Valence: 84.02
Our method	CNN+LSTM	joy, liking, fear, sadness, dislike, despair	92.98

About 15% and 4% higher than the two methods that also use CNN models, respectively. In emotion classification basis, in addition to the common binary arousal classification and binary valence classification, Sharma et al. [21] also conduct the classification of four composite emotional states. Although Pane et al. [17] classify four specific categories of emotions from a discrete perspective, they essentially divide four composite emotions based on the valence-arousal plane by mapping each composite emotion to happy, sad, angry, and relaxed. Our method performs more fine-grained emotion division, the performance of our model surpassed the methods employed by Pane and Sharma.

5. Conclusions

In this research, we use a clustering approach in combination with the linguistic resource WordNet-Affect in order to realize the mapping of the emotion dimension model to a discrete model, thereby constructing emotion classification labels. We extract EEG

frequency-domain features from different frequency bands and organize them into a three-dimensional structure, combining them with electrode location information. These features are then fed into a hybrid deep learning model of CNN and LSTM, enabling the recognition of six distinct types of emotions.

We conduct an empirical study on the DEAP dataset using this method and obtain satisfactory results. Specifically, our method achieves an average accuracy of 92.98% with a standard deviation of 1.77% for each subject. Additionally, we review current EEG analysis techniques for recognizing human emotions. Compared with similar studies, our method provides an improved emotion classification basis, broader practical applications, and a significant improvement in categorization accuracy. We propose that our research findings hold significant implications for improving mental health diagnostics and understanding human emotions. In order to translate our research into practical policy and application, we recommend that policymakers explore ways to integrate our emotion classification model into mental health assessments, enabling effective diagnosis and personalized treatment.

Due to the cumbersome collection of EEG data, the current scale of EEG emotion databases is relatively small. Although we can perform segment-level emotion recognition and use sliding windows to segment samples, these scales are insufficient for deep learning methods that require a large number of samples for training. Problems that need to be solved in the future include the impact of insufficient training data on the model, and a more natural library of emotion-inducing materials needs to be established to support high-quality data collection, thereby building a larger-scale EEG emotion database. In addition, there is a high degree of correlation among some emotion categories represented by the discrete emotion model, posing a certain challenge for discrete emotion recognition.

Author Contributions: Conceptualization, Y.W. and L.Z.; methodology, B.X. and Y.W.; software, B.X.; validation, B.X. and Y.W.; formal analysis, B.X. and W.Z.; investigation, B.X.; resources, L.Z.; data curation, W.Z. and Y.H.; writing—original draft preparation, B.X.; writing—review and editing, L.Z., B.X. and Y.H.; visualization, L.Z.; supervision, L.Z. and Y.W.; project administration, L.Z.; funding acquisition, L.Z. and Y.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goshvarpour, A.; Goshvarpour, A. EEG spectral powers and source localization in depressing, sad, and fun music videos focusing on gender differences. *Cogn. Neurodyn.* **2019**, *13*, 161–173. [CrossRef] [PubMed]
2. Fiorini, L.; Mancioppi, G.; Semeraro, F.; Fujita, H.; Cavallo, F. Unsupervised emotional state classification through physiological parameters for social robotics applications. *Knowl. Based Syst.* **2020**, *190*, 105217. [CrossRef]
3. Zeng, H.; Yang, C.; Dai, G.; Qin, F.; Zhang, J.; Kong, W. EEG classification of driver mental states by deep learning. *Cogn. Neurodyn.* **2018**, *12*, 597–606. [CrossRef] [PubMed]
4. Dabas, H.; Sethi, C.; Dua, C.; Dalawat, M.; Sethia, D. Emotion classification using EEG signals. In Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, Shenzhen, China, 8–10 December 2018; pp. 380–384.
5. Russell, J.A. Affective space is bipolar. *J. Personal. Soc. Psychol.* **1979**, *37*, 345. [CrossRef]
6. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]
7. Zheng, W.L.; Lu, B.L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]
8. Strapparava, C.; Valitutti, A. WordNet Affect: An Affective Extension of WordNet. In Proceedings of the International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26–28 May 2004.

9. Warriner, A.B.; Kuperman, V.; Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **2013**, *45*, 1191–1207. [CrossRef]

10. Sammler, D.; Grigutsch, M.; Fritz, T.; Koelsch, S. Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* **2007**, *44*, 293–304. [CrossRef]

11. Aftanas, L.; Reva, N.; Savotina, L.; Makhnev, V. Neurophysiological correlates of induced discrete emotions in humans: An individually oriented analysis. *Neurosci. Behav. Physiol.* **2006**, *36*, 119–130. [CrossRef]

12. Dadebayev, D.; Goh, W.W.; Tan, E.X. EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques. *J. King Saud-Univ. Inf. Sci.* **2022**, *34*, 4385–4401. [CrossRef]

13. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [CrossRef]

14. Luo, Y.; Fu, Q.; Xie, J.; Qin, Y.; Wu, G.; Liu, J.; Jiang, F.; Cao, Y.; Ding, X. EEG-based emotion classification using spiking neural networks. *IEEE Access* **2020**, *8*, 46007–46016. [CrossRef]

15. Yoon, H.J.; Chung, S.Y. EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Comput. Biol. Med.* **2013**, *43*, 2230–2237. [CrossRef]

16. Atkinson, J.; Campos, D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl.* **2016**, *47*, 35–41. [CrossRef]

17. Pane, E.S.; Wibawa, A.D.; Purnomo, M.H. Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters. *Cogn. Process.* **2019**, *20*, 405–417. [CrossRef]

18. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [CrossRef] [PubMed]

19. Topic, A.; Russo, M. Emotion recognition based on EEG feature maps through deep learning network. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 1442–1454. [CrossRef]

20. Chen, J.; Zhang, P.; Mao, Z.; Huang, Y.; Jiang, D.; Zhang, Y. Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access* **2019**, *7*, 44317–44328. [CrossRef]

21. Sharma, R.; Pachori, R.B.; Sircar, P. Automated emotion recognition based on higher order statistics and deep learning algorithm. *Biomed. Signal Process. Control.* **2020**, *58*, 101867. [CrossRef]

22. Li, C.; Bao, Z.; Li, L.; Zhao, Z. Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. *Inf. Process. Manag.* **2020**, *57*, 102185. [CrossRef]

23. Li, R.; Ren, C.; Zhang, X.; Hu, B. A novel ensemble learning method using multiple objective particle swarm optimization for subject-independent EEG-based emotion recognition. *Comput. Biol. Med.* **2022**, *140*, 105080. [CrossRef] [PubMed]

24. Hasan, M.; Rokhshana-Nishat-Anzum; Yasmin, S.; Pias, T.S. Fine-grained emotion recognition from eeg signal using fast fourier transformation and cnn. In Proceedings of the 2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 16–20 August 2021; pp. 1–9.

25. Bălan, O.; Moise, G.; Petrescu, L.; Moldoveanu, A.; Leordeanu, M.; Moldoveanu, F. Emotion classification based on biophysical signals and machine learning techniques. *Symmetry* **2019**, *12*, 21. [CrossRef]

26. Oostenveld, R.; Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* **2001**, *112*, 713–719. [CrossRef] [PubMed]

27. Ekman, P.; Sorenson, E.R.; Friesen, W.V. Pan-cultural elements in facial displays of emotion. *Science* **1969**, *164*, 86–88. [CrossRef] [PubMed]

28. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.

29. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

30. Morris, J.D. Observations: SAM: The Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. *J. Advert. Res.* **1995**, *35*, 63–68.

31. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, Portland, OR, USA, 2–4 August 1996; AAAI Press: Portland, OR, USA, 1996; pp. 226–231.

32. Yu, G.; Cai, Z.; Wang, S.; Chen, H.; Liu, F.; Liu, A. Unsupervised online anomaly detection with parameter adaptation for KPI abrupt changes. *IEEE Trans. Netw. Serv. Manag.* **2019**, *17*, 1294–1308. [CrossRef]