



Singular Learning of Deep Multilayer Perceptrons for EEG-Based Emotion Recognition

Weili Guo^{1,2}, Guangyu Li^{1,2*}, Jianfeng Lu² and Jian Yang^{1,2*}

¹PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Jiangsu Key Lab of Image and Video Understanding for Social Security, Nanjing University of Science and Technology, Nanjing, China, ²School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

OPEN ACCESS

Edited by:

Xiaopeng Hong,
Xi'an Jiaotong University, China

Reviewed by:

Lianyong Qi,
Qufu Normal University, China
Liangfei Zhang,
University of St Andrews,
United Kingdom

*Correspondence:

Guangyu Li
guangyu.li2017@njust.edu.cn
Jian Yang
csjyang@njust.edu.cn

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Computer Science

Received: 30 September 2021

Accepted: 05 November 2021

Published: 21 December 2021

Citation:

Guo W, Li G, Lu J and Yang J (2021)
Singular Learning of Deep Multilayer
Perceptrons for EEG-Based
Emotion Recognition.
Front. Comput. Sci. 3:786964.
doi: 10.3389/fcomp.2021.786964

Human emotion recognition is an important issue in human-computer interactions, and electroencephalograph (EEG) has been widely applied to emotion recognition due to its high reliability. In recent years, methods based on deep learning technology have reached the state-of-the-art performance in EEG-based emotion recognition. However, there exist singularities in the parameter space of deep neural networks, which may dramatically slow down the training process. It is very worthy to investigate the specific influence of singularities when applying deep neural networks to EEG-based emotion recognition. In this paper, we mainly focus on this problem, and analyze the singular learning dynamics of deep multilayer perceptrons theoretically and numerically. The results can help us to design better algorithms to overcome the serious influence of singularities in deep neural networks for EEG-based emotion recognition.

Keywords: emotion recognition, EEG, deep multilayer perceptrons, singular learning, theoretical and numerical analysis

1 INTRODUCTION

Emotion recognition is a fundamental task in affective computing and has attracted many researchers' attention in recent years (Mauss and Robinson, 2009). Human emotion can be expressed through external signals and internal signals, where external signals usually include facial expressions, body actions, and speeches, and electroencephalograph (EEG) and galvanic skin response (GSR) are typical internal signals. EEG is the method to measure electrical activities of the brain by using electrodes along the scalp skin and it is rather reliable; therefore, EEG has played a more significant role in investigating human emotion recognition problem in recent years (Yin et al., 2021).

For the emotion recognition problem based on EEG signals, researchers mainly investigate this issue from two aspects: how to extract better features from EEG signals and how to construct a model with better performance. For aspect 1, researchers have investigated the feature extraction methods of EEG signals from a time domain, frequency domain, and time-frequency domain, respectively, and a series of results have been given previously (Fang et al., 2020; Nawa et al., 2020). In this paper, we mainly focus on aspect 2, i.e., the computational model problem, and researchers have proposed many models to recognize emotions through EEG signals (Zong et al., 2016; Yang et al., 2018a; Zhang et al., 2019; Cui et al., 2020). In recent years, deep learning technology has achieved great success in many fields (Yang et al., 2018b; Yang et al., 2019; Basodi et al., 2020; Zhu and Zhang, 2021), and many works are devoted to addressing the EEG emotion recognition issue by applying deep neural networks (DNNs) (Cao et al., 2020; Natarajan et al., 2021), where the performances based on deep

learning also show significant superiority of conventional methods (Ng et al., 2015; Tzirakis et al., 2017; Hassan et al., 2019). However, the learning dynamics of DNNs, including deep multilayer perceptrons (MLPs), deep belief networks and deep convolution neural networks, are often affected by singularities, which exist in the parameter space of DNNs (Nitta, 2016).

Due to the influence of singularities, the training of DNNs often becomes very slow and the plateau phenomenon can often be observed. When the DNNs are applied to EEG-based emotion recognition, the severe negative effect of singularities on the learning process of DNNs is also inevitable, where the efficiency and performance of networks can also not be guaranteed. However, up to now, there are rarely literatures investigating this problem. In this paper, we mainly concern this problem. The main contribution of this paper is to take the theoretical and numerical analysis of singular learning in DNNs for EEG-based emotion recognition. We choose deep MLPs as the learning machine, where deep MLPs are of typical DNNs and the results are also representative for other DNNs. The types of singularities in parameter space are analyzed and the specific influence of the singularities is clearly shown. Based on the obtained results in this paper, we can further design the related algorithms to overcome this issue.

The rest of this paper is organized as follows. A brief review of related work is presented in **Section 2**. In **Section 3**, theoretical analysis of singularities in deep MLPs for EEG-based emotion recognition is taken and then the learning dynamics near singularities are numerically analyzed in **Section 4**. **Section 5** states conclusion and discussion.

2 RELATED WORK

In this section, we provide a brief overview of previous work on EEG-based emotion recognition and singular learning of DNNs.

In recent years, due to the high accuracy and stabilization of EEG signals, EEG-based algorithms have attracted ever-increasing attention in emotion recognition field. To extract better features of EEG signals, researchers have proposed various feature extraction models (Zheng et al., 2014; Zheng, 2017; Tao et al., 2020; Zhao et al., 2021), such as power spectral density (PSD), differential entropy (DE), and differential asymmetry (DASM). By using PSD and DE to extract dimension reduced features of EEG signals, Fang et al. (2020) chose the original features and dimension reduced features as the multi-feature input and verified the validity of the proposed method in the experiment part. Li et al. (2020) integrated psychoacoustic knowledge and raw waveform embedding within an augmented feature space. Song et al. (2020) employed an additional branch to characterize the intrinsic dynamic relationships between different EEG channels and a type of sparse graphic representation was presented to extract more discriminative features. Besides the feature extraction methods, more attention is paid to study the emotion classification. Given that the deep learning technology has excellent capabilities, various types of DNNs have been widely used in emotion classification (Li et al., 2018; Li et al., 2019; Ma

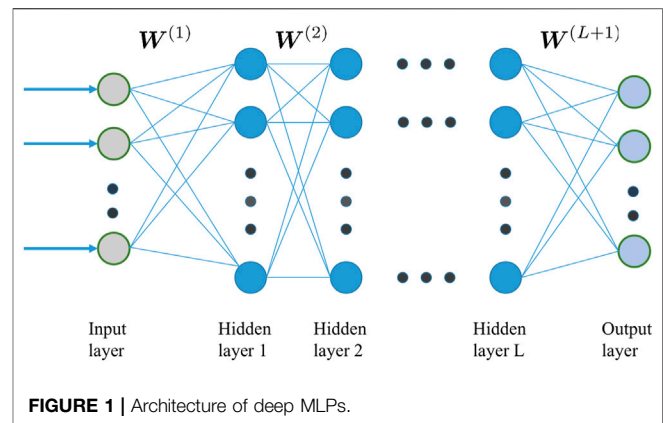


FIGURE 1 | Architecture of deep MLPs.

et al., 2019; Atmaja and Akagi, 2020; Cui et al., 2020; Zhong et al., 2020), including deep convolution neural networks, deep MLPs, long short term memory (LSTM)-based recurrent neural networks, and graph neural networks. The obtained results show that these DNN models can provide superior performance compared to previous models (Yang et al., 2021a; Yang et al., 2021b).

As mentioned above, various DNNs have been widely used in EEG-based emotion recognition; however, the training processes of DNNs often encounter many difficulties. Even if numerous research studies have been developed to conduct explanatory research, it is still very far to revealing the mechanism. As there are singularities in the parameter space of DNNs where the Fisher information matrix is singular, the singular learning dynamics of DNNs have been studied and have attracted more and more attention. As the basis of DNNs, traditional neural networks often suffer from the serious influence of various singularities (Amari et al., 2006; Guo et al., 2018; Guo et al., 2019), and the learning dynamics of DNNs are also easy to be influenced by the singularities. Nitta (2016, 2018) analyzed the types of singularity in DNNs and deep complex-value neural networks. Ainsworth and Shin (2020) investigated the plateau phenomenon in ReLU-based neural networks. By using the spectral information of Fisher information matrix, Liao et al. (2020) proposed an algorithm to accelerate the training process of DNNs.

In view of the serious influence of singularities to DNNs, the training processes of DNNs will also encounter difficulties when applying DNNs to EEG-based emotion recognition. Thus, it is necessary to take the theoretical and numerical analysis to reveal the mechanism and propose related algorithms to overcome the influence of singularities.

3 THEORETICAL ANALYSIS OF SINGULAR LEARNING DYNAMICS OF DEEP MULTILAYER PERCEPTRONS

In this section, we theoretically analyze the learning dynamics near singularities of deep MLPs for the EEG-based emotion recognition.

3.1 Learning Paradigm of Deep Multilayer Perceptrons

Firstly, we introduce a typical learning paradigm of deep MLPs. For a typical deep multilayer perceptrons with L hidden layers (the architecture of the networks is shown in **Figure 1**), assuming M_i is the neuron number of hidden layer i , M_0 is the dimension of the input layer and M_{L+1} is the dimension of the output layer, we denote that: $W_{jk}^{(i)}$ represents the weight connecting from the j th node of the previous layer to the k th node of hidden layer i , and $W_{pq}^{(L+1)}$ represents the weight connecting from the p th node of hidden layer L to the q th node of output layer for $1 \leq i \leq L$, $1 \leq j \leq M_{i-1}$, $1 \leq k \leq M_i$, $1 \leq p \leq M_L$, and $1 \leq q \leq M_{L+1}$. Then $\theta = \{W^{(1)}, W^{(2)}, \dots, W^{(L+1)}\}$ represents all the parameters of the networks, where $W^{(i)} = [W_1^{(i)}, W_2^{(i)}, \dots, W_{M_i}^{(i)}]$ and $W_j^{(i)} = [W_{1j}^{(i)}, W_{2j}^{(i)}, \dots, W_{M_{(i-1)}j}^{(i)}]^T$ for $1 \leq i \leq L + 1$ and $1 \leq j \leq M_i$. In this paper, the widely used log-sigmoid function $\phi(x) = \frac{1}{1+e^{-x}}$ is adopted as the activation of hidden layers and the purelin function $\psi(x) = x$ is adopted as the activation function of output layer, then for the input $x \in \mathbb{R}^{M_0}$, by denoting the input to hidden layer k as $X^{(k-1)}$ for $1 \leq k \leq L$ and the input to output layer as $X^{(L)}$, the mathematical model of the networks can be described as follows:

$$f(x, \theta) = \psi((W^{(L+1)})^T X^{(L)}) = (W^{(L+1)})^T X^{(L)}. \quad (1)$$

For $1 \leq k \leq L$, $X^{(k)}$ can be computed as $X^{(k)} = \phi(X^{(k-1)}, W^{(k)}) = \phi((W^{(k)})^T X^{(k-1)})$ and $X^{(0)}$ is the input x .

We choose the square loss function to measure the error:

$$l(y, x, \theta) = \frac{1}{2}(y - f(x, \theta))^2, \quad (2)$$

and use the gradient descent method to minimize the loss:

$$\theta_{t+1} = \theta_t - \eta \frac{\partial l(y, x, \theta_t)}{\partial \theta_t}, \quad (3)$$

where η is the learning rate.

3.2 Singularities of Deep Multilayer Perceptrons in Electroencephalograph-Based Emotion Recognition

In this paper, we mainly focus on the mechanism of singular learning dynamics of deep MLPs applied to EEG-based emotion recognition domain, not seeking the best performance; therefore, the size of the networks need not to be very large, and an appropriate size that can capture the essence of singular learning dynamics can satisfy the requirement. Without loss of generality, we choose the deep MLPs with two hidden layers and a single output neuron, i.e., $L = 2$ and $M_3 = 1$, i.e., $W^{(3)} = W_1^{(3)} = [W_{11}^{(3)}, W_{21}^{(3)}, \dots, W_{M_2}^{(3)}]^T$, we simply denoted as $W^{(3)} = [W_1^{(3)}, W_2^{(3)}, \dots, W_{M_2}^{(3)}]$. Then, the deep MLPs can be rewritten as:

$$\begin{aligned} f(x, \theta) &= (W^{(3)})^T \phi(\phi(x, W^{(1)}), W^{(2)}) \\ &= \sum_{j=1}^{M_2} W_j^{(3)} \phi(\phi(x, W^{(1)}), W_j^{(2)}). \end{aligned} \quad (4)$$

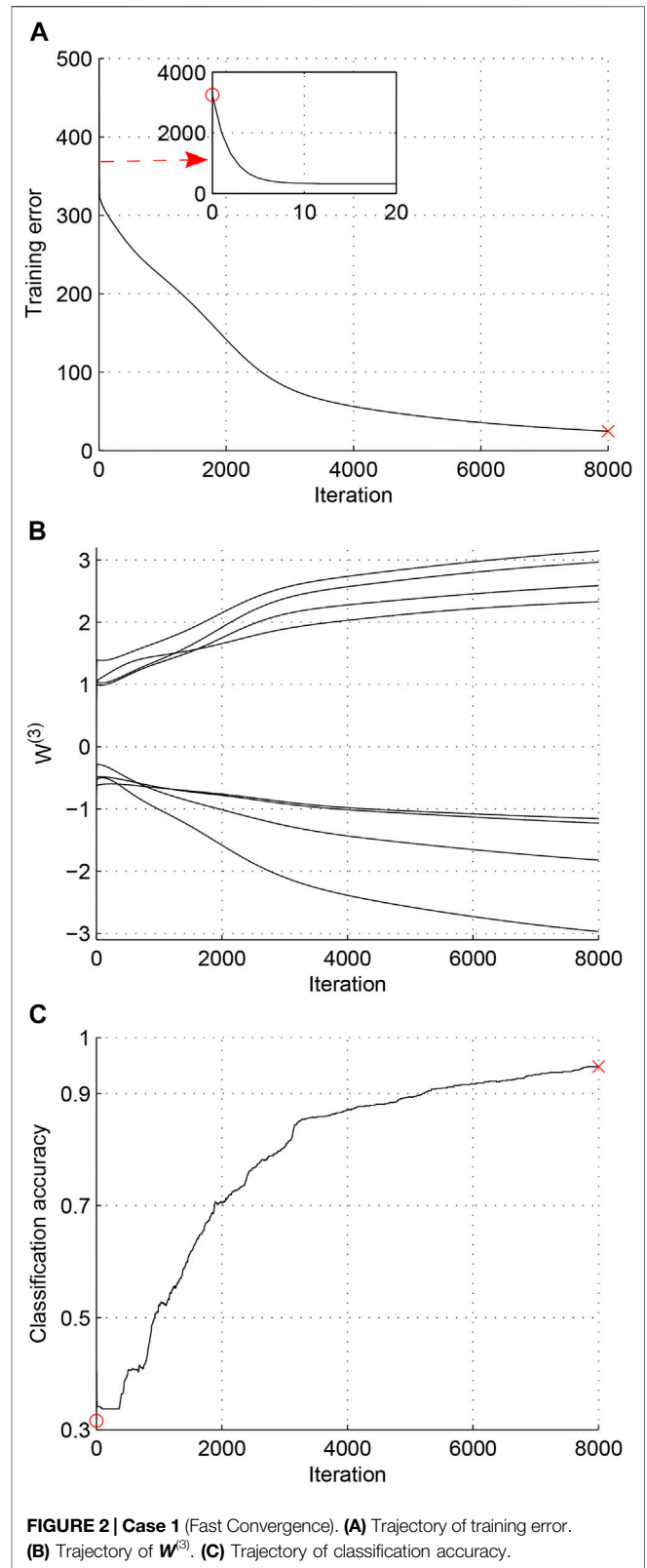


FIGURE 2 | Case 1 (Fast Convergence). (A) Trajectory of training error. (B) Trajectory of $W^{(3)}$. (C) Trajectory of classification accuracy.

Next, we analyze the types of singularities. From **Eq. 4**, we can see that if one output weight equals zero, e.g., $W_j^{(3)} = 0$, whatever the values of $W^{(1)}$ and $W_j^{(2)}$ be, the output of unit j will be always 0

and the unit seems to be vanished. As the values of $W^{(1)}$ and $W_j^{(2)}$ have no effect on the output of the deep MLPs, the training process will encounter difficulties on the subspace $R_1 = \{\theta | W_j^{(3)} = 0\}$. Besides the above singularity, if there are two elements of weight $W^{(2)}$ overlap, e.g., $W_i^{(2)} = W_j^{(2)}$, then

$$\begin{aligned} &W_i^{(3)}\phi(\phi(\mathbf{x}, W^{(1)}), W_i^{(2)}) \\ &+ W_j^{(3)}\phi(\phi(\mathbf{x}, W^{(1)}), W_j^{(2)}) \\ &= (W_i^{(3)} + W_j^{(3)})\phi(\phi(\mathbf{x}, W^{(1)}), W_i^{(2)}) \end{aligned}$$

remains the same value when $W_i^{(3)} + W_j^{(3)}$ takes a fixed value, regardless of particular values of $W_i^{(3)}$ and $W_j^{(3)}$. Therefore, we can identify their sum $W = W_i^{(3)} + W_j^{(3)}$; nevertheless, each of $W_i^{(3)}$ and $W_j^{(3)}$ remains unidentifiable. Thus, the training will also suffer difficulties on the subspace $R_2 = \{\theta | W_i^{(2)} = W_j^{(2)}\}$.

To sum up the above analysis, it can be seen that there are at least two types of singularities:

- (1) Zero weight singularity: $R_1 = \{\theta | W_j^{(3)} = 0\}$,
- (2) Overlap singularity: $R_2 = \{\theta | W_i^{(2)} = W_j^{(2)}\}$.

Till now, we have theoretically analyzed the types of singularity that existed in the parameter space of deep MLPs; in the next section, we will numerically analyze the influence of singularities to solve EEG-based emotion recognition problem.

4 NUMERICAL ANALYSIS OF LEARNING DYNAMICS NEAR SINGULARITIES

In this section, we take the numerical analysis of singularities by taking experiments on the dataset of EEG signals. For the EEG datasets, the SEED dataset is a typical benchmark dataset that is developed by SJTU and has been widely used to evaluate the proposed methods on EEG-based emotion recognition. In this paper, the training process will be carried out using the SEED dataset.

4.1 Data Preprocessing

The SEED dataset (Zheng and Lu, 2015) is collected from 62-channel EEG device and contains EEG signals of three emotions (positive, neutral, and negative) from 15 subjects. Due to the low signal-to-noise ratio of raw EEG signals, it is rather necessary to take the preprocessing step to extract meaningful features. As is known, there are five frequency bands for each EEG channel: delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz). That means, for one subject, the data are the form 5×62 , the dimension of raw EEG signal is very large, and then we use principal component analysis (PCA) (Abdi and Williams, 2010) to extract the features of the EEG signal. After the PCA step, the form of EEG signals becomes 5×5 , and then by putting every element of the data to a vector, the dimension of the input can be finally reduced to be 25.

4.2 Learning Trajectories Near Singularities

Now, we take experiments on the SEED dataset, and the learning dynamics near singularities will be numerically analyzed. We

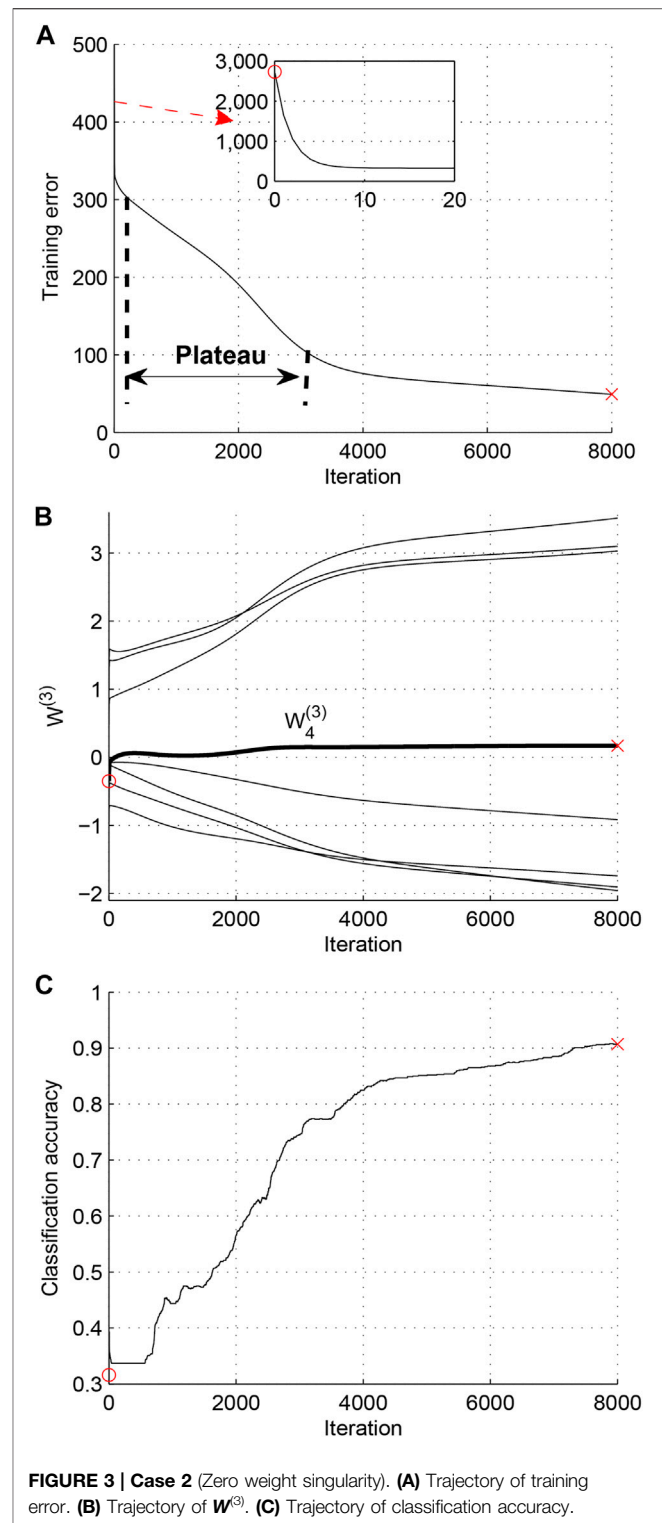
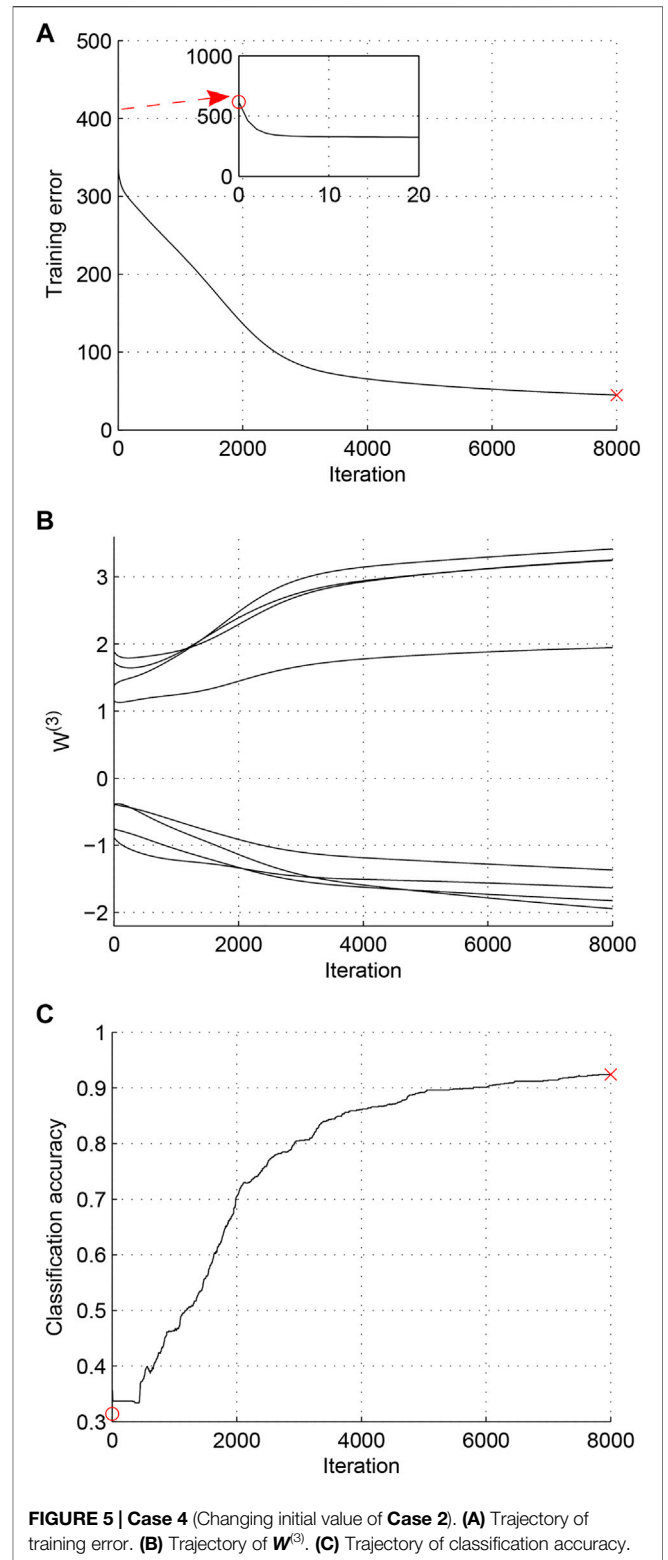
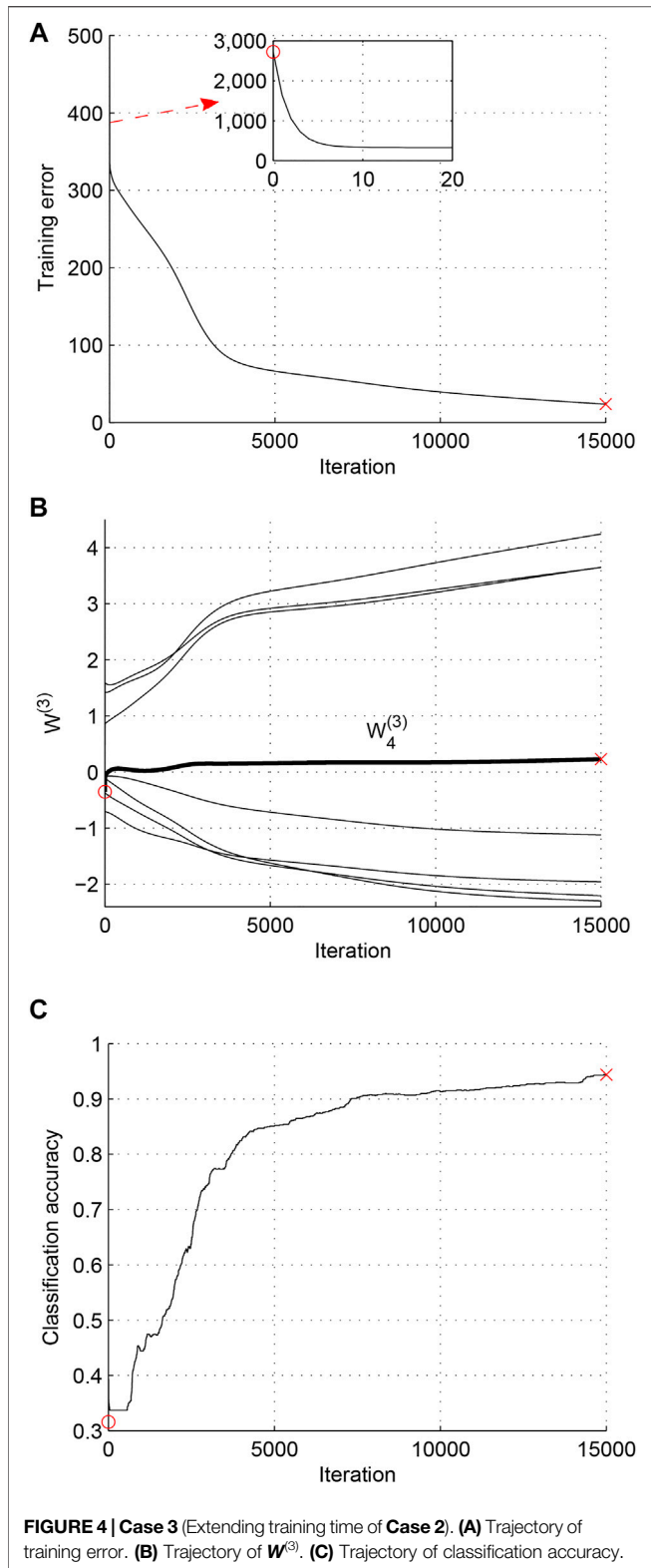


FIGURE 3 | Case 2 (Zero weight singularity). (A) Trajectory of training error. **(B)** Trajectory of $W^{(3)}$. **(C)** Trajectory of classification accuracy.

choose the neuron numbers of two hidden layers as $L_1 = 8$ and $L_2 = 8$; thus, the architecture of the deep MLPs is 25–8–8–1. As there are three emotions in the SEED dataset, we set values 1, 2, and 3 corresponding to labels positive, neutral, and negative, respectively. We choose the training sample number and



testing sample number to be 1,000 and 500, respectively. Then, by setting the learning rate to $\eta = 0.002$, the target error to 0.05, and the maximum epochs to 8,000, we use Eq. 3 to accomplish the experiment. By analyzing the experiment results, two cases of

learning dynamics will be shown. Besides training error, classification accuracy is also used to measure the performance. In the following figures of experiment results, “o” and “x” represent the initial state and final state,

TABLE 1 | Training and testing classification accuracy.

	Iteration number	Training classification accuracy	Testing classification accuracy
Case 1	8,000	0.948	0.941
Case 2	8,000	0.901	0.894
Case 3	15,000	0.944	0.938
Case 4	8,000	0.924	0.920

respectively. The experiments were run by using Matlab 2013a on a PC with an Intel Core i7-9700K CPU @3.60 GHz, 32 GB RAM and NVIDIA GeForce RTX 2070 GPU.

Case 1. Fast convergence: The learning process fast converges to the global minimum.

For this case, the learning dynamics does not suffer from any influence of singularity and the parameters fast converge to the optimal value. The initial value of $\mathbf{W}^{(3)}$ is $\mathbf{W}^{(3)(0)} = [0.8874, 0.6993, 0.5367, -0.9415, -0.8464, -0.9280, 0.3335, -0.7339]^T$ and the final value of $\mathbf{W}^{(3)}$ is $\mathbf{W}^{(3)} = [3.1443, 2.5868, 2.3291, -1.1544, -1.2281, -2.9704, 2.9650, -1.8221]^T$. The experiment results are shown in **Figure 2**, which represent the trajectories of training error, output weights $\mathbf{W}^{(3)}$, and classification accuracy, respectively.

As can be seen from **Figure 2A**, the learning dynamics quickly converge to the global minimum and have not been affected by any singularity.

Case 2. Zero weight singularity: the learning process is affected by the elimination singularity.

For this case, one output weight crosses 0 during the learning process and a plateau phenomenon can be obviously observed. The initial value of $\mathbf{W}^{(3)}$ is $\mathbf{W}^{(3)(0)} = [0.4825, 0.9885, -0.9522, -0.3505, -0.5004, 0.9749, -0.9111, -0.5056]^T$, and the final student parameters are $\mathbf{W}^{(3)} = [3.0297, 3.1006, -1.7413, 0.1717, -1.9567, 3.5131, -1.9037, -0.9143]^T$. The experiment results are shown in **Figure 3**, which represent the trajectories of training error, output weights $\mathbf{W}^{(3)}$ and classification accuracy, respectively.

From **Figure 3B**, we can see that $W_4^{(3)}$ crosses 0 in the learning process and the learning process is affected by elimination singularity. During the stage $W_4^{(3)}$ crosses 0, the plateau phenomenon can be obviously observed (**Figure 3A**). Then, the student parameters escape the influence of elimination singularity. After the training process, we can see that the training error is bigger than that in Case 1 and the classification accuracy is also lower than that in Case 1, which means that the parameters do not reach the optimum.

Case 3. Extending training time of Case 2.

In this experiment, we only increase the training epochs to 15,000, and the rest of the experiment setup remains the same with that in Case 2. The experiment results are shown in **Figure 4**. Compared to **Figure 3**, it can be seen that the learning process that is affected by the zero weight singularity can arrive at the optimum, but it costs much more time. This means that the zero

weight singularities will greatly reduce the efficiency of deep MLPs.

Case 4. Changing initial value of Case 2.

In order to confirm that the plateau phenomenon corresponds to the zero weight singularity, a supplementary experiment is carried out here where only the initial value of $\mathbf{W}^{(3)(0)}$ has been changed and the rest of the experiment setup remains the same. The initial value of $\mathbf{W}^{(3)}$ is $\mathbf{W}^{(3)(0)} = [-0.5056, -0.9111, 1.7749, -0.5004, 1.6495, -0.9522, 0.9885, 1.2825]^T$, and the final student parameters are $\mathbf{W}^{(3)} = [-1.3660, -1.8232, 3.2529, -1.9425, 3.2450, -1.6325, 1.9452, 3.4158]^T$. The experiment results are shown in **Figure 5**, which represent the trajectories of training error, output weights $\mathbf{W}^{(3)}$, and classification accuracy, respectively. As can be seen in **Figure 5**, there is not any weight of $\mathbf{W}^{(3)}$ that becomes zero. Also, no plateau phenomenon can be observed, and the classification accuracy has reached a comparatively high value. By comparing the experiment results shown in **Figures 3, 5**, we can conclude that the plateau phenomenon is indeed caused by zero weight singularity.

Remark 1. From the results shown in **Figures 2–5** and **Table 1**, we can see that the training and testing accuracy in Case 2 is the lowest. This means that when the training process is affected by the zero weight singularity, the parameters cannot achieve the optimum after the same training time with that in fast convergence case. When we extend the training time in Case 2, the parameters can escape the influence of zero weight singularity and finally arrive at the optimum, which is shown in Case 3. Thus, the points in zero weight singularity are saddle points, not local minimum. To sum up, the zero weight singularity will seriously delay the training process, and it is worthy to investigate algorithms to overcome the influence of zero weight singularities.

Remark 2. When taking the experiments, we do not observe the learning dynamics of deep MLPs that are affected by overlap singularities. The results are in accordance with the conclusion where we analyze the learning dynamics of shallow neural networks (Guo et al., 2018); i.e., the overlap singularities mainly influence the neural networks with low dimension and the large-scale networks predominantly suffer from zero weight singularities. Thus, we should pay more attention to how to overcome the influence of zero weight singularities.

In this section, we have numerically analyzed the learning dynamics near singularities of deep MLPs for EEG-based emotion recognition and showed the singular case. We can obtain that the

learning dynamics of deep MLPs are mainly influenced by zero weight singularities and rarely affected by overlap singularities.

5 CONCLUSION AND DISCUSSION

Deep learning technology has been widely used in EEG-based emotion recognition and has shown superior performance compared to traditional methods. However, for various DNNs, there exist singularities in the parameter space, which cause singular behaviors in the training process. In this paper, we investigate the singular learning dynamics of DNNs when applied to EEG-based emotion recognition. By choosing deep MLPs as the learning machine, we firstly take the theoretical analysis of singularities of deep MLPs, and obtained that there are at least two types of singularities: overlap singularity and zero weight singularity. Then, by doing several experiments, the numerical analysis is taken. The experiment results show that the learning dynamics of deep MLPs are seriously influenced by zero weight singularities and rarely affected by overlap singularities. Furthermore, the plateau phenomenon is caused by zero weight singularity. Thus, we should pay more attention to how to overcome the serious influence of zero weight singularity to improve the efficiency of DNNs in EEG-based emotion recognition in the future.

REFERENCES

- Abdi, H., and Williams, L. J. (2010). Principal Component Analysis. *Wires Comp. Stat.* 2, 433–459. doi:10.1002/wics.101
- Ainsworth, M., and Shin, Y. (2020). Plateau Phenomenon in Gradient Descent Training of Relu Networks: Explanation, Quantification and Avoidance.
- Amari, S.-i., Park, H., and Ozeki, T. (2006). Singularities Affect Dynamics of Learning in Neuromanifolds. *Neural Comput.* 18, 1007–1065. doi:10.1162/neco.2006.18.5.1007
- Atmaja, B. T., and Akagi, M. (2020). “Deep Multilayer Perceptrons for Dimensional Speech Emotion Recognition,” in Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Auckland, New Zealand, December 7–10, 2020 (Auckland: IEEE), 325–331.
- Basodi, S., Ji, C., Zhang, H., and Pan, Y. (2020). Gradient Amplification: An Efficient Way to Train Deep Neural Networks. *Big Data Min. Anal.* 3, 196–207. doi:10.26599/bdma.2020.9020004
- Cao, Q., Zhang, W., and Zhu, Y. (2020). Deep Learning-Based Classification of the Polar Emotions of “moe”-Style Cartoon Pictures. *Tsinghua Sci. Technology* 26, 275–286.
- Cui, H., Liu, A., Zhang, X., Chen, X., Wang, K., and Chen, X. (2020). Eeg-based Emotion Recognition Using an End-To-End Regional-Asymmetric Convolutional Neural Network. *Knowledge-Based Syst.* 205, 106243. doi:10.1016/j.knsys.2020.106243
- Fang, Y., Yang, H., Zhang, X., Liu, H., and Tao, B. (2020). Multi-feature Input Deep forest for Eeg-Based Emotion Recognition. *Front. Neurobotics* 14, 617531.
- Guo, W., Ong, Y.-S., Zhou, Y., Hervas, J. R., Song, A., and Wei, H. (2019). Fisher Information Matrix of Unipolar Activation Function-Based Multilayer Perceptrons. *IEEE Trans. Cybern.* 49, 3088–3098. doi:10.1109/tcyb.2018.2838680
- Guo, W., Wei, H., Ong, Y., Hervas, J. R., Zhao, J., Wang, H., et al. (2018). Numerical Analysis Near Singularities in RBF Networks. *J. Mach. Learn. Res.* 19, 1–39.
- Hassan, M. M., Alam, M. G. R., Uddin, M. Z., Huda, S., Almogren, A., and Fortino, G. (2019). Human Emotion Recognition Using Deep Belief Network Architecture. *Inf. Fusion* 51, 10–18. doi:10.1016/j.inffus.2018.10.009
- Li, J.-L., Huang, T.-Y., Chang, C.-M., and Lee, C.-C. (2020). A Waveform-Feature Dual branch Acoustic Embedding Network for Emotion Recognition. *Front. Comput. Sci.* 2, 13. doi:10.3389/fcomp.2020.00013

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

WG and GL: Methodology. WG and JY: Validation and investigation. WG: Writing—original draft preparation. GL, JL, and JY: Formal analysis, data curation. JL and JY: Writing—reviewing and editing, and supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61906092, 61802059, 62006119, and 61876085, the Natural Science Foundation of Jiangsu Province of China under Grant Nos. BK20190441, BK20180365, and BK20190444, and the 973 Program No. 2014CB349303.

- Li, J., Zhang, Z., and He, H. (2018). Hierarchical Convolutional Neural Networks for Eeg-Based Emotion Recognition. *Cogn. Comput.* 10, 368–380. doi:10.1007/s12559-017-9533-x
- Li, P., Liu, H., Si, Y., Li, C., Li, F., Zhu, X., et al. (2019). EEG Based Emotion Recognition by Combining Functional Connectivity Network and Local Activations. *IEEE Trans. Biomed. Eng.* 66, 2869–2881. doi:10.1109/tbme.2019.2897651
- Liao, Z., Drummond, T., Reid, I., and Carneiro, G. (2020). Approximate Fisher Information Matrix to Characterize the Training of Deep Neural Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 15–26. doi:10.1109/tpami.2018.2876413
- Ma, J., Tang, H., Zheng, W., and Lu, B. (2019). “Emotion Recognition Using Multimodal Residual LSTM Network,” in Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, October 21–25, 2019, 176–183. doi:10.1145/3343031.3350871
- Mauss, I. B., and Robinson, M. D. (2009). Measures of Emotion: A Review. *Cogn. Emot.* 23, 209–237. doi:10.1080/02699930802204677
- Natarajan, Y., Srihari, K., Chandragandhi, S., Raja, R. A., Dhiman, G., and Kaur, A. (2021). Analysis of Protein-Ligand Interactions of Sars-Cov-2 against Selective Drug Using Deep Neural Networks. *Big Data Min. Anal.* 4, 76–83.
- Nawaz, R., Cheah, K. H., Nisar, H., and Yap, V. V. (2020). Comparison of Different Feature Extraction Methods for Eeg-Based Emotion Recognition. *Biocybernetics Biomed. Eng.* 40, 910–926. doi:10.1016/j.bbe.2020.04.005
- Ng, H., Nguyen, V. D., Vonikakis, V., and Winkler, S. (2015). “Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, November 09–13, 2015. Editors Z. Zhang, P. Cohen, D. Bohus, R. Horaud, and H. Meng, 443–449. doi:10.1145/2818346.2830593
- Nitta, T. (2016). “On the Singularity in Deep Neural Networks,” in Proceedings of 23rd International Conference on Neural Information Processing, Kyoto, Japan, October 16–21, 2016 (Kyoto: Lecture Notes in Computer Science), 389–396. doi:10.1007/978-3-319-46681-1_47
- Nitta, T. (2018). Resolution of Singularities via Deep Complex-Valued Neural Networks. *Math. Meth Appl. Sci.* 41, 4170–4178. doi:10.1002/mma.4434
- Song, T., Liu, S., Zheng, W., Zong, Y., and Cui, Z. (2020). “Instance-adaptive Graph for EEG Emotion Recognition,” in Proceedings of the Thirty-Fourth AAAI

- Conference on Artificial Intelligence, New York, February 7–12, 2020, 2701–2708. doi:10.1609/aaai.v34i03.5656Aaai34
- Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., et al. (2020). Eeg-based Emotion Recognition via Channel-wise Attention and Self Attention. *IEEE Trans. Affective Comput.*, 1. doi:10.1109/TAFFC.2020.3025777
- Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., and Zafeiriou, S. (2017). End-to-end Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE J. Sel. Top. Signal. Process.* 11, 1301–1309. doi:10.1109/jstsp.2017.2764438
- Wei-Long Zheng, W., and Bao-Liang Lu, B. (2015). Investigating Critical Frequency Bands and Channels for Eeg-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi:10.1109/tamd.2015.2431497
- Yang, Y., Fu, Z., Zhan, D., Liu, Z., and Jiang, Y. (2021a). Semi-supervised Multi-Modal Multi-Instance Multi-Label Deep Network with Optimal Transport. *IEEE Trans. Knowl. Data Eng.* 33, 696–709.
- Yang, Y., Wu, Q. M. J., Zheng, W.-L., and Lu, B.-L. (2018a). Eeg-based Emotion Recognition Using Hierarchical Network with Subnetwork Nodes. *IEEE Trans. Cogn. Dev. Syst.* 10, 408–419. doi:10.1109/tcds.2017.2685338
- Yang, Y., Wu, Y., Zhan, D., Liu, Z., and Jiang, Y. (2018b). “Complex Object Classification: A Multi-Modal Multi-Instance Multi-Label Deep Network with Optimal Transport,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, August 19–23, 2018, 2594–2603.
- Yang, Y., Zhan, D., Wu, Y., Liu, Z., Xiong, H., and Jiang, Y. (2021b). Semi-supervised Multi-Modal Clustering and Classification with Incomplete Modalities. *IEEE Trans. Knowl. Data Eng.* 33, 682–695.
- Yang, Y., Zhou, D., Zhan, D., Xiong, H., and Jiang, Y. (2019). “Adaptive Deep Models for Incremental Learning: Considering Capacity Scalability and Sustainability,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, August 4–8, 2019, 74–82.
- Yin, Y., Zheng, X., Hu, B., Zhang, Y., and Cui, X. (2021). EEG Emotion Recognition Using Fusion Model of Graph Convolutional Neural Networks and LSTM. *Appl. Soft Comput.* 100, 106954. doi:10.1016/j.asoc.2020.106954
- Zhang, T., Zheng, W., Cui, Z., Zong, Y., and Li, Y. (2019). Spatial-temporal Recurrent Neural Network for Emotion Recognition. *IEEE Trans. Cybern.* 49, 839–847. doi:10.1109/tcyb.2017.2788081
- Zhao, L., Yan, X., and Lu, B. (2021). “Plug-and-play Domain Adaptation for Cross-Subject Eeg-Based Emotion Recognition,” in Proceedings of thirty-Fifth AAAI Conference on Artificial Intelligence, Virtual Event, February 2–9, 2021 863–870.
- Zheng, W. (2017). Multichannel Eeg-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis. *IEEE Trans. Cogn. Dev. Syst.* 9, 281–290. doi:10.1109/tcds.2016.2587290
- Zheng, W., Zhu, J., Peng, Y., and Lu, B. (2014). Eeg-based Emotion Classification Using Deep Belief Networks. *IEEE Int. Conf. Multimedia Expo*, 1–6. doi:10.1109/icme.2014.6890166
- Zhong, P., Wang, D., and Miao, C. (2020). Eeg-based Emotion Recognition Using Regularized Graph Neural Networks. *IEEE Trans. Affective Comput.*, 1. doi:10.1109/TAFFC.2020.2994159
- Zhu, K., and Zhang, T. (2021). Deep Reinforcement Learning Based mobile Robot Navigation: A Review. *Tsinghua Sci. Technol.* 26, 674–691. doi:10.26599/tst.2021.9010012
- Zong, Y., Zheng, W., Cui, Z., and Li, Q. (2016). Double Sparse Learning Model for Speech Emotion Recognition. *Electron. Lett.* 52, 1410–1412. doi:10.1049/el.2016.1211

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Guo, Li, Lu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.