This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,

Memristor-Based Hierarchical Attention Network for Multimodal Affective Computing in Mental Health Monitoring

Zhekang Dong, *Member*, *IEEE*, Xiaoyue Ji, *Student Member*, *IEEE*, Chun Sing Lai, *Senior Member*, *IEEE*, Donglian Qi, *Senior Member*, *IEEE*, Guangdong Zhou and Loi Lei Lai, *Life Fellow*, *IEEE*

***Abstract*—We present a circuit design of the hierarchical attention network for multimodal affective computing, which can be used in mental health monitoring. Specifically, a kind of cost-effective memristor is fabricated using the albumen protein, and the corresponding testing performance is conducted to ensure its efficiency and stability. Then, considering the hierarchical mechanism inspired by the human limbic system, the nanoscale memristors arranged in a crossbar array configuration are further applied to construct a compact hierarchical attention network that can perform the multimodal affective computing. Furthermore, based on the wearable technology and flexible electronics technology, a mental health monitoring system with low privacy invasiveness, low energy consumption, and low fabrication cost can be designed. Based on the mapping relationship between the multimodal affective computing and mental health, the mental health state of the current user can be monitored. This study is expected to help achieving the deep integration of neuromorphic electronics and mental health monitoring system, further promoting the development of next-generation consumer healthcare technology in smart city.**

***Index Terms*—Hierarchical attention network, multimodal affective computing, human limbic system, mental health monitoring**

I. Introduction

With the progressive increase of stress, anxiety and depression in social environment, mental health problems are as prevalent as other common physical disease or injury, and these problems appear to be increasing in recent years [1]. So far, the mental health problems have been the leading causes of behavioural adjustment difficulties or even disability in many countries [2]. Generally, clinicians evaluate the mental health of people through psychological evaluations and questionnaires, which suffer from subjectivity and timeliness [3]. Mental health monitoring approaches have attracted more and more attention in recent years since they can effectively establish the relationship of signal processing and mental health [4-8]. In [4], researchers proposed a highly integrated multimodal signal system for mental health prediction, which can infer the user's current mental state. In [5], researchers designed a machine learning algorithm to distinguish between happy and unhappy, which evaluated the relationship between different opponents of emotion including happiness, health, energy, alertness and stress. Chernykhet al. achieved emotion classification by using Long Short-Term Memory (LSTM) network [6], and Hanet.al built an improved LSTM network through the DenseNet structure to further enhance the accuracy [7]. A novel solution to targeted aspect-based sentiment analysis was proposed, which solved the challenges of sentiment analysis by exploiting common sense knowledge [8]. However, traditional health monitoring systems are facing challenges and barriers. Specifically, the implementation and operation of such traditional mental health monitoring methods are costly and energy intensive. Meanwhile, these methods are not particularly designed for people in the early stages of developing mental health problems and may raise severe privacy concerns in the present society [9]. Furthermore, traditional health monitoring systems based on von Neumann computer architecture still suffer from computationally challenging problems with unattainable energy efficiencies. To realize mental health monitoring in smart home/city, advanced artificial intelligence (AI) technologies such as nanotechnology, intelligent sensing, and deep learning can be seamlessly incorporated into the developed mental health monitoring system, to facilitate higher-dimensional and more depth data analysis.

The emerging of the nanoscale memristor provides a new approach to realize the mental health monitoring, offering benefits in terms of good privacy, low cost, and easily deployable software [10, 11]. As the fourth basic circuit element, memristor was first proposed by L. O. Chua in 1971 [10] and was further associated with physical devices by R. Stanley Williams and his team from Hewlett-Packard Labs in 2008 [11]. With the development of memristor theory and technology, this nanoscale electronics device has been proved

Manuscript received December 31, 2021.

This work was supported in part by the National Natural Science Foundation of China under Grant U1909201, Grant 62001149, and Natural Science Foundation of Zhejiang Province under Grant LQ21F010009 *(Corresponding author: Chun sing Lai).*

Z. Dong is with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China, 310018, and also with the College of Electrical Engineering, Zhejiang University, Hangzhou, China, 310027, (e-mail: englishp@hdu.edu.cn).

X. Ji and D. Qi are with the College of Electrical Engineering, Zhejiang University, Hangzhou, China, 310027, (e-mail: ji.xiaoyue@zju.edu.cn; qidl@zju.edu.cn).

C. S. Lai is with the Department of Electronic and Electrical Engineering, Brunel University London, London, UB8 3PH, UK and also with the School of Automation, Guangdong University of Technology, Guangzhou, China, 510006, (email: chunsing.lai@brunel.ac.uk).

G. Zhou is with the College of Artificial Intelligence, Southwest University, Chongqing, China, 400715, (e-mail: zhougd@swu.edu.cn).

L. L. Lai is with the School of Automation, Guangdong University of Technology, Guangzhou, China 510006, (e-mail: l.l.lai@ieee.org).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,

effective in the fields of intelligent sensing and deep learning. For example, a memristor-based artificial sensory neuron with visual-haptic fusion was fabricated in [12], which enabled the manipulating of skeletal myotubes and a robotic hand. A novel memristor-based edge computing system for image recognition was proposed in [13], which had advantages in computing resources, calculation time, and energy consumption. In [14], researchers demonstrated in situ training of both the spatial shared-weight Convolutional Neural Network (CNN) and the spatial-temporal shared-weight Convolutional LSTM (ConvLSTM) using memristor crossbar arrays. Therefore, combining memristor technology with intelligent sensing and deep-learning methods for mental health monitoring is desirable and achievable, which can reduce the system complexity and the computational burden, while maintaining low cost and comfort. The main contributions are briefly summarized as below:

1) The circuit designs of LSTM module and self-attention module are proposed using the cost-effective, reliable, flexible, and eco-friendly albumen protein memristor, enabling a parallel-computed and highly integrated neuromorphic system.

2) Inspired by the human limbic system, a memristor-based hierarchical attention network is designed which can perform the multimodal affective computing. Compared with the existing mainstream methods, the proposed network achieves good performance especially in processing time.

3) Based on the wearable technology and flexible electronics technology, a mental health monitoring system with low privacy invasiveness, low energy consumption, and low fabrication cost is designed.

The outline of the paper is organized as follows: Section II briefly describes the fabrication process of the albumen protein memristor and the testing performance is conducted to explore its electrical characteristics. In Section III, the memristor-based hierarchical attention network is designed according to the human limbic system. Section IV describes the specific process of the multimodal affective computing, and provides the classification results compared with existing mainstream methods. In Section V, a schematic of mental health monitoring is proposed combined with the classification results. Finally, Section VI concludes the entire work.

II. Memristor Fabrication and Testing Performance

So far, a variety of memristor models with different physical mechanisms and materials have been proposed. As a natural biological material, albumin protein does not require additional chemical extraction and purification processes, which reduces manufacturing costs and simplifies processing steps [15]. Considering the renewability, biocompatibility and pollution-free property, a kind of organic memristor is fabricated using albumin protein, as shown in Fig. 1.

The specific preparation process is provided as below:

Step 1: H_2O_2 solutions with concentrations of 5%, 10%, 15% and 30% are prepared, respectively.

Step 2: The pristine egg albumen solutions and H_2O_2 solutions with different concentrations are dispersed into deionized water at a volume ratio of 1:10 and continuously stirred using a magnetic stirrer for 30 minutes.

Step 3: A white flocculate appears with stirring, which can be used to fabricate the precursor after filtering operation.

Step 4: The prepared precursor is continuously spin-coated on a flexible ITO plastic substrate at speed of 4,500 rpm for 60 seconds.

Step 5: The flexible ITO substrate is transferred to a muffle furnace and annealed at 97°C in ambient atmosphere for 3 hours, and then the H_2O_2-egg-albumen-based active film is formed.

Step 6: Ag electrodes with a diameter of 200μm and a thickness of 120±10nm is fabricated by physical magnetron sputtering to cover the active egg albumen. In this way, the Ag/Egg Albumin/ITO memristor is fabricated.

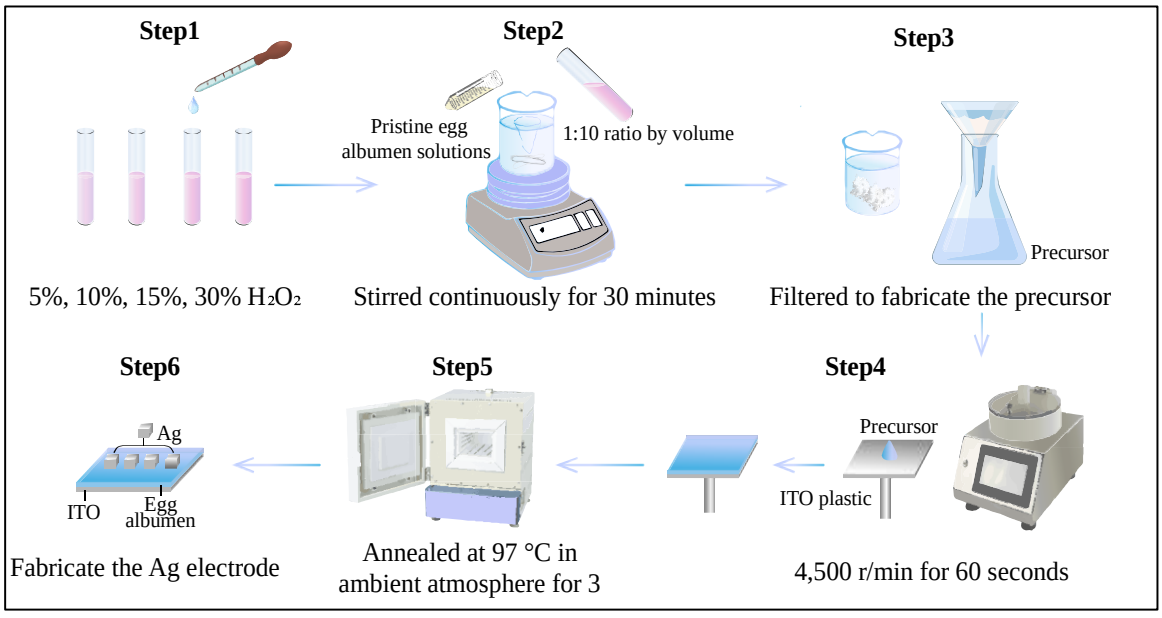
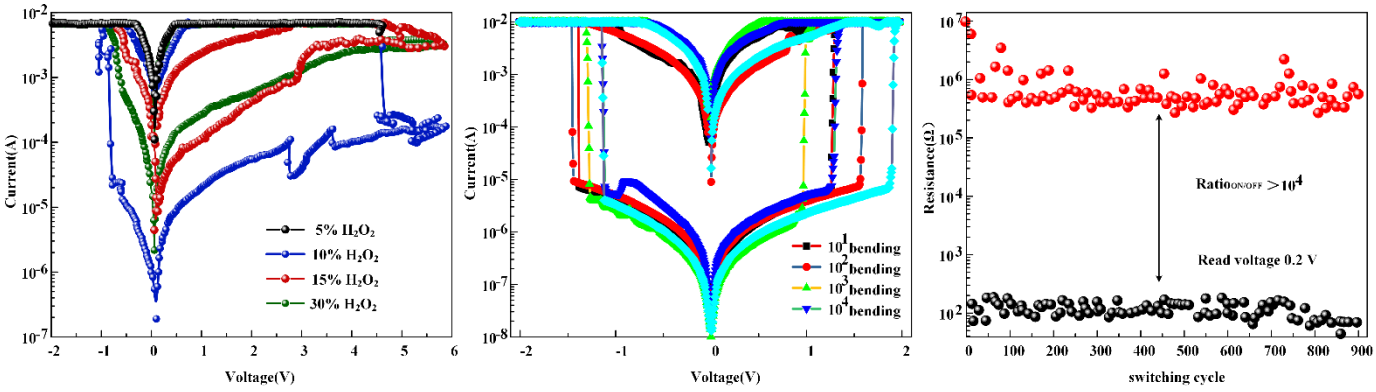


Fig. 1 The specific preparation process of albumen protein memristor

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,



(a) (b) (c)

Fig. 2 The performance test of albumen protein memristor. (a) The I-V curves of the Ag/Egg Albumin/ITO memristor with concentrations of 5%, 10%, 15% and 30%; (b) The I–V curves of the Ag/Egg Albumin/ITO memristor with concentration of 10% are measured after mechanical bending 10^1, 10^2, 10^3, 10^4 times; (c) The retention endurance of an Ag/Egg Albumin/ITO memristor.

In this work, the electrochemical workstation CHI-600D is used to test the performance of the Ag/Egg Albumin/ITO memristor at room temperature. The typical I-V hysteresis curves of Ag/Egg Albumin/ITO memristor demonstrate that resistive switching behaviour can be observed in all Ag/Egg Albumin/ITO memristors with different H_2O_2 concentrations, as shown in Fig. 2(a). Specifically, the Ag/Egg Albumin/ITO memristor with concentration of 10% exhibits a significant resistive switching effect. While the Ag/Egg Albumin/ITO memristors with concentration of 5%, 15% and 30% show poor resistive switching behaviour. Therefore, an optimized memristor is developed using the 10% H_2O_2 solution modified egg albumen as a functional film at room temperature. In order to test the flexibility of the device, the I–V curves are measured after the device mechanically bending 10^1, 10^2, 10^3, 10^4 times, as shown in Fig. 2(b). The resistive switching behaviour can be maintained after the device mechanically bending over 10^4 times, indicating that the Ag/Egg Albumin/ITO memristor have good performance against mechanical bending. To study the endurance of the device, a 0.2 V reading voltage is applied both in the high resistive state and low resistive state for 900 switching cycles, as shown in Fig. 2(c). Notably, the specific time of a single cycle can be achieved by electrochemical workstation using the cyclic voltammetry technique [16]. In this work, a single cycle can be calculated and defined as 120 seconds.

A resistance ratio between the high resistive state and low resistive state of over 10^4 can be maintained during the retention time indicating that the prepared memristor has good stability and endurance.

III. MEMRISTOR-BASED HIERARCHICAL ATTENTION NETWORK FOR AFFECTIVE COMPUTING

With the development of the basic brain neuroscience, the knowledge of the correlation between the human brain and its emotion/behaviour control have been developed recent years. It is currently accepted that the human limbic system (as shown in Fig. 3(a)), mainly composed of the amygdala, hippocampus, insula, anterior cingulate cortex, and prefrontal cortex, participates in almost all of the emotional processes [17]. Normally, there are two basic forms of sensory information existing in the emotional process [18]: The first form occurs through the lateral cortex, which mainly refers to the nonspecific sensory activities dominated by the dorsal thalamus. The sensory information will be analysed roughly, and the thalamus can provide a decision whether the external stimulus is a known one within a short time. The second form occurs through the cingulate cortex and hippocampus, which is directly related to hypothalamic activity. The sensory information will be transmitted slower, and a detailed and comprehensive analysis can be given. Inspired by this mechanism, the fabricated memristor is utilized to construct a

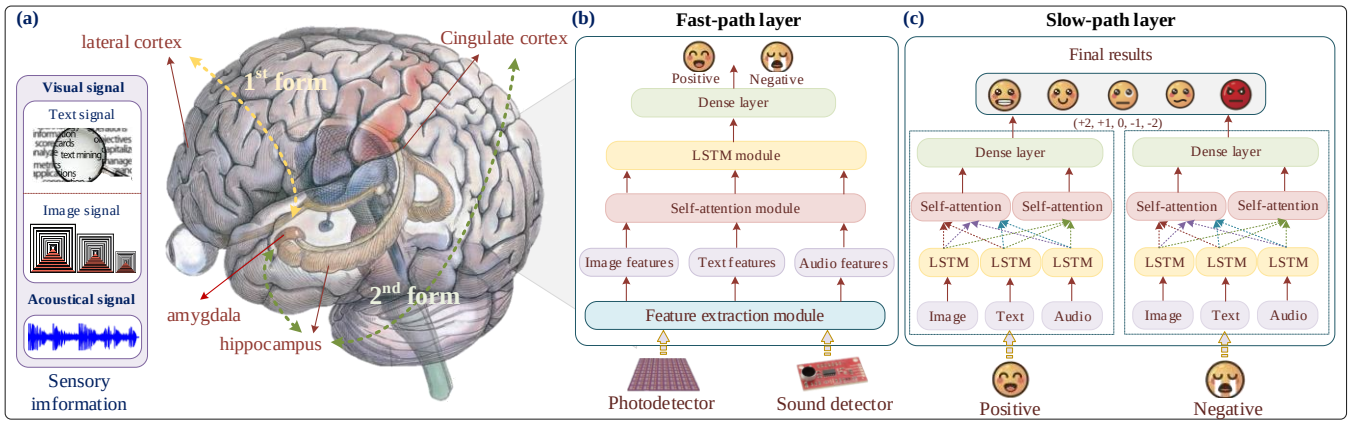


Fig. 3 The overall framework of the memristor-based hierarchical attention network. (a) The limbic system of human brain; (b) The structure of the fast-path layer; (c) The structure of the slow-path layer

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,

hierarchical attention network in this section, which can be further applied to perform the multimodal affective computing. Specifically, the overall framework of the memristor-based hierarchical attention network is provided as below:

From Fig. 3, the entire memristor-based hierarchical attention network consists of two parts, i.e., the fast-path layer (corresponding to the 1st form, as shown in Fig. 3(b)) and the slow-path layer (corresponding to the 2nd form, as shown in Fig. 3(c)). The former is used to realize a binary classification task (i.e., positive and negative), while the latter is used to perform the five-class classification task (i.e., joy, anger, sorrow, happiness, and neutrality). Notably, the structures of the fast-path layer and the slow-path layer mainly conclude four components, i.e., the LSTM module, the self-attention module, the dense layer module, and the feature extraction module. Particularly, the circuit implementation of the dense layer module and the feature extraction module is relatively simple and easy to achieve [19, 20]. The specific implementation scheme can be found in [19, 20], and we will not repeat any details here. This section mainly focuses on the circuit design of LSTM module and self-attention module, due to the fact that the relative research is insufficient and challenging. The specific process description is provided as below.

A. *Circuit design of LSTM module*

LSTM network is a subclass of recurrent neural network (RNN) that specializes in machine learning tasks involving sequences, especially in sentiment analysis. A standard LSTM cell contains three gates, i.e., the forget gate, the input gate, and the output gate, as shown in Fig. 4.

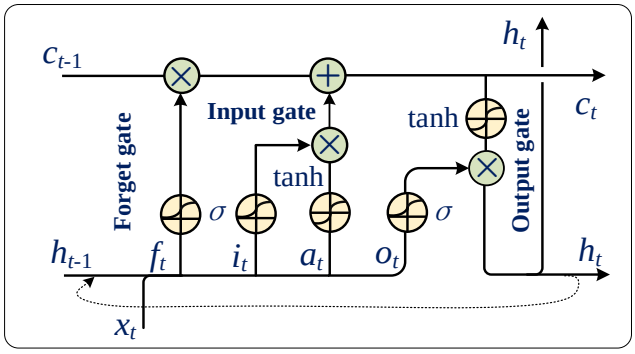


Fig. 4. The internal structure of an LSTM cell

From Fig. 4. x_t and c_t denote the input vector and the cell state at the present step, respectively. h_t and h_{t-1} are the output vectors at the present and previous time steps, respectively. σ is the logistic sigmoid function, which yields i_t, f_t, and o_t via the network parameters weight $\boldsymbol{W}$ (W_a, W_f, W_i, and W_o), recurrent weight $\boldsymbol{U}$ (U_a, U_f, U_i, and U_o), and bias $\boldsymbol{b}$ (b_a, b_f, b_i, and b_o). It is noted that the internal operating mechanism of an LSTM cell can be characterized by two main phases, i.e., the linear matrix operation and the gated nonlinear activation [6].

Based on this, an LSTM unit consisted of two sub-circuits, i.e., the linear matrix operation circuit and the nonlinear activation circuit, is designed in Fig. 5(a). Meanwhile, it is clear that an LSTM cell can be constructed by four LSTM units, the specific circuit design of an LSTM cell is provided in Fig. 5(b). Particularly, there are many LSTM variants (e.g., the Bi-LSTM [21], GRU [22], ConvLSTM [14]) have been proposed recently. For different LSTM variants, the corresponding circuit structure can also be design by recombining the fundamental linear matrix operation circuit and the activation circuit.

To better illustrate the circuit design scheme in Fig. 5, the relevant circuit analysis is provided as below:

The linear matrix operation circuit: the linear matrix operation circuit is consisted of two albumen protein memristor crossbar arrays, an auxiliary circuit, and a biasing circuit. The memristor crossbar array is used to perform the matrix-vector multiplication in LSTM, and the biasing circuit is responsible for the sum operation.

Assuming $R_{a1}=R_{a2}=R_{a3}=R_{a4}$ (the resistors in auxiliary circuit) and $R_{b1}=R_{b2}=R_{b3}=R_{b4}=0.5R_{b5}$ (the resistors in biasing circuit), the output of the matrix operation circuit V_{ob} can be mathematically expressed by $V_{ob}=V_xW+V_hU+V_b$. Here, V_x and V_h denote the applied normalized voltage, bias voltage V_b is the input of the biasing circuit (labeled by the green rectangle), representing the bias parameter in LSTM. $W=G_{w,ij+}-G_{w,ij-}$ and $U=G_{u,ij+}-G_{u,ij-}$ denote the weights and recurrent weights, respectively. Notably, $G_{w,ij}$ and $G_{u,ij}$ are the conductance of the albumen protein memristor.

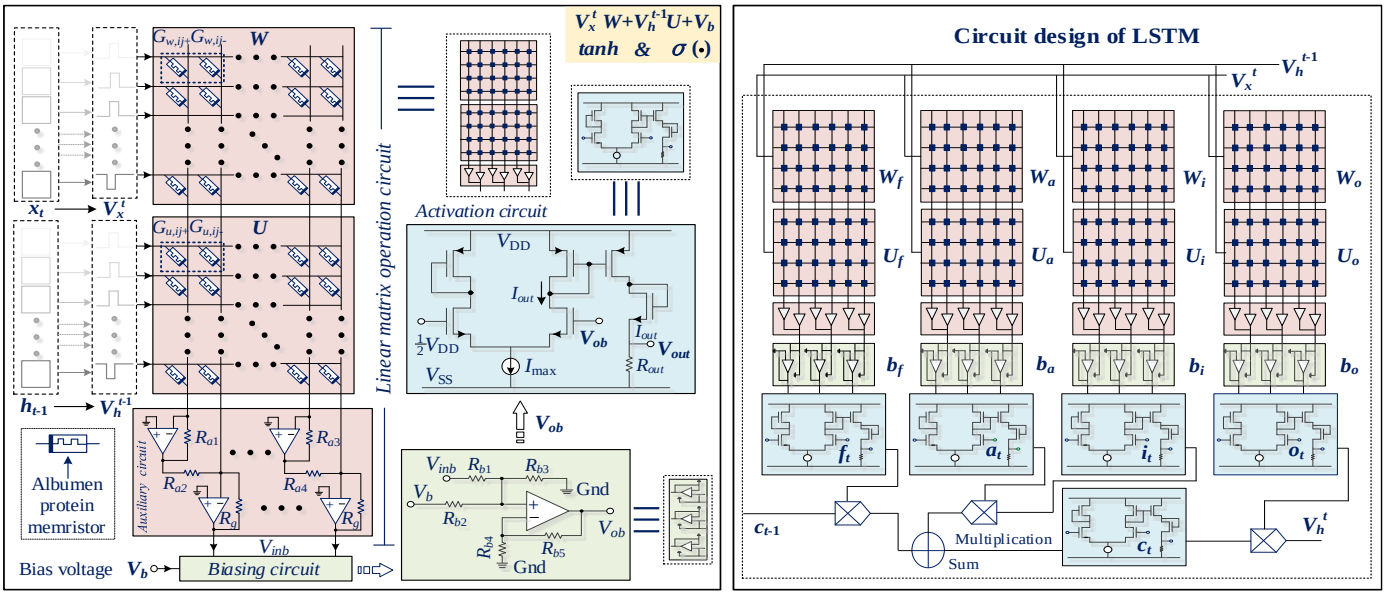
The activation circuit: the activation circuit (labeled by the blue rectangle) is used to realize the nonlinear activation function. The input voltage V_{ob} is connected to one side of an NMOS source-coupled pair, biased with a current sink $I_{\max}$. The output current I_{out} is computed by the product of the normalized current I_n and the current sink $I_{\max}$. Notably, the normalized current I_n with a piecewise form can be achieved by [23]. According to the Ohm's law, the output voltage V_{out} can be calculated by $V_{\text{out}}=I_n\cdot I_{\max}\cdot R_{\text{out}}+V_{\text{ss}}$. If we assume the load resistance $R_{out}=1\Omega$ is a constant resistor, the output voltage V_{out} can be rewritten by $V_{\text{out}}=I_n\cdot I_{\max}+V_{\text{ss}}$. Notably, this activation circuit can be used to realize both the logistic sigmoid function and the hyperbolic tangent function by tuning circuit parameters.

B. *Circuit design of Self-attention module*

Self-attention mechanism can effectively identify the relations among input data, which aims at reducing the fundamental constraint of sequential computation in LSTM network [24]. The self-attention mechanism can be described as mapping a query and a set of key-value pairs, where the query (Q), keys (K), values (V), and output are all vectors. as shown in the bottom left of Fig. 6. Notably, unlike the LSTM module, the input of the self-attention module is voltage, while the output of the self-attention module is current.

From Fig. 6, $x\in R^{N\times d}$ is the input matrix, parameter N and parameter d are the sequence length and the hidden dimension, respectively. Connection weights (W^q, W^k, W^v and W) $\in R^{N\times}$ are all learnable parameters. Following the weights (W^q, W^k, and W^v), the input matrix x can be converted into the attention query, key and value matrices Q, K, and $V\in R^{N\times d}$. In this work, we compute the additive attention between each query vector and each key vector and apply a SoftMax function to normalize the additive attention scores. Then, we compute the dot product of the value matrix V and the corresponding normalized attention score as the output.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,



(a) (b)

Fig. 5. The specific circuit design scheme. (a) Circuit design of a LSTM unit; (b) Circuit design of a LSTM cell

Similarly, the self-attention mechanism can also be characterized by two main phases, i.e., the similarity computation and the dot product operation. Based on this, a self-attention circuit consisted of two sub-circuits, i.e., the similarity operation circuit and the dot product operation circuit are designed in Fig. 6. To better illustrate the circuit design scheme in Fig. 6, the relevant circuit analysis is provided as below:

The similarity computation circuit: the similarity computation circuit is consisted of three albumen protein memristor crossbar arrays, a *tanh* circuit and a SoftMax circuit. The memristor crossbar arrays are used to compute and store the parameters of each row in the key matrix *K*, the query matrix *Q*, and the matrix *W*, respectively. The *tanh* circuit can directly use the above-mentioned activation circuit, realizing the nonlinear activation function. The SoftMax circuit is responsible for the normalization operation [25].

The dot product operation circuit: the dot product operation circuit is consisted of an albumen protein memristor crossbar array and several multiplying units. The albumen protein memristor crossbar array is used to compute and store the parameters of each row in the value matrix *V*, and multiplying units compute the weighted sum of rows in value matrix *V* and the corresponding SoftMax circuit outputs, i.e., the final outputs of the self-attention circuit.

IV. Multimodal Affective Computing

In this section, the proposed memristor-based hierarchical

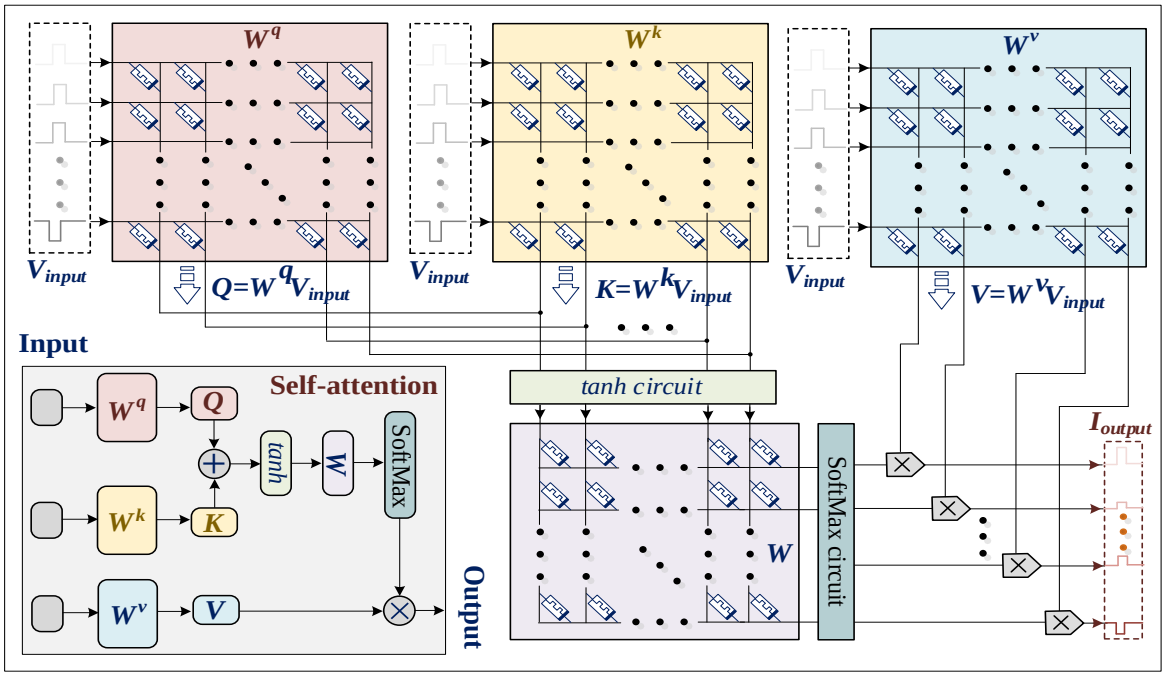


Fig. 6. The circuit design of self-attention module

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,

attention network (as shown in Fig. 3) is used to perform the multimodal affective computing, and the entire process can be divided into two phases: a binary classification phase conducted by the fast-path layer, and a five-class classification phase conducted by the slow-path layer. The details are provided as follows:

A binary classification phase:

As shown in Fig. 3(b), when the photodetector [26] and sound detector [27] collect the sensory information, the visual and acoustical signal in the form of voltage pulses can be obtained successfully. Notably, the visual signal contains two parts here, i.e., the image part and the text part. Then, the hierarchical feature representation (i.e., the image features, text features, and audio features) can be achieved by the feature extraction module [20]. Next, these hierarchical features are fused using the self-attention module and further injected to the LSTM module. Finally, the outputs of the LSTM module are input to the dense layer, the classification result 0/1 (0 represents negative, while 1 represents positive) is obtained. The binary cross entropy loss function [28] is used as the performance function to train and optimize the parameters in LSTM.

A five-class classification phase:

According to the classification results obtained by the fast-path layer, the original hierarchical features (image features, text features, and audio features) can be further divided into two subsets, i.e., the negative subset and the positive subset. Actually, the slow-path layer can be deemed as a dual-channel structure, as shown in Fig. 3(c). Notably, these two channels are the same with each other, the only difference is the input signal (one is the negative subset, the other is the positive subset). Then, the hierarchical features are injected to the three respective LSTM modules, and three initial classification results based on different modals can be achieved. Next, a combined set formed by any two initial classification results are input to a self-attention module for the fusion operation, and a combined set containing all three initial classification results are input to another self-attention module for realizing the same task. Finally, the outputs of the two self-attention modules are further injected to the dense layer, and the final classification result with five possible values (+2, +1, 0, -1, and -2) can be obtained. Here, the categorical cross entropy loss function [29] is utilized as the performance function to train and optimize the parameters in LSTM.

Based on the above-mentioned two phases, an effective verification process is necessary. So far, there have been multiple standard datasets [30] for multimodal sentiment analysis, for example, YouTube, ICT-MMMO, CMU-MOSI, MOUD, and CMU-MOSEI datasets. Considering the applicability of dataset, CMU-MOSI and CMU-MOSEI dataset are used for verification purpose. Notably, CMU-MOSI contains 2199 utterance-video segments sliced from 93 videos played by 89 distinct narrators. Each segment is manually annotated with a real-valued score within the range of [-3, +3]. As the next generation of CMU-MOSI dataset, CMU-MOSEI dataset contains 23453 annotated segments and 250 topics from approximately 1000 different speakers. Each utterance in CMU-MOSEI dataset corresponds to an integer label from -3 to +3, indicating seven types of emotional categories. Notably, since up to five different classes (i.e., joy, anger, sorrow, happiness, and neutrality) need to be recognized in this work, the data with the label greater than 2 or less than -2 will not be considered.

Furthermore, comparison experiments are conducted and the final classification results are collected in Table I.

From Table 1, the classification results on the test set of CMU-MOSI and CMU-MOSEI dataset are exhibited. Firstly, considering the control variable rule, the modality of all the sentiment recognition methods is audio, image, and text. Clearly, the proposed method achieves the best performance, compared with the other competitors. Specifically, for CMU-MOSI dataset, the accuracy of the proposed method (five-class classification task) is 68.31% that is slightly higher ~1% than that of the second place. For CMU-MOSEI dataset, the accuracy is higher 1.2% than that of the second place. Meanwhile, although the proposed method is a two-step strategy, the total time is much less than the other one-step methods, which provides the efficiency advantage in subsequent mental health monitoring. Notably, the overall accuracy of the five-class classification task is insufficiently high (~70%), which is an inherent issue of existing mainstream affective computing methods especially when the number of categories is more than 4. According to [38], this problem can be addressed by building a more universal dataset and a more efficient feature fusion mechanism.

Table I. The collection of the classification results obtained by different methods

Methods	Modalities	CMU-MOSI			CMU-MOSEI		
		Procedure	Accuracy	Time	Procedure	Accuracy	Time
The proposed method	Audio + image + text	Fast-path	81.20%	0.73 Sec	Fast-path	84.31%	0.98 Sec
	Audio + image + text	Slow-path	68.31%	1.15 Sec	Slow-path	70.67%	1.16 Sec
Reference [31]	Audio + image + text	–	65.54%	4.23 Sec	–	66.39%	4.31 Sec
Reference [32]	Audio + image + text	–	62.19%	5.42 Sec	–	64.22%	5.19 Sec
Reference [33]	Audio + image + text	–	67.43%	4.65 Sec	–	68.54%	4.48 Sec
Reference [34]	Audio + image + text	–	66.58%	5.19 Sec	–	67.31%	5.21 Sec
Reference [35]	Audio + image + text	–	65.41%	6.21 Sec	–	66.88%	6.34 Sec
Reference [36]	Audio + image + text	–	63.22%	5.22 Sec	–	67.31%	5.53 Sec
Reference [37]	Audio + image + text	–	61.16%	5.44 Sec	–	67.31%	5.61 Sec

Note: The proposed method is a two-step method, the accuracy of fast-path process is the result of binary classification task, correspondingly, the accuracy of the slow-path process is the result of five-class classification task.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/MMUL.2019.2954405,

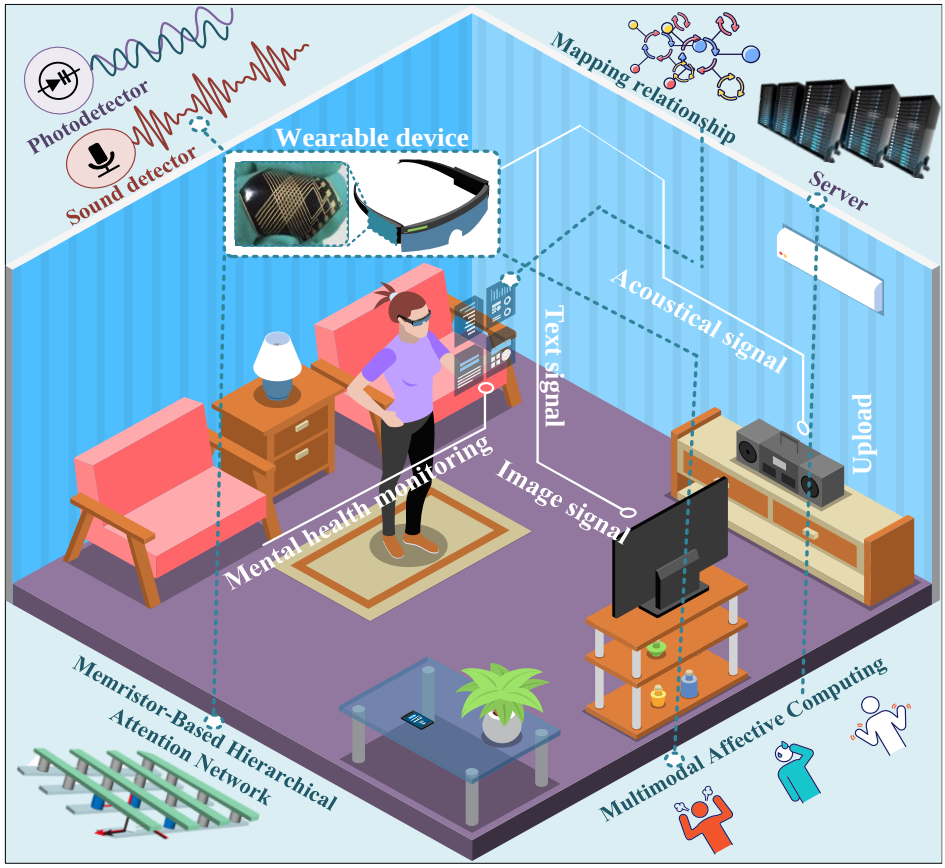


Fig. 7. The schematic of mental health monitoring system using wearable technology and flexible electronics technology

V. Mental Health Monitoring Using Wearable Technology and Flexible Electronics Technology

In this section, a schematic of mental health monitoring system integrating the wearable technology and flexible electronics technology is illustrated in Fig. 7, which can be used in the smart city/home areas for various intelligent monitoring and healthcare.

From Fig. 7, the multisensory module and the memristor-based hierarchical attention network can be easily integrated into a small rectangle neuromorphic computing film and further installed in a wearable device (e.g., glasses). Notably, the multisensory module is used to collect sensory information and the memristor-based hierarchical attention network is used to perform multimodal affective computing. The multisensory module essentially encompasses two multimodal sensory channels: the visual, and the audio channel, which can convert multimodal information into voltage signals. The multisensory module on the glasses can effectively extend the sensing range to detect the variation of information (e.g., the image signal, text signal, and acoustical signal) in home environment, offering a more private way for affective computing in smart glasses with the advantages of low cost and eco-friendly fabrication. With aides of memristor-based hierarchical attention network, the emotion from the user can be recognized effectively. Next, all affective states obtained by the smart glasses will be uploaded to a server (storing the mapping relationship between the mental health state and the multimodal affective computing) for further analysis and decision making. Notably, this mapping relationship can be achieved through many machine learning methods [4] or expert experience and knowledge. After that, the corresponding result will be shown on guardian's (or user's) mobile device for long-term health monitoring and the valuable clinical mental health diagnosis. Notably, the processor of the hierarchical attention network is mainly composed by LSTM module and self-attention module. Based on this, the proposed hierarchical attention network can be regarded as a general intelligent computing system that has the potential to perform some LSTM/self-attention-based recognition applications.

The mental health monitoring system using wearable technology and flexible electronics technology can handle the multimodal data from the collected data on the hardware side to the artificial intelligence comprehension on the software side, and finally can be combined with Internet of Things (IOT) to promote the development of smart home/city.

VI. Conclusions

This paper focuses on the investigation of mental health monitoring method using neuromorphic sensing devices, brain neuroscience and computer science theory. Specifically, a kind of organic memristor is firstly prepared using the albumen protein, the corresponding testing performance demonstrate the electrical characteristics of the nanoscale device. Then, the albumen protein memristor pairs with the crossbar array configuration are used to design the LSTM module and the

self-attention module. Combining with the existing dense layer module and the feature extraction module, a memristor-based hierarchical attention network is designed for multimodal affective computing. Notably, inspired by the human limbic system, the structure of the entire network can be further divided to two parts, i.e., the fast-path layer and the slow-path layer. The former one can be used to perform the binary classification task, and the latter one can be used to perform the five-class classification task. Compared with the existing mainstream affective computing methods, the proposed method achieves the best performance in terms of accuracy (slightly increase ~1%) and processing time (~62.4% reduction). Furthermore, a mental health monitoring system integrating the wearable technology and the flexible electronics technology is designed, where the constructed memristor-based hierarchical attention network and the necessary multisensory module can be easily integrated into a small rectangle neuromorphic computing film and further installed in a wearable device (e.g., glasses). Based on the mapping relationship between the multimodal affective computing and mental health, the mental health state of the current user can be continuously monitored. This work provides a new avenue for the wearable mental health monitoring system development, which is expected to promote the development of consumer healthcare services in smart city.

Notably, mental health monitoring system is still in an infancy stage with abundant opportunities and challenges. To further develop and improve the performance of mental health monitoring system, several aspects can be considered in future research: 1) At the data level: A complete and universal dataset is necessary and important for validation process; 2) At the device level: More reliable and eco-friendly memristor models are required. The memristive mechanism needs to be more thoroughly comprehended to further improve the device performance; 3) At the circuit level: Memristive circuits are desired to be effectively enlarged with efficient read/write schemes. Equipping memristors with transistor-based selectors or rectifying capabilities can reduce the sneak path currents; 4) At the system level: To overcome the device variations, low-precision algorithms and systems can be further studied. Meanwhile, brain-inspired intelligence algorithm can be further explored to solve different practical problems.

REFERENCES

[1] E. Kim, A. Coumar, W. B. Lober and Y. Kim, "Addressing Mental Health Epidemic Among University Students via Web-based, Self-Screening, and Referral System: A Preliminary Study," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 301-307, March 2011, doi: 10.1109/TITB.2011.2107561.

[2] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris and H. Christensen, "We Feel: Mapping Emotion on Twitter," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1246-1252, July 2015, doi: 10.1109/JBHI.2015.2403839.

[3] Q. Yang, "A New Approach to Evidence-Based Practice Evaluation of Mental Health in Psychological Platform under the Background of Internet + Technology," *2019 International Conference on Electronic Engineering and Informatics (EEI)*, 2019, pp. 321-323, doi: 10.1109/EEI48997.2019.00076.

[4] D. Zhou, J. Luo, V. M. Silenzio, Y. Zhou, J. Hu, G. Currier, and Kautz, H, "Tackling mental health by integrating unobtrusive multimodal sensing," in *Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1401-1408, 2015, doi: abs/10.5555/2887007.2887201.

[5] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano and R. Picard, "Predicting students' happiness from physiology, phone, mobility, and behavioral data," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 222-228, 2015, doi: 10.1109/ACII.2015.7344575.

[6] V. Chernykh and P. Prikhodko, "Emotion recognition from speech with recurrent neural networks", 2017, arXiv preprint arXiv:1701.08071.

[7] K. J. Han, A. Chandrashekaran, J. Kim and Lane, I, "The CAPIO 2017 conversational speech recognition system," 2017, arXiv preprint arXiv:1801.00059.

[8] W. Jiao, M. Lyu and I. King, "Real-time emotion recognition via attention gated hierarchical memory network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8002-8009, April 2020, doi: 10.1609/aaai.v34i05.6309.

[9] T. Nakamura, Y. D. Alqurashi, M. J. Morrell and D. P. Mandic, "Hearables: Automatic Overnight Sleep Monitoring With Standardized In-Ear EEG Sensor," in *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 203-212, Jan. 2020, doi: 10.1109/TBME.2019.2911423.

[10] L. Chua, "Memristor-The missing circuit element," in *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 507-519, September 1971, doi: 10.1109/TCT.1971.1083337.

[11] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," in *Nature*, vol. 453, no. 7191, pp. 80–83, May 2008, doi: 10.1038/nature06932.

[12] C. Wan, P. Cai, X. Guo, M. Wang, N. Matsuhisa, L. Yang, et al, "An artificial sensory neuron with visual-haptic fusion," in *Nature communications*, vol. 11, no. 1, pp. 1-9, September 2020, doi: 10.1038/s41467.-020-18375-y.

[13] H. Ran et al., "Memristor-Based Edge Computing of Blaze Block for Image Recognition," in *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2020.3045029.

[14] Z. Wang, C. Li, P. Lin, M. Rao, et al, "In situ training of feed-forward and recurrent convolutional memristor networks," in *Nature Machine Intelligence*, vol. 1, no. 9, pp. 434-442, September 2019, doi: 10.1038/s42256-019-0089-1.

[15] X. Ji, Z. Dong, C. S. Lai and D. Qi, "A Brain-Inspired In-Memory Computing System for Neuronal Communication via Memristive Circuits," in *IEEE Communications Magazine*, vol. 60, no. 1, pp. 100-106, January 2022, doi: 10.1109/MCOM.001.21664.

[16] S. D. Adams, E. H. Doeven, S. J. Tye, K. E. Bennet, M. Berk and A. Z. Kouzani, "TinyFSCV: FSCV for the Masses," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 1, pp. 133-142, Jan. 2020, doi: 10.1109/TNSRE.2019.2956479.

[17] K. V. Baev, "Functions of the limbic system in the brain," *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC) held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA) Intell*, 1998, pp. 522-524, doi: 10.1109/ISIC.1998.713717.

[18] R. Ravi and Mija S.J, "Design of Brain Emotional Learning Based Intelligent Controller (BELBIC) for uncertain systems," *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 2014, pp. 1089-1093, doi: 10.1109/ICACCCT.2014.7019265.

[19] L. Zhang, D. Borggreve, F. Vanselow and R. Brederlow, "Quantization Considerations of Dense Layers in Convolutional Neural Networks for Resistive Crossbar Implementation," *2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2020, pp. 1-6, doi: 10.1109/MOCAST49295.2020.9200280.

[20] S. Razmpour, A. M. Sodagar, M. Faizollah, M. Y. Darmani and M. Nourian, "Reconfigurable biological signal co-processor for feature extraction dedicated to implantable biomedical microsystems," *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 861-864, doi: 10.1109/ISCAS.2013.6571983.

[21] J. Jeon, B. Jeong, S. Baek and Y. -S. Jeong, "Hybrid Malware Detection Based on Bi-LSTM and SPP-Net for Smart IoT," in *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2021.3119778.

[22] M. Xia, H. Shao, X. Ma and C. W. de Silva, "A Stacked GRU-RNN-Based Approach for Predicting Renewable Energy and Electricity Load for Smart Grid Operation," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7050-7059, Oct. 2021, doi: 10.1109/TII.2021.3056867.

[23] Z. Dong, C. S. Lai, Z. Zhang, D. Qi, M. Gao and S. Duan, "Neuromorphic extreme learning machines with bimodal memristive synapses," in *Neurocomputing*, vol.453, pp.38-49, April 2021, doi:

10.1016/j.neucom.2021.04.049.

[24] P. Gao, K. Lu, J. Xue, L. Shao and J. Lyu, "A Coarse-to-Fine Facial Landmark Detection Method Based on Self-attention Mechanism," in *IEEE Transactions on Multimedia*, vol. 23, pp. 926-938, 2021, doi: 10.1109/TMM.2020.2991507.

[25] F. Spagnolo, S. Perri and P. Corsonello, "Aggressive Approximation of the SoftMax Function for Power-Efficient Hardware Implementations," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, doi: 10.1109/TCSII.2021.3120495.

[26] S. Chen, Z. Lou, D. Chen, and G. Shen, "An Artificial Flexible Visual Memory System Based on an UV-Motivated Memristor," in *Advanced Materials*, vol. 30, no. 7, pp. 1705400, January 2018, doi: 10.1002/adma.201705400.

[27] H. Tan, Y. Zhou, Q. Tao, J. Rosen and S. van Dijken, "Bioinspired multisensory neural network with crossmodal integration and recognition," in *Nature communications*, vol. 12, no. 1, pp. 1-9, February 2021, doi: 10.1038/s41467-021-21404-z.

[28] X. Li, L. Yu, D. Chang, Z. Ma and J. Cao, "Dual Cross-Entropy Loss for Small-Sample Fine-Grained Vehicle Classification," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4204-4212, May 2019, doi: 10.1109/TVT.2019.2895651.

[29] Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," in *IEEE Access*, vol. 8, pp. 4806-4813, 2020, doi: 10.1109/ACCESS.2019.2962617.

[30] M. G. Huddar, S. S. Sannakki and V. S. Rajpurohit, "A survey of computational approaches and challenges in multimodal sentiment analysis," in *International Journal of Computer Sciences and Engineering*, vol.7, no.1, pp. 876-883, January 2019, doi: 10.26438/ijcse/v7i1.876883.

[31] A. Agarwal, A. Yadav and D. K. Vishwakarma, "Multimodal Sentiment Analysis via RNN variants," *2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)*, 2019, pp. 19-23, doi: 10.1109/BCD.2019.8885108.

[32] T. Kim and B. Lee, "Multi-attention multimodal sentiment analysis," *In Proceedings of the 2020 International Conference on Multimedia Retrieval*, pp. 436-441, June 2020, doi: 10.1145/3372278.3390698.

[33] Y. Li, K. Zhang, J. Wang and X. Gao, (2021). "A cognitive brain model for multimodal sentiment analysis based on attention neural networks," *Neurocomputing*, vol. 430, pp. 159-173, March 2021, doi: 10.1016/j.neucom.2020.10.021

[34] J. Yu, J. Jiang and R. Xia, "Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 429-439, 2020, doi: 10.1109/TASLP.2019.2957872.

[35] A. Rao, A. Ahuja, S. Kansara and V. Patel, "Sentiment Analysis on User-generated Video, Audio and Text," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. 24-28, doi: 10.1109/ICCCIS51004.2021.9397147.

[36] H. Wang, A. Meghawat, L. -P. Morency and E. P. Xing, "Select-additive learning: Improving generalization in multimodal sentiment analysis," *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 949-954, doi: 10.1109/ICME.2017.8019301.

[37] Z. Luo, H. Xu, and F. Chen, "Utterance-based audio sentiment analysis learned by a parallel combination of cnn and lstm," 2018, arXiv preprint arXiv: 1811.08065v1.

[38] W. Peng, X. Hong and G. Zhao, "Adaptive Modality Distillation for Separable Multimodal Sentiment Analysis," in *IEEE Intelligent Systems*, vol. 36, no. 3, pp. 82-89, 1 May-June 2021, doi: 10.1109/MIS.2021.3057757.