Stress analysis based on feature-level late fusion

Eun-Bin Choi *
Department of Artificial Intelligence Convergence
Chonnam National University
Gwangju, South Korea
iidmsqlss@gmail.com

Hong-Hai Nguyen *
Department of Artificial Intelligence Convergence
Chonnam National University
Gwangju, South Korea
honghaik14@gmail.com

Trong-Nghia Nguyen
Department of Artificial Intelligence Convergence
Chonnam National University
Gwangju, South Korea
trongnghia7171@gmail.com

Soo-Hyung Kim +
Department of Artificial Intelligence Convergence
Chonnam National University
Gwangju, South Korea
shkim@jnu.ac.kr

ABSTRACT

Sentiment analysis is an aspect momentous for application in the future. Especially, emotional stress estimation helps improve our life. In this paper, we solved the problem of stress sub-challenge of Multimodal Sentiment Analysis 2022 (MuSe2022). The aim of the stress task is to predict continuous levels of arousal and valence in continuous times based on audio, video, text and biological signals. In this work, we conducted two steps: feature late fusion and stress estimation. First, various features are provided by the sub-challenge. So we take the strategy of combining features to extract feature fusion. We combined DEEPSPECTRUM and eGeMAPS for arousal; DEEPSPECTRUM, FAUS and Physiological signals for valence by Gated Recurrent Unit (GRU) for multimodal fusion. Second, the valence and arousal prediction were based on combining three ingredients such as Local Attention, GRU, and Bayesian NetWork. The Concordance Correlation Coefficient (CCC) loss function was applied during model training. Our method achieved the good CCC performance of 0.640 on a combination of valence and arousal on the challenge of the development set, which outperforms the baseline method with a corresponding CCC of 0.618 (Audio and Video) on a combination of arousal and valence.

KEYWORDS

ULM-TSST, GRU, Bayesian Network, Local Attention, Stress analysis.

1 INTRODUCTION

Stress is triggered by a physiological and/or psychological event [16] Stress is a serious problem in modern society. If stress develops chronically, it can have negative health issues, such as decreased mental performance, raised blood pressure, lower body recovery, etc [30]. As a result, it can cause low productivity and dissatisfaction with life. Therefore, it is important to manage continuous stress. Also, the effects of stress on health and performance can vary depending on an emotion of a person about stress [8]. Thus, continuous emotional stress estimation help improves our life.

Biological signals are measures of human body processes over time [18] which was divided into physical and physiological signals. The measures of body change are the physical biological signals such as muscle activity, eyes, blinks, head, body, facial expressions, voice, etc. Physiological signals are related to some signals of the body, such as Electrocardiogram (ECG) related to cardiac activity, Respiration (RESP), Heart Rate (BPM), and Electrodermal activity (EDA) related to the exocrine activity. Electrocardiogram(ECG) is the value of the activity of the heart contractile activity. Heart Rate (BPM) is the number of heart beats per minute, which measured in bpm and the most widely adopted and measure to stress estimate. Respiration (RESP) is measured by breath sensors, which estimate the number of air exchange of the lungs. Most studies use RESP and combine it with bio-signals for calculating the stress level of the objects. Electrodermal activity (EDA) is measured as electricity flow through the skin. Even slight changes in the amount of sweat which change the electrical conductivity of the skin [13].

Sentiment Analysis (SA) is the computational study of people's sentiments, emotions and attitudes toward an entity [22]. In other words, SA can be used to automatically estimate the value of the valence or polarity, whether it is positive, negative, or neutral [23]. SA achieves various benefits in our life. For example, SA can measure customer satisfaction, check customer sentiments when shopping, and forecast movie sales [17, 20, 27, 33]. Therefore, SA is applying in many domains and a growing field of research influenced by other domains such as text mining, machine learning, natural language processing, etc [34]. Therefore, SA is a momentous aspect for application in the future.

The Multimodal Sentiment Analysis Challenge (MuSe) 2022 is dedicated to multimodal sentiment and emotion recognition. There are three sub-challenges : Humor Detection Sub-Challenge (MuSe-Humor), Emotional Reactions Sub-Challenge (MuSe-Reaction) and Multimodal Emotional Stress Sub-challenge (MuSe-Stress). The goal of MuSe-Humor is to detect the presence of humor in the football press conference recordings based on the novel Passau Spontaneous Football Humor (Passau-SFCH) dataset. MuSe-Reaction aims to predict the intensities of seven

* Both authors contributed equally to this research.
+ Corresponding author.

self-reported emotions (Adoration, Amusement, Anxiety, Disgust, Empathic Pain, Fear, Surprise) from user-generated reactions to emotionally evocative videos. It is based on the novel Hume-Reaction dataset. MuSe-Stress predicts the valence level and psycho-physiological arousal level over time continuously from audio-visual recordings. It is based on the Ulm-Trier Social Stress Test dataset (ULM-TSST) about people in stressful dispositions [7]. And It consists of audio, video, text and biological signals. In this paper, we present our solutions to the Multimodal Emotional Stress Sub-challenge (MuSe-Stress).

In this study, we solved the problems of stress estimation. Our contributions are summarized as the following:

- Using features which were extracted from the deep learning model.
- Using late fusion strategy to enhance performance.
- Combination of GRU, local attention and Bayesian Neural Network for sentiment analysis.
- Conducting experimental with a lot of method to compare with the baseline and our method.

This paper is structured as follows : section 2 is related work, section 3 is proposed method and section 4 is the experimental results and finally, the conclusion in section 5.

2 RELATED WORK

This section focuses on the extract features, models, and multimodal fusion used in Audio/Visual Emotion Challenge (AVEC), which is the predecessor of MuSe.

2.1 Extract features

Multi-modal data is helpful to sentiment analysis as it provides more information. MuSe's predecessor, AVEC has been investigating various multimodal features. Before the deep learning area, handcrafted features were used for regression. Sánchez-Lozano et al. [29] have extracted the Local Binary Patterns (LBP) and Gabor from the visual modality and low-level descriptors (LLD) to the audio modality. As deep learning developed, various excellent models could be developed. Shizhe Chen et al. [5] used different hand-crafted and deep learned features from acoustic, visual, and textual modalities in AVEC 2017. Zhao et al. [35] introduced the learned audio features using the pretrained VGGish model in AVEC 2018. Chen et al. [4] combined a pre-trained 2D-CNN and a 1D-CNN for learning deep spatial-temporal features from video and audio in AVEC 2019.

2.2 Models

Support Vector Machine Regression (SVR) was often used in the past [25]. However, other methods have been devised because they cannot deal with temporal and contextual features on emotion recognition or sentiment analysis. It is the Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). RNN and LSTM consider the previous state because the connection between units has a cyclic structure. In recent AVEC, 1st winners have been used the LSTM for continuous sentiment analysis. Also, Sánchez-Lozano et al. [5] used LSTM and SVR to overcome the overfitting problem in AVEC 2017. Sun et al.[32], the winners of MuSe2020, used the LSTM recurrent neural network, adding the self-attention mechanism. Convolutional Neural Network (CNN) is a great model for extracting features from tasks. Xingchen Ma et al. [21] used a combination of CNN and LSTM to use more efficient audio representations.

2.3 Multimodal Fusion

Because emotions are recognized through multiple modalities, it is important to use multi-modal fusion. There are three ways to combine the multiple modalities: early fusion, model fusion and late fusion. Early fusion is the insertion of all features into a single model at once, and late fusion is the combination of the predictions of multiple models from each combination of features. Huang et al. [15] showed that late fusion is better than early fusion for predicting arousal/valence in AVEC 2018. Chen et al. [4] proposed a method to utilize both early fusion and late fusion using Bidirectional Deep Long Short-Term Memory networks (BDLSTM) in AVEC 2019.

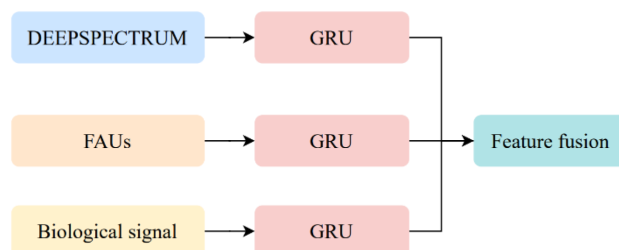

Figure 1: Feature late fusion for Valence.

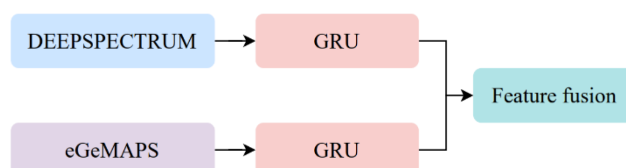

Figure 2: Feature late fusion for Arousal.

3 PROPOSED METHOD

In this section, we present the proposed method for level continues emotional stress estimation. Our method proposed based on two steps: Feature fusion and estimation the value of the stress. For step 1, the DEEPSPECTRUM and eGeMAPS features were combined by feature late fusion strategy to create new features for arousal prediction, illustrated in Figure 1. Similarly, new features of the valence were combined by the DEEPSPECTRUM, FAUS and physiological signal features, illustrated in Figure 2. For step 2, Our method conducted a combination of ingredients, such as Local Attention, GRU, and Bayesian Network (Figure 3) for the stress prediction.

3.1 Gated Recurrent Unit

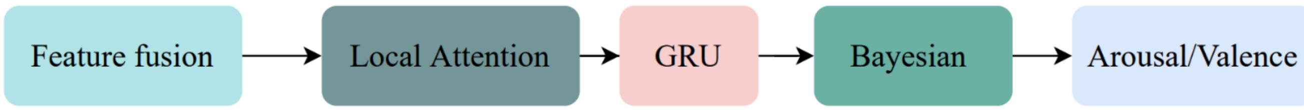

Figure 3: An overview of valence/arousal prediction system.

In 2014, Cho et al. [6] introduced the Gated Recurrent Unit (GRU) to solve the problem about vanishing gradient, which appears from a standard Recurrent Neural Network (RNN). The vanishing gradient problem of a standard RNN was solved by GRU, which was used to update gate and reset gate. These two gates decide on the information to be transmitted to the output. The training process will help keep and remember information over time and remove unnecessary information. Because of the simpler construction, GRUs are quicker to train.

3.2 Bayesian Neural Network

The main idea of a Bayesian Neural Network [3, 10] is to use weights from a normal distribution (Equation 2, 3, 4) instead of using a defined weight (Equation 1). So that the parameters of the network are decided by calculating mean and variance. We have the formula as below:

$$a^{(i+1)} = W^{(i+1)}.z^{(i)} + b^{(i+1)} \tag{1}$$

$$W_{(n)}^{(i)} = N(0,1) * \log\left(1 + \rho^{(i)}\right) + \mu^{(i)} \tag{2}$$

$$b_{(n)}^{(i)} = N(0,1) * \log\left(1 + \rho^{(i)}\right) + \mu^{(i)} \tag{3}$$

$$a_{(n)}^{(i)} = W_{(n)}^{(i)} \text{ x } z_{(n)}^{(i)} + b_{(n)}^{(i)} \tag{4}$$

Where a is the output, b is the biases, W is the wrights, z is the activate function, ρ is standard deviation and μ is the mean on the nth sample of the layer ith.

Bayesian learns about probability distributions of a neural network on the weights, which are used in regression problems to gather confidence interval for data.

3.3 Local Attention

Local attention layer [2, 24, 26] used the window attention to slide on processing. The size of the window is fixed and overlap. This process helps to learn local information and combine information in overlapping.

4 EXPERIMENTS

4.1 Dataset

The MuSe Stress provides by MuSe challenge 2022 [7], which predicts valence and arousal on the continuous levels. The dataset is based on the Ulm-Trier Social Stress Test dataset (ULM-TSST). It comprises 69 participants (20 males and 49 females) and includes about 6 hours (cf. Table 1). The dataset includes set features such as the audio, video, text and physiological signals (ECG, RESP and BPM). Moreover, the MuSe-Stress in this year uses the labels of the previous years of the MuSe-Physio task [31] as the arousal gold standard.

Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) [11] was extracted from the OPENSILE toolkit [12], the acoustic features that can catch affective changes in voice.

DEEPSPECTRUM [1] was utilised pre-trained Convolutional Neural Networks. The MuSe-Stress used DENSENET121 [14] pre-trained on ImageNet [28]. The DEEPSPECTRUM is effective for sentiment analysis.

With Facial Action Units (FAUs) was related with expression emotion which was proposed by Ekman et al. [9]. 20 different FAUs were extracted by using PY-FEAT.

Table 1: Details of partitions in MuSe-Stress dataset.

Partition	ULM-TSST	
	Number of videos	Stress duration
Train	41	3:25:56
Devel	14	1:10:50
Test	14	1:10:41
Total	69	5:47:27

4.2 Experimental setup

The Pytorch Deep Learning toolkit was used in stress task. In the training process, we used a window length of 200 steps and a hop length of 100. The GRU applied 2 layers and 64-dimensional hidden states. Our methods are trained in 100 epochs with an initial learning rate of 0.002 and Adam optimizer [19]. We repeated the training with 20 seeds to get the best CCC score. The reduce learning rate was used for training all stages.

In sub-challenge, we used the Concordance Correlation Coefficient (CCC) loss function the same as the MuSe2022 challenge [7]. The CCC loss was defined:

$$\mathcal{L} = 1 - CCC$$

$$CCC = \frac{2\rho\sigma_{\hat{Y}}\sigma_Y}{\sigma_{\hat{Y}}^2 + \sigma_Y^2 + (\mu_{\hat{Y}} - \mu_Y)^2} \tag{1}$$

where $\boldsymbol{\mu_{\hat{Y}}}$ was the mean of the predict $\hat{\boldsymbol{Y}}$, $\boldsymbol{\mu_Y}$ was the mean of the label $\boldsymbol{Y}$, $\boldsymbol{\rho}$ was the Pearson Correlation Coefficient between $\hat{\boldsymbol{Y}}$ and $\boldsymbol{Y}$, $\boldsymbol{\sigma_{\hat{Y}}}$ and $\boldsymbol{\sigma_Y}$ were the corresponding standard deviations.

4.3 Results

Table 2 presents the results of step 1 of our method. We used only the GRU model for each feature to prediction. We changed to window size (200) and hop size (100). The result demonstrates the audio feature gives better than video and physiological signal. Specifically, DEEPSPECTRUM and eGeMaps got 0.540, 0.541

higher than baseline respectively 0.494, 0.460. Moreover, DEEPSPECTRUM gave the best results with unimodal.

With feature fusion, we conducted combine DEEPSPECTRUM and eGeMAPS for arousal prediction; DEEPSPECTRUM, FAUs and Physiological Signal for valence prediction. We have experiments based on the combination of 3 components: Local Attention, GRU, and Bayesian Network. The results are presented in Table 3. The arousal achieved the best results when combining GRU and Bayesian Network, 0.641, respectively. The valence achieved the best results of the baseline, 0.691, respectively. The combination of valence/arousal achieves the best results when combining Local Attention, GRU, and Bayesian (LAGB) or Local Attention and GRU (LAG), 0.640, respectively. The LAGB and the LAG are better than baseline, 0.640 and 0.618, respectively. The results of LAGB and LAG are equal, however, the valence of the LAGB is better than the LAG. Although it is only a small change, there is potential to improve the prediction.

Table 2: Feature late fusion: Predicted results of Valence, Arousal of MuSe-Stress, Combined ($0.5Arousal + 0.5Valence$), As feature sets, we used DEEPSPECTRUM, eGeMAPS for audio (A); FAUs for video (V); and Physiological Signal with development set.

Features	Proposed method			Baseline [7]		
	Valence	Arousal	Combined	Valence	Arousal	Combined
eGeMAPS	0.554	0.528	0.541	0.411	0.509	0.460
DEEPSPECTRUM	0.536	0.556	**0.546**	0.414	0.574	0.494
FAUs	0.422	0.411	0.416	0.519	0.475	0.497
Physiological signal	0.331	0.252	0.292	0.392	0.436	0.412

Table 3: Stress estimation: Predicted results of Valence, Arousal of MuSe-Stress, Combined ($0.5Arousal + 0.5Valence$), As feature late fusion, we used DEEPSPECTRUM, eGeMAPS for audio ; FAUs for video; and Physiological Signal with development set, where A is Audio, V is Video, T is text, and P is Physiological signal features.

Methods	Features late fusion	Arousal	Valence	Combined
The baseline [7]	A + T	0.448	0.524	0.486
The baseline [7]	T + V	0.461	0.330	0.396
The baseline [7]	A + T + V	0.506	0.441	0.474
The baseline [7]	A + V	0.544	**0.691**	0.618
GRU	A +V +P	0.639	0.625	0.632
GRU + Bayesian	A +V +P	**0.641**	0.631	0.636
Local Attention + GRU	A +V +P	0.628	0.655	**0.640**
Local Attention + GRU + Bayesian	A +V +P	0.614	0.667	**0.640**

5 CONCLUSIONS

In summary, this paper argued that we conducted experimental to the MuSe-Stress 2022. We used features from deep representations of common modalities (i.e., audio, video, and physiological signal). GRU applied to keep information over time for sentiment analysis. Feature fusion strategy improves performance to predict valence/ arousal. To enhance of the model, we combined GRU, Local Attention and Bayesian Network as model prediction. Experimental results show that our proposed model outperforms the baseline method. We demonstrated that our results were better when the multimodal features were combined since each feature contributed differently to the sentiment analysis process.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1A4A1019191) and also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A3A04036408).

The corresponding author is Soo-Hyung Kim.

REFERENCES

[1] Shahin Amiriparian, Maurice Gerczuk, Sandra Ottl, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, and Björn Schuller. 2017. Snore sound classification using image-based deep spectrum features. (2017).

[2] Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The

LongDocument Transformer. arXiv:2004.05150 [cs.CL]
[3] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural network. In International conference on machine learning. PMLR, 1613–1622.
[4] Haifeng Chen, Yifan Deng, Shiwen Cheng, Yixuan Wang, Dongmei Jiang, and Hichem Sahli. 2019. Efficient spatial temporal convolutional features for audiovisual continuous affect recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop. 19–26
[5] Shizhe Chen, Qin Jin, Jinming Zhao, and Shuai Wang. 2017. Multimodal multitask learning for dimensional and continuous emotion recognition. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. 19–26.
[6] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).
[7] Lukas Christ, Shahin Amiriparian, Alice Baird, Panagiotis Tzirakis, Alexander Kathan, Niklas Müller, Lukas Stappen, Eva-Maria Meßner, Andreas König, Alan Cowen, Erik Cambria, and Björn W. Schuller. 2022. The MuSe 2022 Multimodal Sentiment Analysis Challenge: Humor, Emotional Reactions, and Stress. In Proceedings of the 3rd Multimodal Sentiment Analysis Challenge. Association for Computing Machinery, Lisbon, Portugal. Workshop held at ACM Multimedia 2022, to appear
[8] Alia J Crum, Peter Salovey, and Shawn Achor. 2013. Rethinking stress: the role of mindsets in determining the stress response. Journal of personality and social psychology 104, 4 (2013), 716.
[9] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. Environmental Psychology & Nonverbal Behavior (1978).
[10] Piero Esposito. 2020. BLiTZ - Bayesian Layers in Torch Zoo (a Bayesian Deep Learing library for Torch). https://github.com/piEsposito/blitz-bayesian-deeplearning/.
[11] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE transactions on affective computing 7, 2 (2015), 190–202.
[12] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia. 1459–1462.
[13] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. 2019. Review on psychological stress detection using biosignals. IEEE Transactions on Affective Computing 13, 1 (2019), 440–460.
[14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 4700–4708.
[15] Jian Huang, Ya Li, Jianhua Tao, Zheng Lian, Mingyue Niu, and Minghao Yang. 2018. Multimodal continuous emotion recognition with data augmentation using recurrent neural networks. In Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop. 57–64.
[16] Gillian H Ice and Gary D James. 2007. Measuring stress in humans: A practical guide for the field. Cambridge university press.
[17] Peng Jiang, Chunxia Zhang, Hongping Fu, Zhendong Niu, and Qing Yang. 2010. An approach based on tree kernels for opinion mining of online product reviews. In 2010 IEEE International Conference on Data Mining. IEEE, 256–265.
[18] Eugenijus Kaniusas. 2012. Fundamentals of biosignals. In Biomedical signals and sensors I. Springer, 1–26.
[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
[20] Sheng-Tun Li and Fu-Ching Tsai. 2013. A fuzzy conceptualization model for text mining with application in opinion polarity classification. Knowledge-Based Systems 39 (2013), 23–33.
[21] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. 2016. Depaudionet: An efficient deep model for audio based depression classification. In Proceedings of the 6th international workshop on audio/visual emotion challenge. 35–42.
[22] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal 5, 4 (2014), 1093–1113.
[23] Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Emotion measurement. Elsevier, 201–237.
[24] Jack Rae and Ali Razavi. 2020. Do Transformers Need Deep Long-Range Memory?. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online. https://www. aclweb.org/anthology/2020.acl-main.672.
[25] Fabien Ringeval, Björn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, Stefan Scherer, Sharon Mozgai, Nicholas Cummins, Maximilian Schmitt, and Maja Pantic. 2017. Avec 2017: Real-life depression, and affect recognition workshop and challenge. In Proceedings of the 7th annual workshop on audio/visual emotion challenge. 3–9.
[26] Aurko Roy*, Mohammad Taghi Saffar*, David Grangier, and Ashish Vaswani. 2020. Efficient Content-Based Sparse Attention with Routing Transformers. https://arxiv.org/pdf/2003.05997.pdf
[27] Huaxia Rui, Yizao Liu, and Andrew Whinston. 2013. Whose and what chatter matters? The effect of tweets on movie sales. Decision support systems 55, 4 (2013), 863–870.
[28] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. International journal of computer vision 115, 3 (2015), 211–252.
[29] Enrique Sánchez-Lozano, Paula Lopez-Otero, Laura Docio-Fernandez, Enrique Argones-Rúa, and José Luis Alba-Castro. 2013. Audiovisual three-level fusion for continuous estimation of russell's emotion circumplex. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge. 31–40.
[30] Suzanne C Segerstrom and Gregory E Miller. 2004. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. Psychological bulletin 130, 4 (2004), 601.
[31] Lukas Stappen, Alice Baird, Lukas Christ, Lea Schumann, Benjamin Sertolli, Eva-Maria Messner, Erik Cambria, Guoying Zhao, and Björn W Schuller. 2021. The MuSe 2021 multimodal sentiment analysis challenge: sentiment, emotion, physiological-emotion, and stress. In Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge. 5–14.
[32] Licai Sun, Zheng Lian, Jianhua Tao, Bin Liu, and Mingyue Niu. 2020. Multi-modal continuous dimensional emotion recognition using recurrent neural network and self-attention mechanism. In Proceedings of the 1st International on Multimodal Sentiment Analysis in Real-life Media Challenge and Workshop. 27–34.
[33] Tatiana Von Landesberger, Sebastian Bremm, Matthias Kirschner, Stefan Wesarg, and Arjan Kuijper. 2013. Visual analytics for model-based medical image segmentation: Opportunities and challenges. Expert Systems with Applications 40, 12 (2013), 4934–4943.
[34] Abdallah Yousif, Zhendong Niu, John K Tarus, and Arshad Ahmad. 2019. A survey on sentiment analysis of scientific citations. Artificial Intelligence Review 52, 3 (2019), 1805–1838.
[35] Jinming Zhao, Ruichen Li, Shizhe Chen, and Qin Jin. 2018. Multi-modal multicultural dimensional continues emotion recognition in dyadic interactions. In Proceedings of the 2018 on audio/visual emotion challenge and workshop. 65–72.