MuSe 2022 Challenge: Multimodal Humour, Emotional Reactions, and Stress

Shahin Amiriparian
University of Augsburg
Augsburg, Germany

Lukas Christ
University of Augsburg
Augsburg, Germany

Andreas König
University of Passau
Passau, Germany

Eva-Maria Meßner
University of Ulm
Ulm, Germany

Alan Cowen
Hume AI
New York, USA

Erik Cambria
Nanyang Technological University
Singapore

Björn W. Schuller
Imperial College London
London, UK

ABSTRACT

The 3rd **Mu**ltimodal **Se**ntiment Analysis Challenge (MuSe) focuses on multimodal affective computing. The workshop is held in conjunction with ACM Multimedia'22. Three datasets are provided as part of the challenge: (i) the Passau Spontaneous Football Coach Humor (Passau-SFCH) dataset which contains humour-tagged audio-visual data of German football coaches, (ii) the Hume-Reaction dataset, which contains annotations on how people respond to emotional stimuli in terms of seven different emotional expression intensities, and (iii) the Ulm-Trier Social Stress Test (Ulm-TSST) dataset, which consists of audio-visual recordings labelled with continuous emotion values of individuals in stressful circumstances. Based on these datasets three affective computing challenges are defined: 1) Humor Detection Sub-Challenge (MuSe-Humor), for spontaneous humour recognition, 2) Emotional Reactions Sub-Challenge (MuSe-Reaction), for prediction of seven fine-grained 'in-the-wild' emotions, and 3) Emotional Stress Sub-Challenge (MuSe-Stress), for continuous prediction of stressed emotion values. In this summary, we describe the motivation behind the challenge, participation and its conditions, as well as the outcomes. The complete MuSe'22 workshop proceedings are available at: https://dl.acm.org/doi/proceedings/10.1145/3551876.

CCS CONCEPTS

• **Information systems** → **Multimedia and multimodal retrieval**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Multimodal Sentiment Analysis; Affective Computing; Humor Detection; Emotion Recognition; Multimodal Fusion; Challenge; Summary Paper

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10.
https://doi.org/10.1145/3503161.3551792

ACM Reference Format:
Shahin Amiriparian, Lukas Christ, Andreas König, Eva-Maria Meßner, Alan Cowen, Erik Cambria, and Björn W. Schuller. 2022. MuSe 2022 Challenge: Multimodal Humour, Emotional Reactions, and Stress. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3503161.3551792

1 INTRODUCTION

The 3rd edition of the MuSe has been organised in conjunction with the 30th ACM International Conference on Multimedia, held in Lisbon, Portugal, 10 – 14 October 2022. The MuSe 2022 addressed the tasks of **Humour**, **Emotional Reactions**, and **Stress**. It aimed at comparing machine learning and multimedia processing methods for affective computing from audio-visual recordings and transcriptions obtained from the audio modality. The first and second editions of the MuSe [12, 13] have effectively forged ties between the text-based sentiment analysis community [5] and audio-visual affective computing community [1, 8, 11, 12].

The main objective of this year's challenge was to offer a standard benchmark test set for those who process multimodal data and to bring together the audio-visual and Natural Language Processing (NLP) communities working on affective computing problems. Further, MuSe 2022 aimed to compare the relative merits of approaches to humour, continuous emotion recognition and stress under clearly specified and strictly comparable conditions, and to determine the complementarity of different modalities and approaches when they are fused together [7]. The need to improve the recognition of spontaneous, 'in-the-wild' affective data and the preparation of such machine learning systems for the real-world application [2–4] was a further driving factor for MuSe 2022.

This year, we called for participation in three sub-challenges. In the **MuSe-Humor** sub-challenge, participants detected the presence of humour in football press conference recordings. The dataset for this sub-challenge contains recordings of press conferences of ten German Bundesliga football coaches. The participants should predict the presence of spontaneous humour.

For the **MuSe-Reaction** sub-challenge, emotional reactions are investigated by introducing a large-scale, multimodal (audio and video) dataset containing 70+ hours of in-the-wild recordings from

2,222 subjects. Participants of this challenge trained multi-output regression models to predict the intensities of seven self-reported emotions: *Adoration*, *Amusement*, *Anxiety*, *Disgust*, *Empathic Pain*, *Fear*, and *Surprise*.

In **MuSe-Stress**, participants developed regression models to continuously predict emotional values in terms of arousal and valence. This sub-challenge is based on the Ulm-TSST dataset, which includes data from individuals in a stress-inducing scenario following the Trier Social Stress Test (TSST) [9]. The MuSe-Stress sub-challenge is inspired by the ubiquity of stress and its negative consequences in modern societies [6]. Besides the audio-visual and text features, Ulm-TSST comprises three biological signals Electrocardiogram (ECG), heart rate (BPM), and Respiration (RESP) which are captured at a sampling rate of 1 kHz.

In addition to participation in the challenge, we encouraged contributions that address multimodal representation learning for emotion recognition, text-based sentiment analysis, transfer and self-supervised learning for audio-visual affect sensing, and mobile and low-resource machine learning applications for real-time emotion recognition.

2 CHALLENGE PROCEDURE

Each team needed to register for the challenge via the official challenge's website[1]. Subsequently, in order to gain access to the data, a team leader who holds an academic position had to sign an end-user licence agreement (EULA) for the intended sub-challenges. For MuSe-Humor and MuSe-Stress, registered teams were given a link to the research data site zenodo.org after examining the EULA. Using the provided link, the participants could download the challenge packages, including raw multimodal data, metadata, feature sets, and ground truth labels. Regarding MuSe-Reaction, participants obtained the data directly from Hume AI after filling the respective EULA. For the test partition, no ground truth labels were provided. For the sake of reproducibility of the baseline results, the baseline code was made publicly available on GitHub[2]. The repository contains information on the installation of the baseline model (a Long Short-Term Memory Recurrent Neural Network regressor), experimental settings and details on reproducing the unimodal and fusion results.

Evaluation of the teams' predictions on the confidential test set was carried out via the Codalab[3] platform. Registered participants received the respective links and could then upload their predictions, immediately being informed about their results. For both MuSe-Humor and MuSe-Reaction, at most five submissions were allowed, while in MuSe-Stress up to 20 submissions were possible. Invalid uploads due to malformed submission files did not count towards the maximum number of allowed predictions. The leaderboard was anonymised.

3 PARTICIPATION

Overall, 41 teams from 16 nations and 31 academic institutions registered in response to the call for participation. About 40 % of the teams provided their predictions on the test partition, resulting in a total of 243 test set submissions. Participants were encouraged to describe their developed approaches and highlight their contribution to the challenge in a paper of up to 8 pages (plus 1-2 pages for references). In total, 17 valid paper submissions were received. The review process was double-blind, and each paper was reviewed by at least three members of the programme committee w. r. t. its novelty, technical correctness, and clarity of presentation. Key statistics for each sub-challenge in terms of the number of registrations and number of teams who actually submitted test predictions are provided in Table 1.

Table 1: Statistics on the number of registered teams (Team Reg.) and teams that have submitted results on the test partition (Test Sub.). Further, baseline results and the results of each challenge's winner and their approaches are provided. ρ: Pearson's Correlation Coefficient; *CCC*: Concordance Correlation Coefficient; *AUC*: Area under the ROC Curve; *TX*: Transformers; *ATT*: Attention mechanism.

	MuSe-Humor	MuSe-Reaction	MuSe-Stress
Team Reg. (#, %)	20, 48.76 %	12, 29.27 %	32, 78.05 %
Test Sub. (#, %)	7, 35 %	6, 50 %	12, 37.5 %
Baseline [7]	.8480 AUC	.2801 ρ	.4761 CCC
Challenge's Winner	.9065 AUC	.3879 ρ	.6829 CCC
Winner's approach	CNNs, TX	CNNs, TX	CNNs, TX, ATT

4 CHALLENGE OUTCOME

All competitive baseline results could be exceeded. Participants often used the provided baseline features to train their machine learning models. The majority of teams, however, extracted additional feature sets and some teams have applied data augmentation techniques. Popular additional features were obtained by applying pre-trained Transformers to all modalities.

In MuSe-Stress, some teams experimented with additional features based on the physiological signals (e. g., NeuroKit2 [10]). The vast majority of teams employed a range of Transformer architectures, typically combined with more "conventional" neural network components such as convolutional and recurrent neural networks. More exotic model types, including Bayesian neural networks and Graph Convolutional Networks, were also employed. In MuSe-Stress, many teams have experimented with the physiological signals and 2 submissions focused primarily on them. Multiple papers made use of fusion strategies, from sophisticated early fusion approaches (by using Transformers) to simpler decision-level fusions. An interesting approach was provided by one of the teams which obtained another modality from the video data of MuSe-Reaction by applying a video captioning approach. Baselines and the results of the winner teams are provided in the Table 1.

5 WORKSHOP ORGANISATION

MuSe takes place as a full-day workshop in Lisbon, Portugal. The workshop organisers deliver brief opening and closing remarks for the program, which also includes oral presentations of the approved papers and two invited keynote speeches. We appreciate the reviewers' efforts and are grateful to the data chairs Alexander Kathan (University of Augsburg, GER), Alice Baird (Hume AI, USA), and

[1] https://www.muse-challenge.org/
[2] https://github.com/EIHW/MuSe2022
[3] https://codalab.lisn.upsaclay.fr/

Panagiotis Tziraks (Hume AI, USA) as well as the program committee for their valuable support: Joost Broekens (University of Leiden, NL), Paul Buitelaar (NUI Galway, IRL), Felix Burkhardt (TU Berlin, DE), Shaun Canavan (University of South Florida, USA), Guillaume Chanel (University of Geneva, CH), Akshi Kumar (Manchester Metropolitan University, UK), Mani Kumar (University of Nottingham, UK), Siddique Latif (University of Southern Queensland, AUS), Gelareh Mohammadi (UNSW Sydney, AUS), Symeon Papadopoulos (ITI , GR), Ioannis Patras (Queen Mary University London, UK), Fabien Ringeval (Universite Grenoble Alpes, FRA), Peter Robinson (University of Cambridge, UK), Albert Ali Salah (Utrecht University, NL), John See (Heriot-Watt University Malaysia, MYS), Mohammad Soleymani (USC Institute for Creative Technologies, USA), Matthew Turk (Toyota Technological Institute at Chicago, USA), Rui Xia (Nanjing University of Science and Technology, CN), and Ziping Zhao (Tianjin Normal University, CN).

REFERENCES

[1] Shahin Amiriparian. 2019. *Deep Representation Learning Techniques for Audio Signal Processing*. Ph.D. Dissertation. Technische Universität München.

[2] Shahin Amiriparian, Nicholas Cummins, Sandra Ottl, Maurice Gerczuk, and Björn Schuller. 2017. Sentiment Analysis Using Image-based Deep Spectrum Features. In *Proceedings 2nd International Workshop on Automatic Sentiment Analysis in the Wild (WASA 2017) held in conjunction with the 7th biannual Conference on Affective Computing and Intelligent Interaction (ACII 2017)*. AAAC, IEEE, San Antonio, TX, 26–29.

[3] Shahin Amiriparian, Maurice Gerczuk, Lukas Stappen, Alice Baird, Lukas Koebe, Sandra Ottl, and Björn Schuller. 2020. Towards Cross-Modal Pre-Training and Learning Tempo-Spatial Characteristics for Audio Recognition with Convolutional and Recurrent Neural Networks. *EURASIP Journal on Audio, Speech, and Music Processing* 2020, 19 (2020), 1–11.

[4] Shahin Amiriparian, Tobias Hübner, Vincent Karas, Maurice Gerczuk, Sandra Ottl, and Björn W. Schuller. 2022. DeepSpectrumLite: A Power-Efficient Transfer Learning Framework for Embedded Speech and Audio Processing From Decentralized Data. *Frontiers in Artificial Intelligence* 5 (2022), 10. https://doi.org/10.3389/frai.2022.856232

[5] Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. 2017. Affective computing and sentiment analysis. In *A practical guide to sentiment analysis*. Springer, 1–10.

[6] Yekta Said Can, Bert Arnrich, and Cem Ersoy. 2019. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics* 92 (2019), 22.

[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] Maurice Gerczuk, Shahin Amiriparian, Sandra Ottl, and Björn Schuller. 2022. EmoNet: A Transfer Learning Framework for Multi-Corpus Speech Emotion Recognition. *IEEE Transactions on Affective Computing* 13 (2022).

[9] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.

[10] Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods* 53, 4 (2021), 1689–1696.

[11] Björn W. Schuller, Anton Batliner, Shahin Amiriparian, Christian Bergler, Maurice Gerczuk, Natalie Holz, Sebastian Bayerl, Korbinian Riedhammer, Adria Mallol-Ragolta, Maria Pateraki, Harry Coppock, Ivan Kiskin, and Stephen Roberts. 2022. The ACM Multimedia 2022 Computational Paralinguistics Challenge: Vocalisations, Stuttering, Activity, & Mosquitos. In *Proceedings of the 30th ACM international conference on Multimedia, ACM MM 2022*. ACM, ACM, Lisbon, Portugal. to appear.

[12] 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*. Association for Computing Machinery, New York, NY, USA, 5–14.

[13] Lukas Stappen, Alice Baird, Georgios Rizos, Panagiotis Tzirakis, Xinchen Du, Felix Hafner, Lea Schumann, Adria Mallol-Ragolta, Bjoern W. Schuller, Iulia Lefter, Erik Cambria, and Ioannis Kompatsiaris. 2020. MuSe 2020 Challenge and Workshop: Multimodal Sentiment Analysis, Emotion-Target Engagement and Trustworthiness Detection in Real-Life Media. In *Proceedings of the 1st International on Multimodal Sentiment Analysis in Real-Life Media Challenge and Workshop*. ACM, Association for Computing Machinery, New York, NY, USA, 35–44.