MuSe 2023 Challenge: Multimodal Prediction of Mimicked Emotions, Cross-Cultural Humour, and Personalised Recognition of Affects

Shahin Amiriparian
University of Augsburg
Augsburg, Germany

Lukas Christ
University of Augsburg
Augsburg, Germany

Andreas König
University of Passau
Passau, Germany

Alan Cowen
Hume AI
New York, USA

Eva-Maria Meßner
University of Ulm
Ulm, Germany

Erik Cambria
Nanyang Technological University
Singapore

Björn W. Schuller
Imperial College London
London, UK

ABSTRACT

The 4th **Mu**ltimodal **Se**ntiment Analysis Challenge (MuSe) focuses on Multimodal Prediction of Mimicked Emotions, Cross-Cultural Humour, and Personalised Recognition of Affects. The workshop takes place in conjunction with ACM Multimedia'23. We provide three datasets as part of the challenge: (i) The Hume-Vidmimic dataset which offers 30+ hours of expressive behaviour data from 557 participants. It involves mimicking and rating emotions: *Approval*, *Disappointment*, and *Uncertainty*. This multimodal resource is valuable for studying human emotional expressions. (ii) The 2023 edition of the Passau Spontaneous Football Coach Humor (Passau-SFCH) dataset comprises German football press conference recordings within the training set, while videos of English football press conferences are included in the unseen test set. This unique configuration offers a cross-cultural evaluation environment for humour recognition. (iii) The Ulm-Trier Social Stress Test (Ulm-TSST) dataset contains recordings of subjects under stress. It involves arousal and valence signals, with some test labels provided to aid personalisation. Based on these datasets, we formulate three multimodal affective computing challenges: (1) Mimicked Emotions Sub-Challenge (MuSe-Mimic) for categorical emotion prediction, (2) Cross-Cultural Humour Detection Sub-Challenge (MuSe-Humour) for cross-cultural humour detection, and (3) Personalisation Sub-Challenge (MuSe-Personalisation) for personalised dimensional emotion recognition. In this summary, we outline the challenge's motivation, participation guidelines, conditions, and results.

CCS CONCEPTS

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

KEYWORDS

Multimodal Sentiment Analysis; Affective Computing; Emotion Mimics, Cross-Cultural Humour Detection; Emotion Recognition; Multimodal Fusion; Challenge; Summary Paper

ACM Reference Format:
Shahin Amiriparian, Lukas Christ, Andreas König, Alan Cowen, Eva-Maria Meßner, Erik Cambria, and Björn W. Schuller. 2023. MuSe 2023 Challenge: Multimodal Prediction of Mimicked Emotions, Cross-Cultural Humour, and Personalised Recognition of Affects. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3581783.3610943

1 INTRODUCTION

The 4th edition of the MuSe took place in conjunction with the 31st ACM International Conference on Multimedia, held in Ottawa, Canada, from October 29, 2023, to November 3, 2023. The objective was to assess machine learning and multimedia processing techniques within the realm of paralinguistics and affective computing, while also encouraging researchers in these fields to apply their latest approaches to our specific sub-challenges. This evaluation encompassed the utilisation of both audio-visual recordings and transcriptions derived from the audio modality. The preceding editions of the MuSe have effectively fostered connections among various research communities. These include the fields of text-based [5] as well as audio-based [2] sentiment analysis, audio-visual affective computing [4, 19, 22], and even strategic management [11, 12].

The primary goal of this year's challenge was twofold: to provide a number of different benchmarks for those engaged in processing multimodal data, and to facilitate collaboration between the audio-visual and Natural Language Processing (NLP) communities focused on contemporary affective computing issues [1]. Given the multifaceted aspects of human affect, MuSe 2023 aimed to evaluate and compare the effectiveness of various multimodal machine learning approaches. This evaluation was conducted within the realms of categorical emotion prediction, cross-cultural humour detection, and personalised dimensional emotion recognition, all executed under precisely comparable conditions. This evaluation sought to

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10.
https://doi.org/10.1145/3581783.3610943

explore how different modalities and methodologies complement each other when integrated [6]. Furthermore, the synthesis of audiovisual and text data within the framework of MuSe 2023 offers a holistic representation of emotional states, leading to more comprehensive insights. Finally, the need to enhance the recognition of spontaneous and real-world affective data often encountered 'in the wild', and to prepare such machine learning systems for practical application was an additional motivating factor behind MuSe 2023 [3, 6].

This year, we invited participation in three sub-challenges. In the **MuSe-Mimic** sub-challenge, participants are tasked with conducting a multi-output regression using features derived from audiovisual and textual data to predict the intensity of three distinct emotional targets: *Approval*, *Disappointment*, and *Uncertainty*. The evaluation metric employed is Pearson's correlation coefficient (ρ).

For the **MuSe-Humour** sub-challenge, each team is tasked with working on an extension of the Passau-SFCH dataset [8]. The primary goal here is to identify spontaneous humour within press conferences, and this challenge brings in a cross-cultural element. Specifically, participants are required to train their models on German recordings and subsequently deploy these models to predict instances of humour within English data.

Within the **MuSe-Personalisation** sub-challenge which employed the Ulm-TSST dataset, teams are prompted to utilise personalisation techniques to predict continuous estimations of valence and arousal. Different from recent years' standard speaker-independent setup [7], participants received labelled data for each subject from the test partition. This approach encourages the exploration of adapting multimodal emotion recognition systems to individuals, accounting for their distinctive characteristics

2 CHALLENGE PROCEDURE

Teams were required to register for the challenge through the official challenge website[1]. Afterwards, access to the data was granted to teams who signed an End-User License Agreement (EULA) for their chosen sub-challenges. For MuSe-Humour and MuSe-Personalisation, registered teams were provided with a link to the research data on zenodo.org[2] [3] after EULA verification. This link enabled participants to download challenge packages, encompassing raw multimodal data, metadata, feature sets, and ground truth labels. In the case of MuSe-Mimic, participants obtained data directly from Hume AI upon EULA acceptance. Except for the MuSe-Personalisation sub-challenge, no ground truth labels were available for the test partition.

To ensure reproducibility, baseline code was publicly accessible on GitHub[4]. The repository provided installation instructions for the baseline model (a GRU-RNN), experimental details, and guidelines to reproduce unimodal and fusion results.

The teams' predictions on the confidential test set were evaluated through the CodaLab[5] platform. For all sub-challenges, teams could make up to five submissions. Invalid submissions due to file format issues did not contribute to the maximum submission count. The leaderboard maintained anonymity throughout the process.

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; *TF*: Transformers; *ATT*: Attention mechanism.

	MuSe-Mimic	MuSe-Humour	MuSe-Personalisation
Team Reg. (#, %)	15, 46.9 %	10, 31.3 %	61, 65.6 %
Test Sub. (#, %)	9, 60.0 %	9, 90.0 %	11, 52.4 %
Baseline [7]	.4727 ρ	.8310 AUC	.7639 CCC
Challenge's Winner	.7351 ρ	.8889 AUC	.8681 CCC
Winner's team	*xmly*	*IAI-CNSC*	*IAI-CNSC*
Winner's approach	LLMs, LoRA, RNN, TF	CNN, ATT	GRU-RNNs

3 PARTICIPATION

In response to the call for participation, a total of 31 teams from 8 nations and 25 academic institutions registered, resulting in 108 test set prediction submissions overall. Participants were invited to detail their developed methodologies and emphasise their contributions to the challenge in an up to 8-page paper (with an additional 1-2 pages for references). In total, 16 papers were submitted. Conducted in a double-blind manner, each paper underwent assessment by a minimum of two members from the program committee, evaluating its degree of innovation, technical accuracy, and presentation clarity. Comprehensive statistics for each sub-challenge, encompassing registration figures and the number of teams that ultimately submitted test predictions, are outlined in Table 1.

4 CHALLENGE OUTCOME

For all sub-challenges, the baseline results were outperformed. Table 1 lists the baseline results as well as the winner's results and methods. The baseline model consisting of unimodal Gated Recurrent Unit (GRU)-Recurrent Neural Networks (RNNs) and late fusion of them was frequently reused and adapted by participants (e. g., [13, 16, 28]). Participants employing the baseline model focused on extracting further features and developing advanced training frameworks on the basis of the provided code. Besides, one can observe the heavy use of Transformers. First, numerous participants employ pretrained Transformer models to compute additional features. Examples include Whisper [20] for audio [28] and APViT [29] for the video modality [13–15]. Second, Transformer- and attention-based architectures proved to be popular choices for fusing representations of different modalities [9, 15, 24, 31]. Most proposed approaches make use of both audiovisual and, if applicable, textual data. Moreover, three papers on MuSe-Personalisation experiment with the Ulm-TSST dataset's [23] physiological signals [13, 24], with one paper [16] focusing entirely on them. The trend towards pretrained Large Language Models (LLMs) with several billions of parameters is also mirrored in this year's methods. Models such as GLM [10] and Bloom [21] in versions of up to 13B parameters are employed by two teams for MuSe-Humour [27]

[1] https://www.muse-challenge.org/

[2] MuSe-Humour: https://zenodo.org/record/7843401

[3] MuSe-Personalisation: https://zenodo.org/record/7920826

[4] https://github.com/EIHW/MuSe-2023

[5] https://codalab.lisn.upsaclay.fr/

and MuSe-Mimic [30], respectively. In MuSe-Humour, some participants explore automatic translation approaches to address the cross-cultural setting of this challenge [26, 27]. Two papers [18, 26] utilise explainable AI methods [17, 25] to provide insights into their models.

5 WORKSHOP ORGANISATION

MuSe 2023 is a full-day workshop in Ottawa, Canada. The program encompasses oral presentations of the accepted papers and a keynote speech. We appreciate the reviewers' efforts and are thankful to the data chairs Alexander Kathan (University of Augsburg, GER) and Alice Baird (Hume AI, USA) and greatly appreciate the reviewers' efforts. We would like to express our gratitude to the program committee for their much appreciated support: Azam Bastanfard (Karaj Islamic Azad University, IRN), Shaun Canavan (University of South Florida, USA), Guillaume Chanel (University of Geneva, CH), Heysem Kaya (Utrecht University, NL), Vangelis Metsis (Texas State University, USA), Peter Robinson (University of Cambridge, ENG), Mohammad Soleymani (University of Southern California, USA), Ziping Zhao (Tianjin Normal University, CHN).

REFERENCES

[1] Shahin Amiriparian. 2022. The Dos and Don'ts of Affect Analysis. In *Proc. of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge.* ACM, Ottawa, Canada, 3–3.

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

[3] 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.

[4] Shahin Amiriparian, Bjorn W Schuller, Nabiha Asghar, Heiga Zen, and Felix Burkhardt. 2023. Guest Editorial: Special Issue on Affective Speech and Language Synthesis, Generation, and Conversion. *IEEE Transactions on Affective Computing* 14, 01 (2023), 3–5.

[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] Lukas Christ, Shahin Amiriparian, Alice Baird, Alexander Kathan, Niklas Müller, Steffen Klug, Chris Gagne, Panagiotis Tzirakis, Lukas Stappen, Eva-Maria Meßner, Andreas König, Alan Cowen, Erik Cambria, and Björn W. Schuller. 2023. The MuSe 2023 Multimodal Sentiment Analysis Challenge: Mimicked Emotions, Cross-Cultural Humour, and Personalisation. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[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 *Proc. of the 3rd Multimodal Sentiment Analysis Challenge.* ACM, Lisbon, Portugal. Workshop held at ACM Multimedia 2022, to appear.

[8] Lukas Christ, Shahin Amiriparian, Alexander Kathan, Niklas Müller, Andreas König, and Björn W Schuller. 2023. Towards Multimodal Prediction of Spontaneous Humour: A Novel Dataset and First Results. *arXiv preprint arXiv:2209.14272* (2023).

[9] Chaoyue Ding, Daoming Zong, Baoxiang Li, Song Zhang, Xiaoxu Zhu, Guiping Zhong, and Dinghao Zhou. 2023. Multimodal Sentiment Analysis via Efficient Multimodal Transformer and Task-Aware Adaptive Training Loss. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[10] Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. In *Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Association for Computational Linguistics, Dublin, Ireland, 320–335. https://doi.org/10.18653/v1/2022.acl-long.26

[11] Alexander Kathan, Shahin Amiriparian, Lukas Christ, Andreas Triantafyllopoulos, Niklas Müller, Andreas König, and Björn W Schuller. 2022. A personalised approach to audiovisual humour recognition and its individual-level fairness. In *Proc. of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge.* 29–36.

[12] Andreas König, Lorenz Graf-Vlachy, Jonathan Bundy, and Laura M Little. 2020. A blessing and a curse: How CEOs' trait empathy affects their management of organizational crises. *Academy of Management Review* 45, 1 (2020), 130–153.

[13] Jia Li, Wei Qian, Kun Li, Qi Li, Dan Guo, and Meng Wang. 2023. Exploiting Diverse Feature for Multimodal Sentiment Analysis. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[14] Qi Li, Shulei Tang, Feixiang Zhang, Ruotong Wang, Yangyang Xu, Zhuoer Zhao, Xiao Sun, and Meng Wang. 2023. Temporal-aware Multimodal Feature Fusion for Sentiment Analysis. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[15] Qi Li, Yangyang Xu, Zhuoer Zhao, Shulei Tang, Feixiang Zhang, Ruotong Wang, Xiao Sun, and Meng Wang. 2023. JTMA: Joint multimodal feature fusion and Temporal Multi-head Attention for Humor Detection. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[16] Misha Libman and Gelareh Mohammadi. 2023. ECG-Coupled Multimodal Approach for Stress Detection. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[17] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30,* I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[18] Ho-Min Park, Ganghyun Kim, Arnout Van Messem, and Wesley De Neve. 2023. MuSe-Personalization 2023: Feature Engineering, Hyperparameter Optimization, and Transformer-Encoder Re-discovery. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[19] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion* 37 (2017), 98–125.

[20] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning.* PMLR, 28492–28518.

[21] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100* (2022).

[22] Björn W. Schuller, Anton Batliner, Shahin Amiriparian, Alexander Barnhill, Maurice Gerczuk, Andreas Triantafyllopoulos, Alice Baird, Panagiotis Tzirakis, Chris Gagne, Alan S. Cowen, Nikola Lackovic, Marie-José Caraty, and Claude Montacié. 2023. The ACM Multimedia 2023 Computational Paralinguistics Challenge: Emotion Share & Requests. In *Proc. of the 31. ACM International Conference on Multimedia, MM 2023.* ACM, ACM, Ottawa, Canada. 5 pages, to appear.

[23] 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 *Proc. of the 2nd on Multimodal Sentiment Analysis Challenge.* ACM, New York, NY, USA, 5–14.

[24] Haiyang Sun, Zhuofan Wen, Mingyu Xu, Zheng Lian, Licai Sun, Bin Liu, and Jianhua Tao. 2023. Exclusive Modeling for MuSe-Personalisation Challenge. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[25] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning.* PMLR, 3319–3328.

[26] Grósz Tamás, Anja Virkkunen, Dejan Porjazovski, and Mikko Kurimo. 2023. Discovering Relevant Sub-spaces of BERT, Wav2Vec 2.0, ELECTRA and ViT Embeddings for Humor and Mimicked Emotion Recognition with Integrated Gradients. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[27] Heng Xie, Jizhou Cui, Yuhang Cao, Junjie Chen, Jianhua Tao, Cunhang Fan, Xuefei Liu, Zhengqi Wen, Heng Lu, Yuguang Yang, Zhao Lv, and Yongwei Li. 2023. Multimodal Cross-Lingual Features and Weight Fusion for Cross-Cultural Humor Detection. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[28] Mingyu Xu, Shun Chen, Zheng Lian, and Bin Liu. 2023. Humor Detection System for MuSE 2023: Contextual Modeling, Pesudo Labelling, and Post-smoothing. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[29] Fanglei Xue, Qiangchang Wang, Zichang Tan, Zhongsong Ma, and Guodong Guo. 2022. Vision transformer with attentive pooling for robust facial expression recognition. *IEEE Transactions on Affective Computing* (2022).

[30] Guofeng Yi, Yuguang Yang, Yu Pan, Yuhang Cao, Jixun Yao, Xiang Lv, Cunhang Fan, Zhao Lv, Jianhua Tao, Shan Liang, and Heng Lu. 2023. Exploring the Power of Cross-Contextual Large Language Model in Mimic Emotion Prediction. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.

[31] Jun Yu, Wangyuan Zhu, Jichao Zhu, Xiaxin Shen, Jianqing Sun, and Jiaen Liang. 2023. MMT-GD: Multi-Modal Transformer with Graph Distillation for Cross-Cultural Humor Detection. In *Proc. of MuSe '23.* ACM, Ottawa, Canada. to appear.