Emotion Recognition in German Sign Language with Facial Action Units

Cristina Luna-Jiménez*, Lennart Eing*, Sergio Esteban Romero†, Tanja Schneeberger‡, Patrick Gebhard‡, Fabrizio Nunnari‡, Elisabeth André*

*University of Augsburg
{cristina.luna.jimenez, lennart.eing, elisabeth.andre}@uni-a.de
†Universidad Politécnica de Madrid
{sergio.estebanro}@upm.es
‡German Research Center for Artificial Intelligence (DFKI)
{tanja.schneeberger, patrick.gebhard, fabrizio.nunnari}@dfki.de

Abstract

Emotion Recognition research in Sign Languages is still in its infancy. Still today, there exists a lack of knowledge about appropriate annotation guidelines and the impact that facial expressions, body postures and head positions have in recognizing emotions while signing, considering that sign language encompasses manual and non-manual cues with linguistic purposes. In this article, we present an acquisition protocol to record acted emotions in German Sign Language under four scenarios (High-Valence and High-Arousal, High-Valence and Low Arousal, Low-Valence and High-Arousal, and Low-Valence and Low-Arousal). The goal is to provide a reference dataset to explore the use of machine learning techniques for an automated classification of emotions in sign language utterances. As a baseline reference, we trained static models with features extracted from the facial muscle activations. The best model achieved an accuracy of 68.84% and a F1 of 67.96% with a random forest trained on the statistics extracted from Action Units. These results highlight the importance of facial expression in sign language, not only for carrying linguistic information but also for transmitting emotions. Results also indicate challenges in detecting emotions in the High-Valence and Low Arousal scenario, which suggests future investigation lines to explore.

Keywords: Emotion Recognition, Sign Language, Machine Learning

1. Introduction

Emotions play an important role in our communication and decision-making. In spoken conversations, emotional content can be conveyed from our facial expressions, or posture, among others (Luna-Jiménez et al., 2022; Zhang et al., 2023). However, Sign Languages are characterized by the use of manual and non-manual channels to transmit, simultaneously, linguistic and affective content. For example, Pendzich et al. (Pendzich et al., 2022) observed that conditionals in German Sign Language (DGS) can be accompanied by brow raise, chin raiser or upper lid raise expressions. Similarly, mouthing can disambiguate semantics of certain signs in DGS (e.g., the same hand movements depicting a square can refer to 'square', 'page' or 'letter' depending on the mouthing (Konrad et al., 2022)). This richness of facial expressions for conveying linguistic meanings represents a challenge for existing automatic emotion recognizers when the person is signing since they were not trained to differentiate that certain facial expressions are associated to linguistic meanings.

Although previous research exists in Affective Computing for Facial Emotion Recognition (FER) in spoken languages (Livingstone and Russo, 2018; Vlasenko et al., 2007), the studies in Sign Languages is rather limited. Therefore, the main contributions of this article are summarized as:

- A proposal of an acquisition framework to elicit and record acted emotions in Sign Language, following previous research in Affective Computing and adapting it to DGS. This framework was evaluated by performing a corpus acquisition with professional deaf interpreters under controlled conditions, signing different glosses in DGS.

- An analysis of the acquired dataset by proposing a baseline with classical Machine Learning models on Facial Emotion Recognition in Sign Language. The top performance was obtained employing Action Units with a Random Forest, achieving an average accuracy of 68.84% and F1 of 67.96%, following a subject-wise leave-one-out strategy. These results highlight that non-manual features, as the facial expressions, not only carry semantic and linguistic information in DGS but also valuable keys for performing emotion recognition.

To our knowledge, this is the first study that explores how to record an acted dataset for performing emotion recognition in German Sign Language (DGS) and evaluate the contribution of facial features for predicting emotions. This study establishes a first exploration in a controlled scenario to, in the future, extend to other affective computing paradigms (e.g. study of emotions in the wild) and

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 297–305
16 May 2026.

acquisition methodologies given the infancy state of the field of emotion recognition in Sign Language.

The remainder of the paper is organized as follows: Section 2 describes the related work, summarizing previous research in emotion recognition in sign language. Section 3 introduces the corpus acquisition process, including the scenario and recording procedure. Then, Section 4 describes the corpus evaluation with machine learning algorithms, indicating the extracted features and the trained models. Afterwards, Section 5 presents the obtained results and the confusion matrices of the predictions per subject. Finally, Section 6 summarizes the main conclusions and future research directions.

2. Related Works

2.1. Emotions in Sign Language

In sign language linguistics, much work investigated how non-manual features serve both affective and grammatical functions.

For example, in a study on Irish sign language (ISL), Kulshreshtha (2024) found that participants used the doubled-wh signs in the context of anger, surprise, and excitement.

Kirst (2024) reports on an experiment aimed at studying the "layering", i.e., "when multiple phonological and paralinguistic elements co-occur". In this experiment, they observed that signers employ three meta-strategies to cope with the communication of emotions on the top of a message: 1) separation, by sequencing the emotion expression before or after the grammatical information; 2) addition, by introducing, for example, rapid blinking and jaw drops; 3) competition, by increasing muscle activation when, for example, both a question modifier and expression of anger require brows furrowing.

Recently, Schwarzenberg et al. (2024) also investigated the interaction between lexemes and non-manuals, showing inconsistency in the use of facial expressions accompanying a sign for an emotional concept (e.g., sadness).

All of the above studies suggest how the use of facial expressions for emotion communication in sign languages differs from spoken communication.

In addition, in sign languages emotion communications is not limited to hands. Hietanen et al. (2004) conducted a meticulous study on the perception of emotions from hands movement in Finnish sign language. The study was focusing on only two emotions (happy, anger) plus neutral. The emotions were judged by non-signers. The results show indeed a significant capability to distinguish between neutral and anger. If emotions from hands can be perceived by non-signers, surely it is part of the language, and it suggests that it is worth investigating in deeper detail the role of body movement for emotion communication.

2.2. Automatic Emotion Recognition

From a computer science viewpoint, research in emotion recognition has advanced in several lines pursuing to study emotions from different perspectives depending on the naturalness and genuineness of the elicited and expressed emotions.

In the first group, datasets on acted emotions are characterized by clean expressions of emotions, in which actors are hired to display them with high arousal (Livingstone and Russo, 2018; Vlasenko et al., 2007). The second group contains elicited emotions, in which stimulus such as images, sounds or conversations are employed to induce a specific emotion in the participants (McKeown et al., 2012), which after being exposed to the stimulus display a more natural emotional reaction. Finally, the third group of datasets consists of 'spontaneous' emotions captured for example from news or TV shows (Grimm et al., 2008), in which natural reactions to events are recorded from the facial and body expressions of the participants. Although this last set is the most natural or genuine, it suffers often from an imbalance in the available classes.

However, although these datasets exist normally for spoken languages, the same pairs are starting to appear and being investigated only recently for sign language. For example, the eJSL dataset (Funakoshi and Zhu, 2025) enters in the category of 'acted' datasets in which two signers represented emotions in 78 pre-selected utterances, resulting in 1,092 videos, recorded in Japanese Sign Language (JSL). This dataset was recorded by a deaf native professional actor, mimicking seven emotions plus neutral. An example of a 'spontaneous' dataset in DGS is FePh (Alaghband et al., 2020), in which emotions for 3,359 image samples from the RWTH-PHOENIX-Weather 2014 (Koller et al., 2015) were annotated based on facial expressions according to the seven 'universal emotions' proposed by P. Ekman (Ekman, 1999) plus a 'None' label in case the emotion was not in the group of seven. Finally, the EmoSign dataset (Chua et al., 2025), in American Sign Language (ASL), is a subset of 200 videos of an average duration of 4.8 seconds annotated by three deaf signer deaf interpreters. This dataset would be considered as part of the 'spontaneous' emotion dataset since videos were extracted from a larger corpus, the ASLLRP (American Sign Language Linguistic Research Project) (Neidle et al., 2022) which was recorded with real interactions of deaf participants and annotated posteriorly. The subset contains labels of sentiments on 7-point Likert scale, and 10 emotions. Additionally, it includes descriptions of manual and non-manual emotional

indicators such as facial expressions, head, mouth and body movements. In addition to the dataset, they provide baselines for the study of emotions in ASL with Multimodal Large Language Models.

Due to the limit number of available datasets with annotations of emotions, researchers interested on performing automatic emotions recognition are starting to employ them. For example, Nunnari et al. (2025) used a subset of 2,547 frames from FePh to recognize emotions in Sign Language videos with a light CNN architecture to perform emotion recognition. Nonetheless, there exists still open-challenges to address, such as the disentanglement of emotional and grammatical functions of facial expressions (Kimmelman et al., 2020; Kirst, 2024), which requires of further development of datasets and annotation schemes. Additionally, the understanding of whether previous approaches and features developed for facial and body emotion recognition (Luna-Jiménez et al., 2022; Martinez-Martin and Fernández-Caballero, 2025) could be also effective for predicting emotions while signing is still a open research question.

3. Corpus Acquisition

In order to create a scenario for collecting and eliciting emotions, we followed the valence-arousal approach introduced by Posner et al. (2005). In this theory, 'valence' describes how pleasant or unpleasant a situation is. For example, situations with negative valence are perceived as unpleasant or bad, which tends to trigger feelings that we tend to avoid. In contrast, situations with positive valence are perceived as pleasant or good. The second axis of the space is 'arousal', which refers to the level of activation or excitement in a given situation. High arousal means that someone is internally very activated or excited, while low arousal means that someone is internally not very activated (e.g. calm, relaxed, or even sleepy).

Specifically, in our scenario, four sub-regions of the valence-arousal space were considered: High-Valence and High-Arousal (HV-HA), High-Valence and Low Arousal (HV-LA), Low-Valence and High-Arousal (LV-HA), and Low-Valence and Low-Arousal (LV-LA). We gave additional examples and colors to these scenarios to make them easier to understand and to explain. For example, the HV-HA was assigned to the color 'Yellow', the HV-LA was assigned to the color 'Green', the LV-HA with the color 'Red' and LV-LA with the color 'Blue'. Figure 1 shows the valence-arousal diagram designed for the explanation of the four scenarios.

3.1. Material preparation

In order to record samples of signs with different emotional states, the first step was to explain the experiments and the goal of the acquisition. With this aim, we created a serie of slides to follow during the acquisition protocol in order to guide to the signer in the glosse to perform and the emotions to be expressed in each of the specific scenarios. The protocol consisted on the following steps:

- First, the signer was welcomed to the experiment and introduced to the task. The task consisted on recording at least ten repetitions of ten pre-selected glosses in four emotional scenarios. For eliciting the emotions of each scenarios, we asked the participant to put themselves in each of the situations and scenarios. After that, the recordings start.
- After the introduction to the experiment, the second step was to introduce briefly the concept of arousal and valence. The introduction was done with short descriptions as well as with graphic elements consisting on the Self-Assessment Manikin (SAM) scale and images extracted from the OASIS corpus (Kurdi et al., 2017) that demonstrated to elicit the emotional states explained in the description, as shown in Figure 1. For example, the picture of the kids on the left corner elicits high valence feelings; whereas the picture of the garbage on the right corner-down elicits emotions with negative valence.

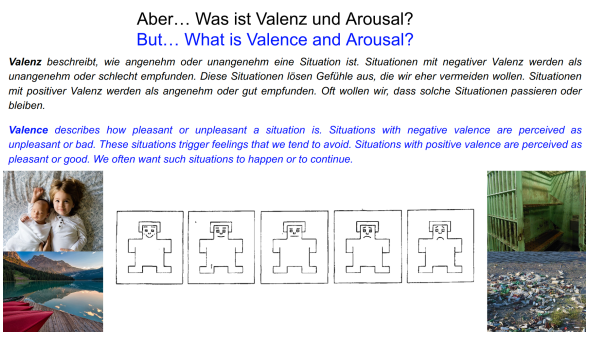

Figure 1: Description of Valence, originally only the German version is provided, the English version is added for the publication.

- After the description of valence and arousal, the next step was to clarify the scenarios and the axis system. For that part, we employed the designed valence-arousal diagram incorporating colors, the SAM images, and additional descriptions, as it can be seen in Figure 2
- Once the experiments were explained, the description of the scenarios starts, as well as the acquisition process. Each of the scenarios

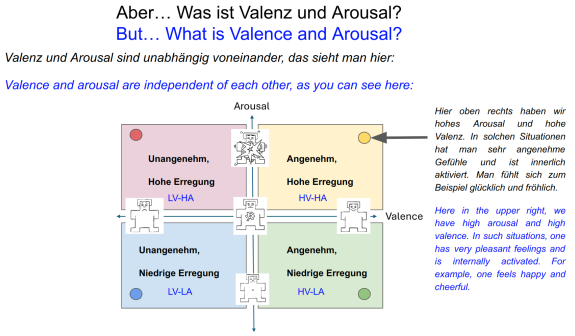

Figure 2: Graph with the Valence-Arousal diagram for explaining each scenario, together with an example, in this case for the yellow scenario of High Valence and High Arousal (HV-HA). The English translation (text in blue) is added for the publication but it was not in the original version.

follow the same structure, first, the valence-arousal graph is shown, indicating in which of the four scenarios we are, then, the scenario is described with a short text and a vignette associated with it (See Figure 3). After that, the recording process starts having a first video with a demo sign with neutral emotion, and followed by the picture of the scenario between sign transitions, in order to let the participant remember the scenario. The same process is repeated for the 4 scenarios.

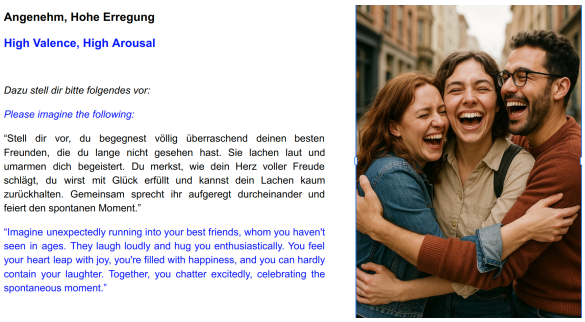

Figure 3: Stimuli for the High Valence and High Arousal scenario (HV-HA). The English translation appears in blue.

- Finally, the acquisition process concludes by thanking the participants for their time.

3.2. Participants and Recordings

For the recordings, we pre-selected a vocabulary of 10 signs in DGS to evaluate the data collection approach proposed and study whether emotions could be distinguished and recognized in this initial set of glosses to verify the validity of the approach. These signs correspond to the meanings of: JA (Yes), NEIN (No), KEINE-AHNUNG (No idea), PERSON (Person), LEUTE (People), AUTO (Car), FEUER (Fire), MEDIZIN (Medicine), ANDERS (Different), MEHR (More).

In order to record high-quality signs, the recording process was performed in the studios of the professional deaf interpreters. In total, three deaf interpreters participated in the recordings, two men and one woman. Only frontal cameras were used during the recording process and a green-screen in the background. All the recordings are in Full HD resolution (1920x1080) at 50 fps. As the recordings of each sign were captured in a continuous way, an annotation step was required. This annotation step was performed by a DGS-proficient professional with linguistic background that was in charge of selecting when signs start and end in the videos and annotate them with the ELAN tool.

After the annotations, videos were split automatically, resulting in 1,540 short-videos of one gloss signed with one of the specific emotions of each scenario. Videos have different durations in length ranging from 0.3 to 3.58 seconds ($\mu = 1.70; \sigma = 0.42$). Figure 4 shows the distribution of samples per subject and emotions in the collected dataset.

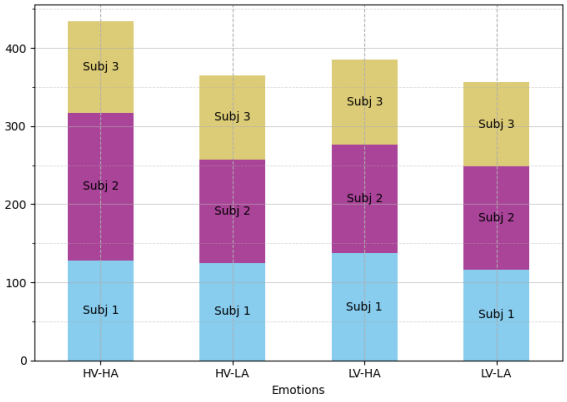

Figure 4: Distribution of the number of files per emotion and subjects. X-axis shows the four scenarios, and the y-axis the number of videos. Finally, colors represent the subjects.

4. Corpus Evaluation on Emotion Recognition

In order to test whether the acquisition methodology and approach was able to capture different emotions across signs in DGS, we trained emotion recognizers on the acquired corpus. As only three subjects were recorded, a subject-wise three-fold Leave-One-Out approach was followed, training in each fold with two subjects and testing the model in the third subject.

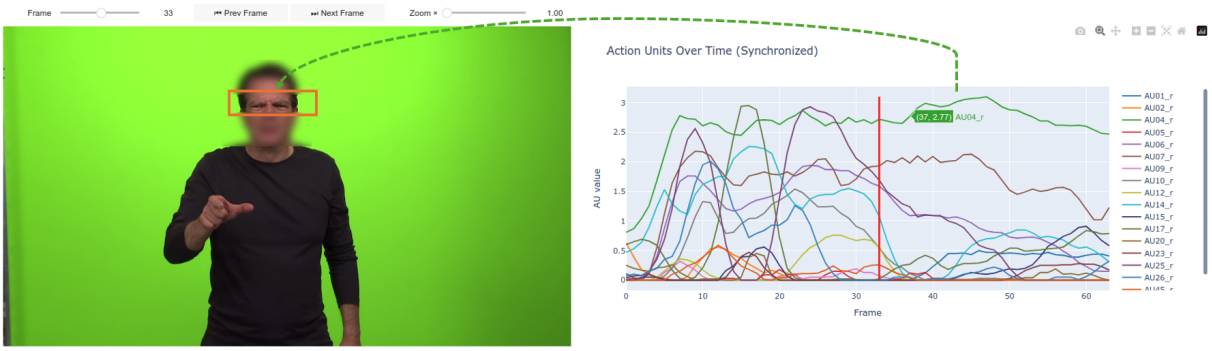

Figure 5: On the right, it can be observed the frame 33 of one of the videos of the 'Angry' scenario of one of the deaf interpreters signing 'PERSON'. On the left, the intensity of the Action Units per frame. Notice that in the AU 04 ('Brow Lowerer') is active with high intensity across the whole video. According to previous studies in facial emotion recognition (Ekman et al., 2002), the AU 04 is related to emotions such as 'Angry' or 'Disgust'.

4.1. Facial Features Extraction

Although pose and manual cues are key in sign language, non-manual features such as facial expressions or mouthing could carry linguistic and emotional information. For this reason, we extracted facial features to evaluate emotion recognition in Sign Language.

To recognize emotions from facial expressions, we employed the Action Units (AUs). AUs constitute the fundamental elements of the Facial Action Coding System (FACS), which categorizes human facial movements based on their observable muscle configurations and intensities. Each AU represents a specific facial muscle activation and is typically linked to changes in facial expression that reflect variations in an individual's psychological or emotional state, and traditionally they achieved considerable good results for recognizing emotions in spoken languages from facial expressions (Luna-Jiménez et al., 2022).

Among the available tools for extracting Action Units from video frames, we selected the OpenFace toolkit (Baltrusaitis et al., 2018). This framework has been extensively employed in facial emotion recognition for spoken languages and provides both the intensity of 17 AUs (AU1,2,4,5,6,7,9,10,12,14,15,17,20,23,25,26,45) and the presence of 18 AUs (same as intensity plus AU28). AU presence is represented as a binary value, equal to one when the AU is detected; whereas AU intensity is a continuous measure ranging from 0 to 5. These outputs are generated by two separate models that share a common pre-processing pipeline. First, facial regions are aligned to derive geometric and appearance-based features; then, these features are fed into two independent SVM classifiers, which estimate the AU presence and intensity for each frame, respectively.

Figure 5 represents the continuous Action Units generated by OpenFace for one of the frames of the videos of the 'Low Valence and High Arousal' scenario, in which the interpreter is signing 'person'. As it can be observed, the AU4 ('Brow Lowerer') is active during the whole video, which usually correlates with emotions that appear in the Low Valence and High Arousal quadrant such as anger, annoying or frustration (Kohler et al., 2004).

After extracting the 35 Action Units of each video frame, we calculated statistical aggregations from them in order to adapt the temporal dimension to static models. To accomplish this, the mean, standard deviation, maximum, minimum, skew, kurtosis and median statistics were calculated for each of the features across their temporal dimension, resulting in a total of 245 features. This approach, employed as our baseline, offers two primary advantages: first, its conceptual simplicity; and second, the robustness conferred by the averaging effect. To illustrate this, consider a video in which several frames deviate from a prototypical emotional expression, e.g., due to the subject closing their eyes. By employing statistics across the temporal dimension, the influence of such anomalous frames on the overall emotion recognition system is mitigated, assuming that the remaining frames predominantly reflect the target emotion.

4.2. Models Training

After adapting the features, we evaluated a range of statistical learning models implemented in the scikit-learn library. Specifically, we considered support vector machines (SVMs) with radial basis function (RBF) and linear kernels, logistic regression, ridge classifiers, decision trees, random forests, and multilayer perceptrons (MLPs). For the SVMs, Logistic

Regression (LR), and Ridge Classifiers (RC), the regularization parameter was varied over the values 1, 0.1, 0.01, 0.001, 0.0001. For the Random Forest (RF) models, we explored different numbers of estimators, namely 10, 20, 30, 40, 50, 100, 200, 300. Similarly, for the Decision Trees (DT), we evaluated different values for the minimum number of samples required to split an internal node using the same set of values (10, 20, 30, 40, 50, 100, 200, 300). For the MLP models, we explored the same set of values for the number of hidden units in the hidden layer (10, 20, 30, 40, 50, 100, 200, 300). All remaining hyperparameter were set to their default values as provided by the scikit-learn library.

5. Results on Facial Emotion Recognition

As can be observed in Table 1, the 245 features extracted from facial expressions seem to be promising features for performing facial expression recognition in Sign Language too. The random forest with 300 estimators, that received at its input the statistics computed from the regression values of the Action Units achieved an accuracy of 68.84% and a F1 of 67.96%, surpassing the Zero-Rule approach by +43.76 percentage points in accuracy and +57.89 in F1. This model is closely followed by the linear-SVM, the Logistic Regressor and the MLP.

Model	Hyperparam.	ACC	F1
ZeroRule	-	25.08	10.07
RBF-SVM	C=1	50.56	49.80
DT	Min. Split = 300	54.65	52.28
RC	C=0.01	61.97	58.93
MLP	Neurons = 20	65.72	64.43
LR	C=0.1	66.89	65.06
Linear-SVM	C=0.001	67.32	65.38
RF	**Estimators=300**	**68.84**	**67.96**

Table 1: Top models configurations - emotion recognition.

In order to understand variations across different subjects on their facial expressions, we evaluated the confusion matrices in the test sets of each of the folds of the subject-wise LOO strategy. Figure 6 shows these confusion matrices in which colors represents each of the subjects. The firs observation suggests that, as in spoken languages, different individuals rely differently in facial expressions to transmit emotions. For illustration, if we analyze subject two (in pink) and three (in orange), we can observe that facial expressions are much more relevant for recognizing emotions in them than for the subject one (in blue), reaching F1 values over 70% versus the 57% of accuracy obtained for the first subject.

Regarding the complexity on recognizing emotions across individuals, it seems that the emotion of the high valence, low arousal (HV-LA) scenario, resembling a calm feeling, is the worst detected by the models with an accuracy of 17.60% (22/125) for the first subject, 40.15% (53/132) for the second subject, and 40.75% (44/108) for the third subject. On the other hand, the high valence, high arousal (HV-HA) scenario, resembling emotions like happy, was the easiest to detect by the models, correctly recognizing the 84.56% (367/434) percent of the samples of the happy class. From the perspective of the data acquisition scenario, these results may indicate that the participants found more complex recording in the calm scenario maybe because the behavioral cues associated to this emotion could be not as familiar as for example those related to emotions such as happy or angry. Indeed, the calm emotion is not that frequently annotated, as commented in (Livingstone and Russo, 2018) and it is not part of the considered theoretical 'universal emotions' described by P. Ekman (Ekman, 1999). However, it is still one of the most important quadrants of the 'circumplex model of affect' (Posner et al., 2005), for this reason, it was considered in the study.

6. Conclusions

As in many languages (spoken and signed), facial expressions carry rich information that is key for fully understanding the transmitted message and the intentions of the communicator. These facial expressions complement the transmitted message and reduce disambiguation. However, although the field of Affective Computing have advanced in spoken languages, in sign languages the exploration of how emotions are expressed while signing has barely started.

In this article, we proposed a method to acquire a dataset of acted emotions, following previous datasets in the literature released for spoken languages, such as RAVDESS (Livingstone and Russo, 2018), in which professional actors are asked to perform emotions 'as if they were feeling them'; in the same way, in this dataset professional deaf interpreters were asked to put themselves in four different emotional scenarios of low arousal and low valence, low arousal and high valence, high arousal and low valence, and high arousal and high valence and perform 10 different signs.

In order to evaluate the suggested approach, we applied traditional methods in facial emotion recognition to extract Action Units across the videos. With this simple approach, the models were able to achieve an average F1 of 67.96% for the face modality, indicating a promising investigation line to perform emotion recognition in sign language,

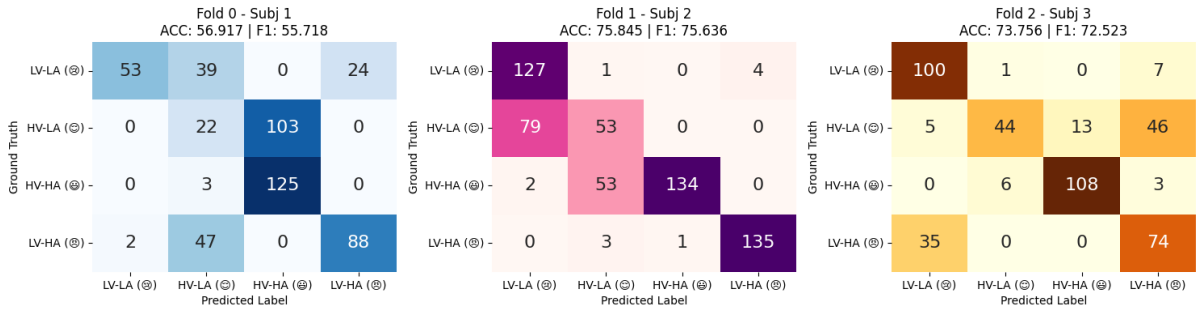

Figure 6: Confusion Matrices of the RF model for facial emotion recognition per subject for the test set of each fold of the LOO. Colors indicate the results of each subject.

and reinforcing the validity of the data acquisition.

As future lines, we would evaluate whether the expressed emotions are perceived in a similar way for deaf and non-deaf individuals, to study perceptual differences when you know the language versus when you do not. Additionally, we will extend the analysis to other non-manual channels (e.g. body posture) to understand the impact of other features and modalities in emotion recognition in Sign Language. Finally, one of the future steps would be to evaluate longer sentences and interactions, as well as annotating them in terms of the transmitted emotions.

7. Limitations

While the study provides preliminary results about the relevance of facial expressions, including more features and subjects in the training and evaluation of the models could benefit to the overall understanding of expressed emotions in Sign Language. However, recruiting deaf participants is challenging, specially considering that recordings can not always being made anonymous given the richness of DGS in terms of the information transmitted by manual and non-manual cues. Additionally, we acknowledge that extracted features are dependent on the libraries, so they could be biases and more accurate for some individuals than for others. Finally, as other datasets in Affective Computing, the first version of our acquisition was based on an acted scenario, in which the deaf interpreters needed to put themselves in hypothetical scenarios to express emotions, hence, models' performance may vary in the real-world, given the nuances associated to emotions. For this reason, these results should be understood as a first step and work in progress in the understanding of expressiveness of emotions in Sign Language.

8. Acknowledgments

This contribution is funded by the German Ministry for Education and Research (BMBF) through the BIGEKO project, grant number 16SV9094. It has also received funding from the project FORSocialRobots, financed by the Bavarian Research Foundation (AZ1594-23).

9. Bibliographical References

Marie Alaghband, Niloofar Yousefi, and Ivan Garibay. 2020. Facial Expression Phoenix (FePh): An Annotated Sequenced Dataset for Facial and Emotion-Specified Expressions in Sign Language. Harvard Dataverse. URL: https://dataverse.harvard.edu/citation?persistentId=doi:10.7910/DVN/358QMQ, doi:10.7910/DVN/358QMQ.

Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, page 59–66, Xi'an, China. IEEE Press. doi:10.1109/FG.2018.00019.

Phoebe Chua, Cathy Mengying Fang, Takehiko Ohkawa, Raja Kushalnagar, Suranga Nanayakkara, and Pattie Maes. 2025. EmoSign: A Multimodal Dataset for Understanding Emotions in American Sign Language. Arxiv Preprint. URL: https://arxiv.org/abs/2505.17090, arXiv:2505.17090.

Paul Ekman. 1999. Basic Emotions. In Tim Dalgleish and M. J. Powers, editors, *Handbook of Cognition and Emotion*, pages 4–5. Wiley. doi:10.1002/0470013494.ch3.

Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. 2002. *Facial Action Coding System: The*

Manual on CD-ROM. Instructor's Guide. Network Information Research Co., Salt Lake City, UT.

Kotaro Funakoshi and Yaoxiong Zhu. 2025. Emotion Recognition in Signers. URL: https://arxiv.org/abs/2512.15376, arXiv:2512.15376.

Michael Grimm, Kristian Kroschel, and Shrikanth Narayanan. 2008. The Vera am Mittag German Audio-Visual Emotional Speech Database. In *2008 IEEE International Conference on Multimedia and Expo*, pages 865–868, Antwerp, Belgium. doi:10.1109/ICME.2008.4607572.

Jari K. Hietanen, Jukka M. Leppänen, and Ulla Lehtonen. 2004. Perception of Emotions in the Hand Movement Quality of Finnish Sign Language. *Journal of Nonverbal Behavior*, 28(1):53–64. URL: http://link.springer.com/10.1023/B:JONB.0000017867.70191.68, doi:10.1023/B:JONB.0000017867.70191.68.

Vadim Kimmelman, Alfarabi Imashev, Medet Mukushev, and Anara Sandygulova. 2020. "Eyebrow Position in Grammatical and Emotional Expressions in Kazakh-Russian Sign Language: A Quantitative Study". *PLoS One*, 15(6):e0233731. doi:10.1371/journal.pone.0233731.

Desirée Kirst. 2024. Eyebrow Competition: Layering Emotion & Grammar in ASL. In *First Workshop on Expressing Emotions in Sign Languages (ExEmSiLa-2024)*, Universität Hamburg. URL: https://www.idgs.uni-hamburg.de/forschung/tagungen/expressing-emotions-in-sign-languages/programm/programm-pdfs/kirst---eyebrow-competition.pdf.

Christian G. Kohler, Travis Turner, Neal M. Stolar, Warren B. Bilker, Colleen M. Brensinger, Raquel E. Gur, and Ruben C. Gur. 2004. Differences in facial expressions of four universal emotions. *Psychiatry Research*, 128(3):235–244. doi:10.1016/j.psychres.2004.07.003.

Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems handling multiple Signers. *Computer Vision and Image Understanding*, 141:108–125. doi:10.1016/j.cviu.2015.09.013.

Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen. doi:10.25592/uhhfdm.10251.

Neha Kulshreshtha. 2024. The Effect of Emotional State on the Position of Wh-signs in Indian Sign Language. In *First Workshop on Expressing Emotions in Sign Languages (ExEmSiLa-2024)*, Universität Hamburg. URL: https://www.idgs.uni-hamburg.de/forschung/tagungen/expressing-emotions-in-sign-languages/programm/programm-pdfs/kulshreshtha---the-effect-of-emotional-state-on-the-position-of-wh.pdf.

Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. 2017. Introducing the Open Affective Standardized Image Set (OASIS). *Behav. Res. Methods*, 49(2):457–470. doi:10.3758/s13428-016-0715-3.

Steven R. Livingstone and Frank A. Russo. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5):1–35. doi:10.1371/journal.pone.0196391.

Cristina Luna-Jiménez, Ricardo Kleinlein, David Griol, Zoraida Callejas, Juan M. Montero, and Fernando Fernández-Martínez. 2022. A proposal for multimodal emotion recognition using aural transformers and action units on ravdess dataset. *Applied Sciences*, 12(1). URL: https://www.mdpi.com/2076-3417/12/1/327, doi:10.3390/app12010327.

Ester Martinez-Martin and Antonio Fernández-Caballero. 2025. Improved Human Emotion Recognition from Body and Hand Pose Landmarks on the GEMEP Dataset using Machine Learning. *Expert Systems with Applications*, 269:126427. URL: https://www.sciencedirect.com/science/article/pii/S0957417425000491, doi:10.1016/j.eswa.2025.126427.

Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder. 2012. The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Transactions on Affective Computing*, 3(1):5–17. doi:10.1109/T-AFFC.2011.20.

Carol Neidle, Augustine Opoku, and Dimitris Metaxas. 2022. ASL Video Corpora & Sign Bank: Resources Available through the American Sign Language Linguistic Research Project (ASLLRP). URL: https://arxiv.org/abs/2201.07899, arXiv:2201.07899.

Fabrizio Nunnari, Alakshendra Singh, and Patrick Gebhard. 2025. Color Histogram Equalization

and Fine-Tuning to improve Expression Recognition of (partially occluded) Faces on Sign Language Datasets. In *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, IVA Adjunct '25, Berlin, Germany. Association for Computing Machinery. doi:10.1145/3742886.3756731.

Nina-Kristi Pendzich, Jens-Michael Cramer, Thomas Finkbeiner, Annika Herrmann, and Markus Steinbach. 2022. How do signers mark conditionals in German Sign Language? Insights from a Sentence Reproduction Task on the use of nonmanual and manual Markers. *Hrvatska revija za rehabilitacijska istraživanja*, 58:206–226. doi:10.31299/hrri.58.si.11.

Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03). URL: https://dx.doi.org/10.1017/s0954579405050340, doi:10.1017/s0954579405050340.

Sarah Schwarzenberg, Jenne Mertens, Alexander Eisenzimmer, Simon Kollien, and Annika Herrmann. 2024. Expressing Emotions in the DGS Corpus: The Interaction of Lexemes and Nonmanuals. In *First Workshop on Expressing Emotions in Sign Languages (ExEmSiLa-2024)*, Universität Hamburg. URL: https://www.idgs.uni-hamburg.de/forschung/tagungen/expressing-emotions-in-sign-languages/programm/programm-pdfs/schwarzenberg-et-al---expressing-emotions-in-the-dgs-corpus.pdf.

Bogdan Vlasenko, Björn Schuller, Andreas Wendemuth, and Gerhard Rigoll. 2007. Combining frame and turn-level information for robust recognition of emotions within speech. In *Interspeech 2007*, pages 2249–2252, Antwerp, Belgium. doi:10.21437/Interspeech.2007-611.

Mingming Zhang, Yanan Zhou, Xinye Xu, Ziwei Ren, Yihan Zhang, Shenglan Liu, and Wenbo Luo. 2023. "Multi-view Emotional Expressions Dataset using 2D Pose Estimation". *Sci. Data*, 10(1):649. doi:10.1038/s41597-023-02551-y.