applied sciences

MDPI

Communication

DevEmo—Software Developers' Facial Expression Dataset

Michalina Manikowska, Damian Sadowski, Adam Sowinski and Michal R. Wrobel *

Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdańsk, Poland
* Correspondence: michal.wrobel@pg.edu.pl

Abstract: The COVID-19 pandemic has increased the relevance of remote activities and digital tools for education, work, and other aspects of daily life. This reality has highlighted the need for emotion recognition technology to better understand the emotions of computer users and provide support in remote environments. Emotion recognition can play a critical role in improving the remote experience and ensuring that individuals are able to effectively engage in computer-based tasks remotely. This paper presents a new dataset, DevEmo, that can be used to train deep learning models for the purpose of emotion recognition of computer users. The dataset consists of 217 video clips of 33 students solving programming tasks. The recordings were collected in the participants' actual work environment, capturing the students' facial expressions as they engaged in programming tasks. The DevEmo dataset is labeled to indicate the presence of the four emotions (anger, confusion, happiness, and surprise) and a neutral state. The dataset provides a unique opportunity to explore the relationship between emotions and computer-related activities, and has the potential to support the development of more personalized and effective tools for computer-based learning environments.

Keywords: emotion recognition; dataset; deep learning

check for updates

Citation: Manikowska, M.; Sadowski, D.; Sowinski, A.; Wrobel, M.R. DevEmo—Software Developers' Facial Expression Dataset. *Appl. Sci.* **2023**, *13*, 3839. https://doi.org/10.3390/app13063839

Academic Editors: Paula Viana, Pedro Carvalho and Teresa Chambel

Received: 7 February 2023
Revised: 5 March 2023
Accepted: 13 March 2023
Published: 17 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The COVID-19 pandemic has forced employees, students and pupils to change the way they carry out their work and teaching activities. Working and studying from home, using computers as the only means of communication, has made it necessary to adapt to the new situation and find ways to work effectively in these unusual conditions. For many people, working remotely has proven to be very advantageous due to the flexibility in organizing working hours and the reduced costs associated with commuting. Even after the restrictions have been lifted, many organizations continue to allow remote work. Additionally, more and more teaching courses are being offered this way. However, working and learning remotely can be emotionally difficult for many people. The lack of direct contact with colleagues can lead to feelings of isolation and loneliness, which can negatively affect mood and motivation.

From the point of view of a teacher who does not have direct contact with students, it is difficult to identify the moments when students have problems with the material presented or with solving tasks. Students' emotions are an important indicator of their understanding and difficulties in learning. Without direct contact, the teacher is unable to perceive these emotions, making it difficult to effectively adapt teaching methods to the students' needs.

One way to support employees and students is to equip and enhance work and learning support tools with emotion recognition mechanisms. Using an affective loop [1], the system can trigger actions to support the user's work and learning when it detects an undesirable emotional state.

Effective training of deep neutral network models, such as those used to recognize the emotions of computer users, requires the use of large datasets [2]. Such datasets must be

appropriately annotated, that is, they must contain information about which emotion is being presented at any given time.

Within the framework of the research described in this paper, a DevEmo dataset of video recordings of computer users at work has been prepared and annotated with respect to the emotions expressed. The collection was based on video recordings of students solving programming tasks. The data were recorded in the students' real work environment.

This dataset can be used to train pre-trained models, such as Xception [3], VGG-16 [4] or ResNet50 [5], using the transfer learning technique to recognize students' emotions as they solve programming problems. In this way, a student's difficulties in solving tasks can be automatically detected, allowing teachers to react quickly and provide appropriate assistance. The model can also be used to analyze the student's learning process to better understand which teaching methods are most effective. However, the application is not limited to distance learning. The reactions are versatile enough that the dataset can be used to train emotion recognition models for all users of computer systems during work and entertainment.

The remainder of the paper is organized as follows: Section 2 gives an overview of related datasets, Section 3 describes data collection, and Section 4 describes the annotation process. In Section 5, the DevEmo dataset is presented, and finally, Section 6 concludes the contribution.

2. Related Work

Among the commonly available datasets used to train artificial intelligence models for emotion recognition from facial expressions, a significant number consists only of images tagged with emotional states [6–14]. In contrast, among the datasets containing video clips, most contain recordings of posed emotions recorded under controlled conditions. Table 1 summarizes the available datasets with video recordings of facial expressions.

Table 1. List of datasets with video recordings of facial expressions of emotions.

Dataset	Number of Recordings	Number of Participants	Conditions
MMI [15]	over 1500	19	posed in controlled environment
CK+ [16]	486	97	posed and spontaneous in controlled environment
DEAP [17]	NA	22 (with facial recordings)	spontaneous in controlled environment
Belfast [18]	1400	256	spontaneous in controlled environment
RECOLA [19]	3.8 h of recordings (only 5 min of recording is annotated)	46	spontaneous in controlled environment
ISED [20]	428	50	spontaneous in controlled environment
Aff-Wild [21]	300	200	spontaneous in uncontrolled environment
RAVDESS [22]	7356	24	posed in controlled environment
DEFE [23]	164	60	spontaneous in controlled environment

The DevEmo dataset is unique among the datasets that were identified in its focus on the spontaneous emotional reactions of students while they are programming in their natural work environment. In contrast, the only other dataset recorded in an uncontrolled environment was the Aff-Wild dataset [21], which focused on a different population. The

other datasets that were identified consisted of either posed facial expressions or spontaneous reactions recorded in controlled environments. However, none of these datasets specifically focused on computer users. The Aff-Wild dataset, for example, contains recordings of stage performances, TV shows, interviews, or lectures. Therefore, the presented DevEmo dataset is more suitable for developing models that focus on recognizing computer users' emotions while working and learning.

3. Recordings Collection

The main purpose of the developed dataset is to provide training data for deep learning models aimed at recognizing the emotions of users of computer systems in their natural work environment. In developing the dataset, we focused specifically on software engineers, including those who are just learning to code.

For data collection, dedicated software was developed to capture video recordings of the users' faces and computer screens as they completed the programming tasks. The software was available online through a website to which participants were given individual access.

During the session, each user had to complete five programming tasks consisting of completing a piece of program code written in the Java programming language. The user interface was prepared in the form of a simplified IDE editor that, among other things, colored the syntax and allowed the user to compile and execute the written program. The correctness of the solved tasks was verified by running pre-prepared unit tests. To evoke emotions, the system periodically generated malicious behavior during programming, such as adding extra characters, deleting semicolons, or duplicating the characters that were entered [24]. Before starting the study, the participant had to allow the website to record video camera and desktop images. These recordings were uploaded to the server after each task was completed.

Computer science students at the engineering level at the Faculty of Electronics, Telecommunications and Informatics at the Gdansk University of Technology were invited to participate in the study. Each participant gave written consent to participate in the study and video recording. The consent form also included an option to give permission to publish the image for scientific purposes. The survey was conducted in late May and early June 2018. A total of 212 participants took part, 117 of whom consented to the publication of a video with their image.

Recordings of 48 participants aged 19–21, including 9 females and 39 males, were selected from the collected set. The first acceptance criterion was the participant's consent to the publication of the image. This was a prerequisite for making the collection available to the scientific community. The second selection criterion was the number of tasks solved and, consequently, the number of videos submitted. At this stage, participants who had not solved all the tasks were rejected. The final list of participants whose recordings were annotated was selected at random.

4. Annotation

The most important task in developing the dataset for training deep learning models was to manually label the recordings. This was the only way to create a valuable dataset. The labeling process was tedious and repetitive. Three people, graduate students, one female and two males, performed the annotation process. Prior to this work, they were trained by learning about emotion recognition, models of emotional states, and behavior analysis. They also had previous experience working with the BORIS program. The process of labeling the recordings started in June 2022 and continued until the end of December 2022. The nearly six months spent on this task perfectly illustrates how time-consuming it was.

4.1. Annotation Procedure

To maintain the credibility of the study, each record had to be annotated by a minimum of three members of the research team. In addition, achieving consistency and the ability to post-analyze and process the results of the tagging required the creation of clear process guidelines in the form of an annotation protocol and adherence to the protocol by all investigators.

The annotation protocol consisted of eight points:

1. Each video is annotated once by all three annotators.
2. Each annotation involves assigning one of eight labels: *Happiness*, *Sadness*, *Anger*, *Disgust*, *Fear*, *Surprise*, *Confusion* and *No face*.
3. Each non-annotated part of the video is interpreted as a *"neutral"* emotional state.
4. The scale of the emotion annotation contains 0 (not exists) and 1 (exists), which means there are no intermediate states.
5. Each annotation has a timestamp for the start and stop of the expression.
6. Expressions do not overlap each other, that is, the previous expression must be annotated as a stop before starting the next annotation.
7. Covering the mouth or at least one eye by the examined people or placing the face outside the frame must be annotated with a *"no face"* label.
8. Emotions expressed as verbal comments may be annotated as well.

4.2. Annotation Process

The BORIS software [25] was used to annotate the recordings, as it does an excellent job of labeling the events that occur in the recordings. Its intuitive interface, as well as the ability to choose the method of labeling (mouse clicks, keystrokes, or an interactive panel), significantly streamlined the team's work. The software allows the recording of point events or sequential events (time interval). The chosen method is sequence-marking in the form of indicating the beginning and end of the interval.

Each team member had their own project file with the extension .boris, in which all information about the labeled recordings and the emotions labeled in them (type of emotion, duration) were stored.

The average time for labeling recordings of one research participant was about 30 min (5 tasks of 5 min each). In particular, the participant's lack of emotion did not speed up the process, as there were cases when a person who did not show emotion for most of the recording decided to show some emotion at the very end. Consequently, each recording had to be meticulously reviewed. A summary of the annotated data is shown in Table 2.

Table 2. Summary of data subjected to annotation.

Number of videos	219
Number of participants	48 (9 female and 39 male)
Age of participants	19–21
Video duration	up to 5 min
Number of ratings per video	0–18
Applied labels	happiness, sadness, disgust, fear, surprise, anger, confusion, no face
Rating values	Discrete scale of 0 and 1

4.3. Annotation Results

A total of 219 recordings were labeled and more than 1600 labels were assigned. Most of the participants did not show any emotion, they were very focused. Quite a few tended to cover part of their face. This mainly happened when they were strongly analyzing a problem they had just encountered. A few participants were so engaged in performing the tasks that they did not notice that a part of their face was coming out of the recorded

frame. During the annotation process, none of the emotions of sadness, disgust and fear were labeled.

5. DevEmo Dataset

The final dataset consists of 217 video clips, 108 of which were labeled with a neutral emotion, while the remaining 109 were tagged with one of the emotions of happiness, surprise, anger and confusion. Sample frames selected from video clips included in the dataset are shown in the Figure 1.



Figure 1. Sample frames from DevEmo dataset video clips.

A summary of the contents of the dataset is presented in the Table 3.

Table 3. Dataset content summary.

Number of video clips	217
Number of participants	33 (5 female and 28 male)
Age of participants	19–21
Average single video clip duration	$\tilde{3}$s
Labels	anger, confusion, happiness, surprise, neutral

The distribution of labeled emotions is not balanced. The most common emotion is confusion, which was assigned to 56 clips. In addition, 21 clips were labeled as surprise and 31 as happiness. Moreover, one clip was tagged with the emotion of anger. The graph in the Figure 2 shows the distribution of labels assigned to video clips from the DevEmo dataset.

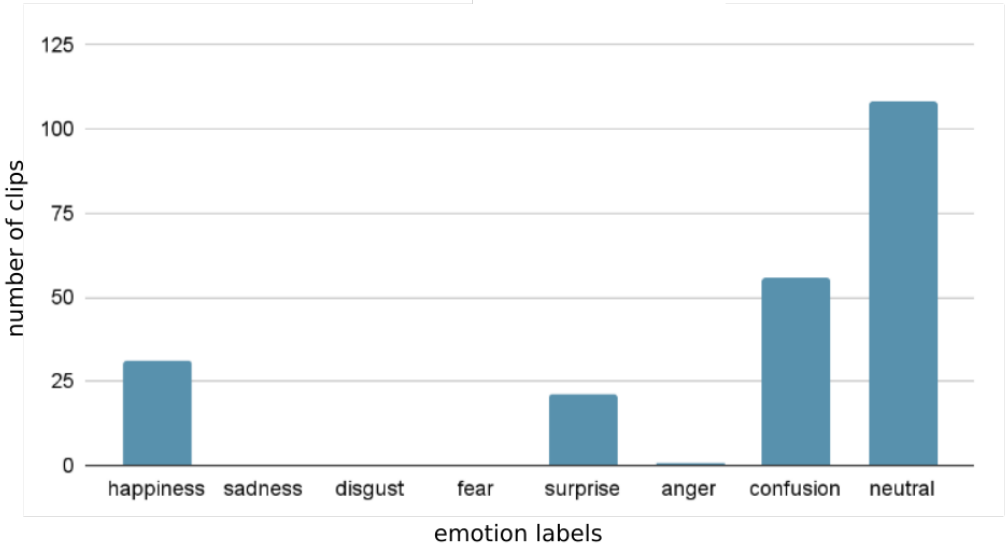


Figure 2. Distribution of labels assigned to video clips.

On average, 3.3 labels were assigned for each participant, with one user's videos being labeled 12 times and 13 users only once. The emotion of confusion was observed among as many as 26 participants. In addition, 16 expressed surprise, 13 happiness, and only one anger.

According to the premise, a video clip was given a label when at least two of the three annotators assigned the same label to it. Details of the number of common annotator observations are shown in Table 4.

Table 4. Summary of annotators' observations.

Number of Annotators that Labeled the Same Emotion	Happiness	Sadness	Disgust	Fear	Surprise	Anger	Confusion
3	20	0	0	0	5	0	17
2	11	0	0	0	16	1	39

5.1. Dataset Preparation

For the purposes of the study, a BORIS program project file was prepared, in which the labels to be used for annotating the source video files were defined in the form of an ethogram. In addition to this file, each annotator received a set of 219 video files. The annotation results for each video file were saved in a separate BORIS project file. Once the annotation was completed, all files were placed in a common repository.

For the purpose of processing the labels of all researchers, a tool has been created that integrates and merges the output files into a single file containing information about fragments annotated with the same labels. The acceptance threshold was at least two of the three researchers annotating the same fragment with the same label.

The next step in preparing the dataset was to generate output video files with fragments labeled with emotions. For this purpose, a tool using the *ffmpeg* program was developed. The following assumptions were adopted:

- Covering fragments marked by 2 or 3 observers are treated as one fragment marked by 3 observers;
- Each fragment has been given a time margin of 1.5 s from the beginning and 0.5 s from the back of the recording;
- Fragments marked *no face* have been rejected; and
- The remaining unlabeled fragments were marked as "neutral" in the number of labels for the subject and the length being the average length of the labels for the subject.

As a result of the script, Matroska Media Container (MKV) video files were created with the original encoding in h264/opus or vp8/opus. The average length of the clips is 6.3 s, and the longest lasts 23 s. In addition, the dataset contains a fragments.csv file describing the dataset that contains the following information: ID examined, task name, label, annotating observers, all observers, duration, filename, video codec. The complete package was compressed into a zip archive file. The dataset can be shared with the scientific community via https://devemo.affectivese.org (accessed on 1 March 2023) website.

5.2. Validation

In order to validate the results, the videos that had been the subject of the annotation process were analyzed in Noldus FaceReader 8.0. This program provides emotion recognition based on facial expression analysis, relying primarily on a subset of Action Units (AUs) provided by the Facial Action Coding System (FACS) [26]. FaceReader is a well-established system in the research community that is known for its relatively high level of accuracy in recognizing emotions [27]. However, it is also known for its vulnerability to poor lighting conditions and for overestimating the emotion of anger.

From the results of the video processing with FaceReader, excerpts corresponding to manually annotated clips from the DevEmo dataset were analyzed. Among the data obtained from FaceReader, the confidence of occurrence of discrete emotions was selected for further analysis. It should be noted that this program recognizes a different set of emotions that do not completely overlap with the labels assigned in the DevEmo set: *Happy*, *Sad*, *Angry*, *Surprised*, *Scared*, *Disgusted* and *Neutral*.

For each excerpt, the maximum confidence of each emotion occurrence was calculated. Figure 3 shows two confusion matrices of annotated labels and FaceReader analysis results. The Y-axis shows labels from the DevEmo dataset, and the X-axis shows labels from FaceReader. The left chart presents a summary for the dominant emotion. For each excerpt, the emotion with the highest maximum confidence value was selected and matched with the emotion selected during the manual annotation process. A similar summary, but for the average of the maximum confidence in recognizing each emotion, is shown in the right chart.

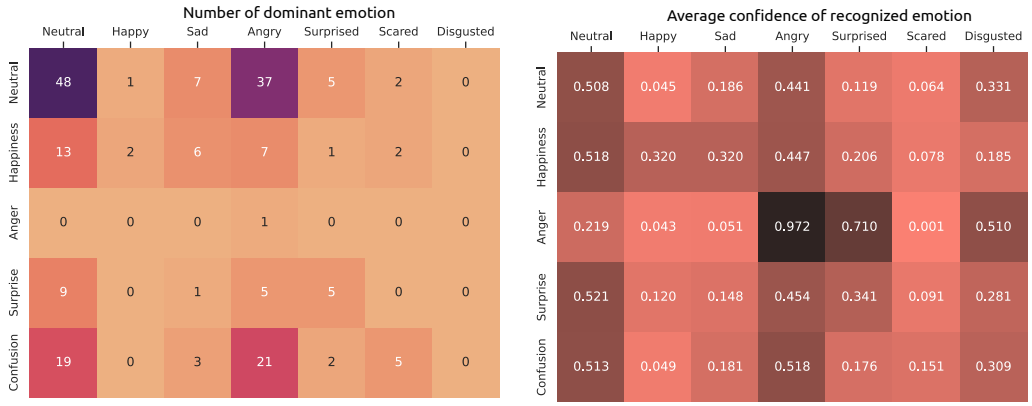


Figure 3. Confusion matrices of annotated labels and FaceReader analysis results.

The results of FaceReader only partially match the assigned labels. The figure clearly shows the over-recognition of the emotion of anger and the neutral state. When these two emotions are ignored, the automatic and manual annotated results become more consistent. Looking at the averages of maximum confidence, *happiness* has the lowest accuracy, being recognized as *sad* actually as often as *happy*. Confusion, which is not an emotion recognized by FaceReader, was most often recognized as *angry* and *disgusted*, which is most reasonable since recognition is mostly based on AU.

Manual analysis of the videos showed that the fragments whose FaceReader results differed the most from manual annotation were mostly recorded in poor lighting conditions

or low resolution. In addition, analysis by a fourth author not involved in the annotation process confirmed the validity of the manual annotation in all cases. This confirms the shortcomings of existing tools and the rationale for sharing datasets of emotion-annotated video clips, such as DevEmo, to develop new emotion recognition tools.

5.3. Limitations

The DevEmo dataset has several limitations in the context of its usage in the deep learning domain. The dataset is relatively small, consisting of only 217 video clips. This can limit the ability of deep learning models to accurately capture and generalize the emotions present in the dataset. Small datasets can lead to overfitting, where models are trained on the specific examples present in the dataset. Therefore, in its current form, the dataset is more suitable for training pre-trained models using the transfer learning technique, to better recognize students' emotions when solving programming tasks.

Finally, the dataset only includes clips annotated with three emotions: happiness, surprise, and confusion. This limited range of emotions may not be representative of the full range of emotions present in real-world situations. This can lead to difficulties in training deep learning models that can accurately detect and classify emotions that may occur in real-life scenarios. However, observations conducted so far on a broader set of recordings, show that indeed, while working with the computer, users show a limited set of emotions.

5.4. Threats to Validity

During the construction of the dataset, a methodological analysis of threats to the validity of the study was carried out in accordance with the approach proposed by Wohlin et al. [28]. This allowed threats to be identified and categorised into three types: internal validity, external validity and construct validity.

The first type of threats is internal validity, which refers to factors that can affect the independent variables [28]. Within this group, the identified threat is the manual annotation of emotions, a subjective assessment of facial expressions that depends on the experience of the annotator, which can lead to misinterpretation of the participant's behaviour (e.g., mistaking a frown for confusion). To minimise this threat, all recordings were assessed by three annotators. In addition, a detailed annotation protocol was prepared which all researchers were required to follow.

The highest number of threats was identified in the external validity group, which refers to factors that limit the ability to generalize results [28]. The first of the potential threats considered is the Hawthorne effect, which means that the participant's awareness of being observed may change the way they behave [29]—either intensifying or weakening the expression of emotions. Participants were aware that the purpose of the study was to identify emotions while solving programming tasks, which may have resulted in some bias. However, the design of the study, which involved participants in their natural working environment, ensured that this threat was not minimized in this study.

However, such a natural working environment raises another threat to external validation. Lighting conditions and the position of the camera capturing participants' facial expressions can affect the reliability of annotations. This threat cannot be mitigated, as the primary objective was to develop the dataset in an uncontrolled environment.

Covering the face is another of the risks identified. However, as with the previous threat, due to the desire to record the natural behaviour of the participants, no instruction was given in any way on how they should behave. In order to minimise this risk, it was decided not to include fragments in the dataset where participants covered their face with their hand or other objects.

Finally, the most important of the risks identified is the sampling of participants. Only 33 Polish students aged between 19 and 21, studying at the same university and department, took part in the study. This means that deep learning models trained on data sets may be less effective at recognising the emotions of people of different ages or races.

The final threat identified belongs to the group of construct validity, which refers to the ability to generalize the results of an experiment to the concept or theory underlying the experiment [28]. An incomplete or inadequate selected emotion model, which does not cover the entirety of the respondents' expressions, may compromise the validity of the labelled data. To address this threat, we relied on industry standards (the Ekman model). However, the results of the study indicate that when it came to recognising students' emotions while solving programming tasks, this was not the most accurate choice. Of the seven basic emotions, only three of them (happiness, surprise and confusion) were identified more than once in the annotation process.

6. Conclusions

The DevEmo dataset is a unique collection of video clips that capture the emotional expressions of students as they perform programming tasks. This dataset is part of a larger study that aims to explore the emotional states of both software developers and novice programmers. The videos were recorded in the students' actual work environment, and the emotions shown were spontaneous, providing an authentic representation of the students' emotional states as they were programming.

The process of creating the DevEmo dataset involved a thorough and systematic labeling process to ensure accurate and reliable data. Each video file was manually labeled by at least three annotators using the annotation protocol. They were responsible for labeling each of the emotional expressions displayed in the video according to the defined labels. In total, each annotator marked approximately 550 labels over the course of all recorded videos. To ensure consistency in the process, the annotators worked independently, and the results of their work were then integrated using a custom-built tool. It allows to merge the labeling lists of all annotators into a single file. The final list contained only co-labeled sequences identified by at least two annotators, ensuring a high degree of accuracy in the labeling process. In addition, neutral sequences of appropriate number and length for each participant were added to the dataset to provide a comprehensive representation of the students' emotional experiences.

After the labeling process was completed, the labeled sequences were extracted into separate video files. In total, 217 video fragments were extracted from the 33 study participants, representing a diverse range of emotional experiences. The results of the labeling process indicated that the dominant emotions displayed by the students were happiness and confusion. The extracted fragments, along with a file describing the contents, can be shared with the scientific community.

The DevEmo dataset can be a valuable resource for future research and studies. The manually labeled video fragments can be used to train pre-trained deep learning models using transfer learning techniques to recognize the emotions of computer users. This would enable the development of new technologies that can detect and analyze emotional responses in real time. The potential applications of the DevEmo dataset are vast, and its use as a source for further research and studies can help advance our knowledge in the field of emotion recognition and distance learning.

In previous studies, we collected more than 1500 videos of recorded facial expressions with more than 400 participants solving programming tasks. In the next steps, we plan to label the next portions of the videos and incrementally extend the DevEmo dataset, with a particular focus on balancing the number of samples for each emotion. Ultimately, the dataset should be large enough to develop deep learning models that effectively recognize students' emotions.

The manual labeling process used to create the DevEmo dataset can be improved in future iterations by refining the set of emotions used to annotate the recordings. One potential improvement would be to discard emotions that are rarely expressed by study participants, such as sadness, fear, or anger. This would allow for a more targeted and less distracting annotation process. In addition, new emotions that are more frequently expressed by participants, such as focus, could be added to the set to provide a more

comprehensive representation of students' emotional experiences. By refining the set of emotions used in the annotation process, subsequent iterations of the DevEmo dataset can provide a more detailed and nuanced understanding of students' emotional experiences as they learn to code.

Author Contributions: Conceptualization, M.R.W.; Methodology, M.M., A.S. and M.R.W.; Software, D.S. and A.S.; Validation, M.M. and D.S.; Formal analysis, D.S. and A.S.; Investigation, M.M. and A.S.; Resources, M.R.W.; Writing—original draft, M.M., D.S., A.S. and M.R.W.; Writing—review & editing, M.R.W.; Visualization, M.R.W.; Supervision, M.R.W.; Project administration, M.R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Written informed consent was obtained from participants for publication of their image.

Data Availability Statement: The dataset is available at https://devemo.affectivese.org (accessed on 1 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Landowska, A.; Szwoch, M.; Szwoch, W. Methodology of affective intervention design for intelligent systems. *Interact. Comput.* **2016**, *28*, 737–759. [CrossRef]
2. Roh, Y.; Heo, G.; Whang, S.E. A survey on data collection for machine learning: A big data-ai integration perspective. *IEEE Trans. Knowl. Data Eng.* **2019**, *33*, 1328–1347. [CrossRef]
3. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
4. Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ. (IJSRP)* **2019**, *9*, 143–150. [CrossRef]
5. Theckedath, D.; Sedamkar, R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **2020**, *1*, 79. [CrossRef]
6. Lyons, M.J.; Akamatsu, S.; Kamachi, M.; Gyoba, J.; Budynek, J. The Japanese female facial expression (JAFFE) database. In Proceedings of the third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 14–16.
7. Goeleven, E.; De Raedt, R.; Leyman, L.; Verschuere, B. The Karolinska directed emotional faces: A validation study. *Cogn. Emot.* **2008**, *22*, 1094–1118. [CrossRef]
8. Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; Baker, S. Multi-pie. *Image Vis. Comput.* **2010**, *28*, 807–813. [CrossRef] [PubMed]
9. Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D.H.; Hawk, S.T.; Van Knippenberg, A. Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* **2010**, *24*, 1377–1388. [CrossRef]
10. Du, S.; Tao, Y.; Martinez, A.M. Compound facial expressions of emotion. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1454–E1462. [CrossRef] [PubMed]
11. Barsoum, E.; Zhang, C.; Ferrer, C.C.; Zhang, Z. Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016; pp. 279–283.
12. Fabian Benitez-Quiroz, C.; Srinivasan, R.; Martinez, A.M. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5562–5570.
13. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.* **2017**, *10*, 18–31. [CrossRef]
14. Dhall, A.; Goecke, R.; Lucey, S.; Gedeon, T. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 2106–2112.
15. Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-based database for facial expression analysis. In Proceedings of the 2005 IEEE international conference on multimedia and Expo, Amsterdam, The Netherlands, 6 July 2005; p. 5.
16. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.

17. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]
18. Sneddon, I.; McRorie, M.; McKeown, G.; Hanratty, J. The belfast induced natural emotion database. *IEEE Trans. Affect. Comput.* **2011**, *3*, 32–41. [CrossRef]
19. Ringeval, F.; Sonderegger, A.; Sauer, J.; Lalanne, D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–8.
20. Happy, S.; Patnaik, P.; Routray, A.; Guha, R. The Indian spontaneous expression database for emotion recognition. *IEEE Trans. Affect. Comput.* **2015**, *8*, 131–142. [CrossRef]
21. Zafeiriou, S.; Kollias, D.; Nicolaou, M.A.; Papaioannou, A.; Zhao, G.; Kotsia, I. Aff-wild: Valence and arousal'In-the-Wild'challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 34–41.
22. Livingstone, S.R.; Russo, F.A. 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* **2018**, *13*, e0196391. [CrossRef] [PubMed]
23. Li, W.; Cui, Y.; Ma, Y.; Chen, X.; Li, G.; Zeng, G.; Guo, G.; Cao, D. A spontaneous driver emotion facial expression (defe) dataset for intelligent vehicles: Emotions triggered by video-audio clips in driving scenarios. *IEEE Trans. Affect. Comput.* **2023**, *14*, 747–760. [CrossRef]
24. Wrobel, M.R. Applicability of emotion recognition and induction methods to study the behavior of programmers. *Appl. Sci.* **2018**, *8*, 323. [CrossRef]
25. Friard, O.; Gamba, M. BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* **2016**, *7*, 1325–1330. [CrossRef]
26. Brodny, G.; Kołakowska, A.; Landowska, A.; Szwoch, M.; Szwoch, W.; Wróbel, M.R. Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions. In Proceedings of the 2016 9th International Conference on Human System Interactions (HSI), Portsmouth, UK, 6–8 July 2016; pp. 397–404.
27. Borsos, Z.; Jakab, Z.; Stefanik, K.; Bogdán, B.; Gyori, M. Test–Retest Reliability in Automated Emotional Facial Expression Analysis: Exploring FaceReader 8.0 on Data from Typically Developing Children and Children with Autism. *Appl. Sci.* **2022**, *12*, 7759. [CrossRef]
28. Wohlin, C.; Runeson, P.; Höst, M.; Ohlsson, M.C.; Regnell, B.; Wesslén, A. *Experimentation in Software Engineering*; Springer Science & Business Media: Berlin, Germany, 2012.
29. Sedgwick, P.; Greenwood, N. Understanding the Hawthorne effect. *BMJ* **2015**, *351*, h4672. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.