Development and Validation of a Facial Emotion Classifier for Applications in the Treatment of Autism Spectrum Disorder

Cathy Hou

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting one in 40 children in the United States and is associated with impaired social interactions, restricted interests, and repetitive behaviors. Currently, making one-to-one individualized therapy more accessible to patients with ASD poses a challenge to the healthcare system. Previous studies have demonstrated the promise of applying mobile systems with real-time emotion recognition to autism therapy, but existing emotion recognition platforms have shown limited performance on children with ASD. The development of a new emotion classifier designed specifically for pediatric populations, trained with images crowdsourced from a mobile charades-style game: *Guess What?*, was proposed. Videos of children with ASD portraying emotions during remote game sessions of *Guess What?* were acquired*,* yielding 6,344 frames that were manually labeled with four of the Ekman universal emotions (happy, scared, angry, sad), a "neutral" class, and "n/a". The data were pre-processed, and a classifier was trained with a transfer-learning and neural-architecture-search approach using the Google Cloud AutoML pipeline. The resulting classifier was evaluated against existing approaches (Microsoft's Azure Face API and Amazon Web Service's Rekognition) using the standard metrics of F1 score and demonstrated superior performance across all evaluated emotions. The results demonstrate that a model trained with a pediatric dataset outperforms existing emotion-recognition approaches for the population of interest. This study supports a new strategy to develop at-home precision therapy for children with ASD. Mobile games like Guess What? not only generate data but also can serve as an AI-based personalized therapeutic intervention.

Keywords: autism spectrum disorder, machine learning, emotion classifier, mobile game

Development and Validation of a Facial Emotion Classifier for Applications in the Treatment of Autism Spectrum Disorder

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting one in 40 children in the United States and is associated with impaired social interactions, restricted interests, and repetitive behaviors [1] [2]. While there is no cure, studies have shown the efficacy of Applied Behavioral Analysis (ABA) therapy, if administered at a young age and customized to address the child's unique deficits [3] [4]. However, caring for a child with ASD can generate a financial burden on the family [5]. Additionally, the increasing prevalence of the condition is resulting in a short supply of certified specialists, further hindering treatment options [6].

Guess What? [7] [8] [9], a charades-style mobile game that delivers at-home social training to children with ASD, was developed to mitigate the high costs and shortage of traditional interventions. To play the game, the child interprets and acts out prompts that are displayed on the screen while the caregiver is tasked with guessing the prompt. Multiple decks and prizes tailor to the child's preferences and help increase engagement.

Guess What? incorporates two teaching methods based on ABA principles: Discrete Trial Training (DTT) and Pivotal Response Treatment (PRT). DTT breaks down the skill into discrete trials that build up the skill step by step [10]. Each trial follows a specific set of steps consisting of an antecedent, prompt, response, reinforcement, and brief pause [10]. Figure 1 shows how *Guess What?* incorporates DDT in the gameplay. PRT is less structured and initiated by the child, emphasizing natural reinforcement and targeting pivotal areas of a child's development instead of specific behaviors [11]. Multiple studies suggest that DTT helps improve emotion recognition and expression [10] and PRT enhances communication skills in children with ASD [11].

Previous studies have demonstrated the promise of applying mobile systems with real-time emotion recognition to ABA therapy [12] [13] [14] [15] [16] [17] [18]. Integrating an automatic emotion classifier into *Guess What?* will provide supplemental reinforcement to the caregiver and allow for the development of additional features integral to ABA therapy: adapting prompts and difficulty to target the child's specific deficits and offering appropriate visual cues to assist the child [3].

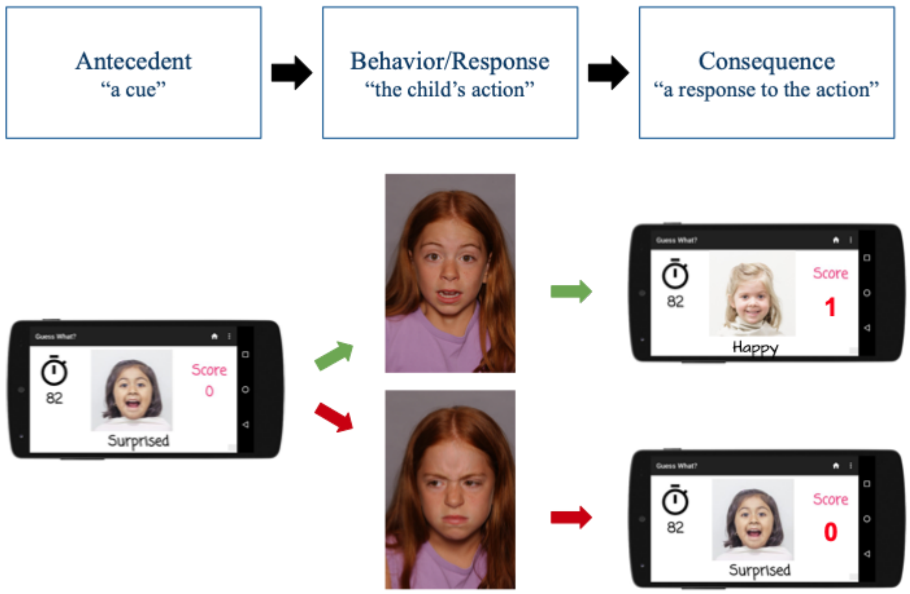


Figure 1: *Guess What*? incorporates DTT, as outlined in the figure, and Pivotal Response Treatment (PRT), two teaching strategies that fall under the umbrella of ABA.

However, existing emotion recognition platforms are not optimized for research on children [19] [20] as a result of being trained on datasets in which pediatric populations are highly underrepresented, such as the CIFAR-100, ImageNet [21], Cohn-Kanade Database [22] and Belfast-Induced Natural Emotion Databases [23]. *Guess What?* can serve as a data acquisition tool and aggregate emotive videos for autism research that can be used to train a more effective automatic emotion recognition platform. This motivates the development of a new emotion classifier designed specifically for pediatric populations, trained with images crowdsourced from *Guess What?*.

Methods

Game Design

To play *Guess What?* [7] [8] [9], the child begins by selecting one of the following themed decks to play: animals, emoji, faces, gestures, jobs, objects, sports, chores, and a special deck created for toddlers. The caregiver will hold the device outwards with the screen facing the child. During the 90-second game session, the child acts out the prompt displayed on the screen while the caregiver guesses. If the caregiver's guess is correct, the child prompts the caregiver, who tilts the phone. The game then rewards the child with a point, resulting in another image appearing on the screen. This process repeats for a fixed amount of time. The entire game session is recorded using the front-facing camera on the device, focusing on the child's actions. If the user grants permission to share this footage, the video is uploaded to a secure and encrypted Amazon Web Services S3 bucket. This data upload and storage process is fully compliant with the Stanford University's High-Risk Application security standards. Additional metadata included with the video includes the prompts used in the session, timing logs, and the number of points awarded. *Guess What?* is available for both Android and iOS platforms.

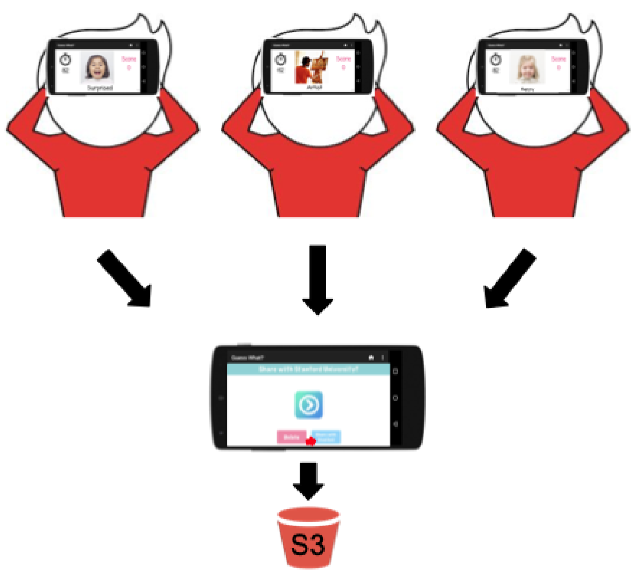


Figure 2. Crowdsourced videos taken during game sessions are stored in an Amazon S3 bucket (with participant's consent).

Data Collection

Figure 2 illustrates the data aggregation process. The two decks that are the most closely associated with emotion recognition and expression are the *emoji* and *faces* decks. These decks contain emoticons (cartoon representations of facial emotions) and real images of children expressing various emotions, respectively. Using crowdsourced videos from fifteen subjects remotely playing these two decks subsampled at 5 frames per second (FPS), a dataset consisting of 6,344 frames was built.

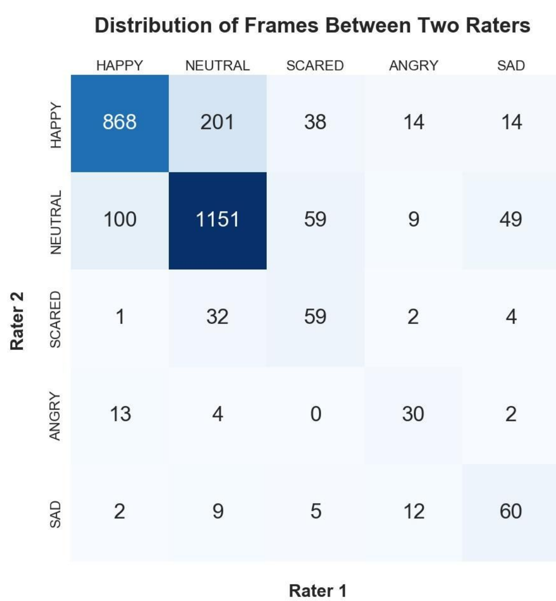


Figure 3. Confusion matrix of the two raters' emotion labels.

Data Processing

To establish ground truth, two raters manually labeled each of the 6,344 frames with either one of four Ekman universal emotions (happy, sad, scared, angry) [24] or a neutral label. In cases where there were no faces in the frame or the face was too blurry to discern, the raters labeled the frame with "n/a." To filter the data, all frames with rater disagreement or labeled as "n/a" were discarded. Faces were then extracted from the remaining frames using the OpenCV library [25] and yielded 757 frames. Figure 3 is a confusion matrix illustrating the distribution of the raters' labels. The Cohen's Kappa statistic for inter-rater reliability [26], a metric which accounts for agreements due to chance, was 0.8, indicating a high level of reliability between the two raters.

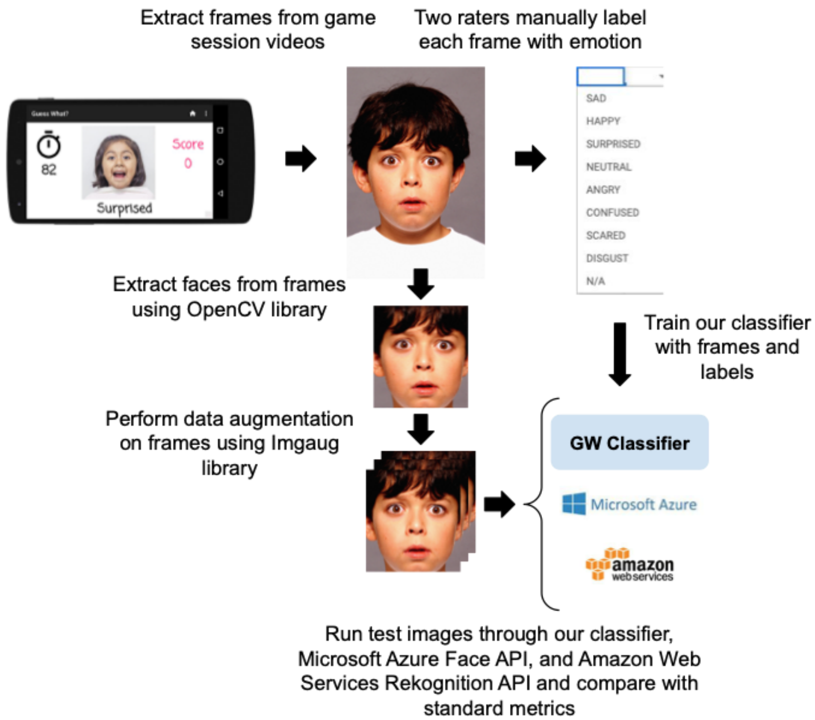


Figure 4. The classifier's training procedure.

Classifier Training

The proposed emotion classifier was trained using Google Cloud's AutoML pipeline, which leverages Google's transfer learning and neural architecture search technologies to automate the determination of the strongest network architecture and optimal hyperparameter configurations to minimize the loss functions [27] [28]. Due to an uneven distribution of emotions, data augmentation methods from the Imgaug library [29] were performed to increase the number of viable frames to 989, with roughly 200 corresponding to each of the five emotions. The specific methods performed included horizontally flipping the image, cropping the image, blurring the image, improving or worsening contrast, adding Gaussian noise to the image, brightening or darkening the image, and applying affine transformations to the image, all performed in random order and of varied magnitudes [29]. 861 frames were used for training and validation, and the remaining 128 frames were used for testing. Figure 4 illustrates the data processing and classifier training procedure.

Data Analysis

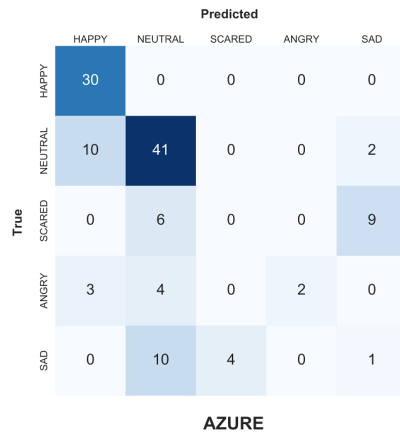
To evaluate the performance of the models, the F1 score was calculated as follows:

$$P = \frac{t_p}{t_p + f_p} \qquad R = \frac{t_p}{t_p + f_n} \qquad F1 = 2 \times \frac{P \times R}{P + R}$$

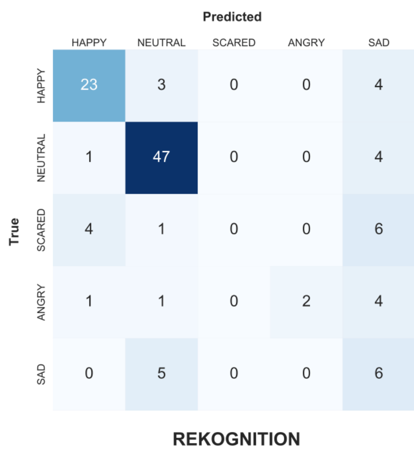
P stands for precision, R stands for recall, F1 stands for F1 score, t_p stands for true positive, f_p stands for false positive, and f_n stands for false negative.

Results and Discussion

a.



b.



c.

Figure 5. Confusion matrices of the proposed classifier, Azure, and Rekognition.

The proposed classifier was compared to two existing emotion recognition platforms: Microsoft's Azure Face API [30] and Amazon Web Services' Rekognition [31]. The same 128 frames that were tested on the proposed classifier were tested on these two classifiers. Figure 5b and 5c show that these classifiers can recognize happy and neutral but perform poorly on angry, sad, and scared classes. These results suggest the need for a new classifier that demonstrates stronger performance across all emotions for pediatric populations.

The performance of the proposed classifier is illustrated in Figure 5a. Figure 5a shows that the most discrepancies occurred between differentiating neutral from happy, which contrasts with the performance of the other classifiers. However, this proposed classifier generally showed very strong performance for all five emotions, especially when compared to the other two existing classifiers.

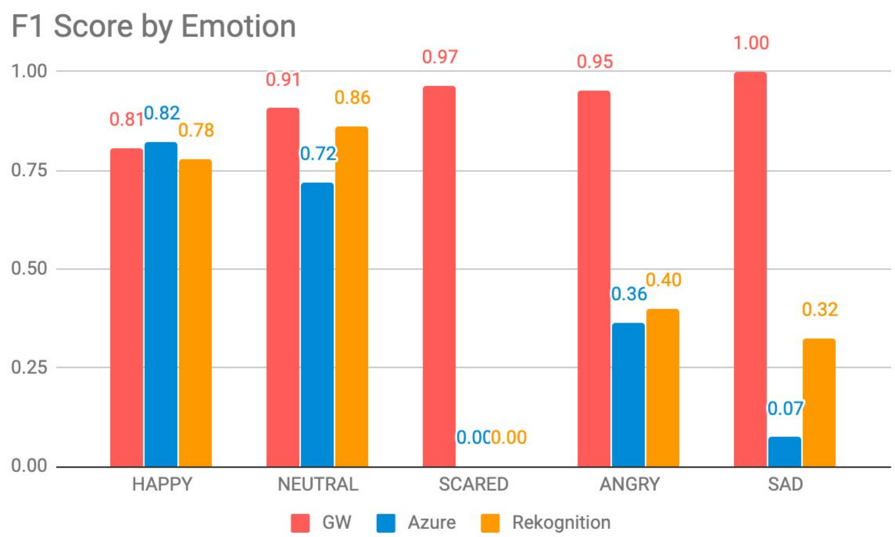


Figure 6. F1 score by emotion by classifier

Figure 6 shows the F1 scores [32] of each classifier separated by emotion. The F1 score is the harmonic mean of the precision and recall. A score of 1 represents perfect precision and recall while a score of 0 represents extremely low precision and recall. The proposed classifier displayed the highest F1 scores across all but one emotion when compared to both existing commercial emotion recognition platforms. The one deviation occurred when the Azure classifier had an F1 score of 0.82 for happy, while the proposed classifier had a comparable F1 score of 0.81. However, the Azure classifier had the lowest F1 scores for all of the other classes.

Conclusion and Future Directions

Both the Azure and Rekognition classifiers had a comparable F1 score with the proposed classifier on happy and neutral frames but failed with other emotions. In addition, the proposed classifier had an F1 score above 0.95 for scared, angry, and sad frames, while the two existing classifiers demonstrated stronger performance on happy and neutral frames. Because heavy data augmentation procedures had to be performed on the scared, angry, and sad classes to evenly distribute the training set, overfitting may have occurred with the proposed classifier. Generating a more diverse dataset of frames to begin with may alleviate this issue and will be addressed in future work. In addition, the proposed classifier will be improved to perform accurately on all emotions: the five emotions addressed in this study as well as disgust and surprise which are two other Ekman emotions. Furthermore, this emotion classifier generalized to children with ASD will be integrated into *Guess What?* to provide supplemental reinforcement and allow for the development of new features, including adapting the game to target specific deficits and providing appropriate guiding feedback, which will enable the development of at-home precision therapy for children with ASD.

Acknowledgements

These studies were supported by awards by the National Institutes of Health (1R21HD091500-01 and 1R01EB025025-01). Additionally, we acknowledge the support of grants from The Hartwell Foundation, the David and Lucile Packard Foundation Special Projects Grant, Beckman Center for Molecular and Genetic Medicine, Coulter Endowment Translational Research Grant, Berry Fellowship, Spectrum Pilot Program, Stanford's Precision Health and Integrated Diagnostics Center (PHIND), Wu Tsai Neurosciences Institute Neuroscience: Translate Program, and Stanford's Institute of Human Centered Artificial Intelligence as well as philanthropic support from Mr. Peter Sullivan. We would also like to acknowledge support from the Thrasher Research Fund and Stanford NLM Clinical Data Science program (T-15LM007033-35). Finally, we acknowledge the Stanford Institutes of Medicine Summer Research Program for funding CH.

References

[1] Kogan, M. D., Vladutiu, C. J., Schieve, L. A., Ghandour, R. M., Blumberg, S. J., Zablotsky, B., ... & Lu, M. C. (2018). The prevalence of parent-reported autism spectrum disorder among US children. *Pediatrics*, *142*(6), e20174161.

[2] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

[3] Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical psychology review*, *30*(4), 387-399.

[4] Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Development and psychopathology, 20(3), 775-803.

[5] Horlin, C., Falkmer, M., Parsons, R., Albrecht, M. A., & Falkmer, T. (2014). The cost of autism spectrum disorders. *PloS one*, *9*(9), e106552.

[6] Ning, M., Daniels, J., Schwartz, J., Dunlap, K., Washington, P., Kalantarian, H., ... & Wall, D. P. (2019). Identification and Quantification of Gaps in Access to Autism Resources in the United States: An Infodemiological Study. *Journal of medical Internet research*, *21*(7), e13094.

[7] Kalantarian, H., Washington, P., Schwartz, J., Daniels, J., Haber, N., & Wall, D. P. (2019). Guess What?. *Journal of Healthcare Informatics Research*, *3*(1), 43-66.

[8] Kalantarian, H., Washington, P., Schwartz, J., Daniels, J., Haber, N., & Wall, D. (2018, June). A Gamified Mobile System for Crowdsourcing Video for Autism Research. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 350-352). IEEE.

[9] Kalantarian, H., Jedoui, K., Washington, P., & Wall, D. P. (2018). A Mobile Game for Automatic Emotion-Labeling of Images. *IEEE Transactions on Games*.

[10] Sigafoos, J., Carnett, A., O'Reilly, M. F., & Lancioni, G. E. (2019). Discrete trial training: A structured learning approach for children with ASD.

[11] Koegel, R. L., & Koegel, L. K. (2006). *Pivotal response treatments for autism: Communication, social, & academic development*. Paul H Brookes Publishing.

[12] Voss, C., Schwartz, J., Daniels, J., Kline, A., Haber, N., Washington, P., ... & Feinstein, C. (2019). Effect of Wearable Digital Intervention for Improving Socialization in Children With Autism Spectrum Disorder: A Randomized Clinical Trial. *JAMA pediatrics*, *173*(5), 446-454.

[13] Daniels, J., Schwartz, J. N., Voss, C., Haber, N., Fazel, A., Kline, A., ... & Wall, D. P. (2018). Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism. *npj Digital Medicine*, *1*(1), 32.

[14] Daniels, J., Schwartz, J., Haber, N., Voss, C., Kline, A., Fazel, A., ... & Wall, D. (2017). 5.13 design and efficacy of a wearable device for social affective learning in children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(10), S257.

[15] Washington, P., Voss, C., Kline, A., Haber, N., Daniels, J., Fazel, A., De, T., Feinstein, C., Winograd, T., & Wall, D.P. (2017). SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. *IMWUT, 1*, 112:1-112:22.

[16] Voss, C., Washington, P., Haber, N., Kline, A., Daniels, J., Fazel, A., ... & Wall, D. (2016, September). Superpower glass: delivering unobtrusive real-time social cues in wearable systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 1218-1226). ACM.

[17] Washington, P., Voss, C., Haber, N., Tanaka, S., Daniels, J., Feinstein, C., ... & Wall, D. (2016, May). A wearable social interaction aid for children with autism. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2348-2354). ACM.

[18] Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., ... & Avramides, K. (2012). Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing*, *16*(2), 117-127.

[19] Kalantarian, H., Jedoui, K., Washington, P., Tariq, Q., Dunlap, K., Schwartz, J., & Wall, D. P. (2019). Labeling images with facial emotion and the potential for pediatric healthcare. *Artificial intelligence in medicine*, *98*, 77-86.

[20] Kalantarian, H., Khaled J., Washington, P., … & Wall, D. (2018, December). The Limitations of Real-Time Emotion Recognition for Autism Research.

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[22] Cohn, J. F. (1999). Cohn-Kanade AU-coded facial expression database. *Pittsburgh University*.

[23] Douglas-Cowie, E., Cowie, R., & Schröder, M. (2000). A new emotion database: considerations, sources and scope. In *ISCA tutorial and research workshop (ITRW) on speech and emotion*.

[24] Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ... & Scherer, K. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, *53*(4), 712.

[25] Wagner, P. (2012). Face recognition with opencv. *Order A J. Theory Ordered Sets Its Appl*, 1-26.

[26] McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276-282.

[27] Li, F. F., & Li, J. (2018). Cloud AutoML: Making AI accessible to every business. *Internet: https://www. blog. google/topics/google-cloud/cloud-automl-making-ai-accessible-everybusiness*.

[28] Wong, C., Houlsby, N., Lu, Y., & Gesmundo, A. (2018). Transfer learning with neural automl. In *Advances in Neural Information Processing Systems* (pp. 8356-8365).

[29] Jung, A. (2017). Imgaug: a library for image augmentation in machine learning experiments.

[30] Face API - Facial Recognition Software: Microsoft Azure. (n.d.). Retrieved from https://azure.microsoft.com/en-us/services/cognitive-services/face/

[31] Amazon Rekognition. (n.d.). Retrieved from https://aws.amazon.com/rekognition/

[32] Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval* (pp. 345-359). Springer, Berlin, Heidelberg.