

Pain Action Unit Detection in Critically Ill Patients

Subhash Nerella
Department of Biomedical
Engineering
University of Florida
Gainesville, USA
subhashnerella@ufl.edu

Julie Cupka
Department of Medicine
University of Florida
Gainesville, USA
jcupuka@ufl.edu

Matthew Ruppert
Department of Medicine
University of Florida
Gainesville, USA
Matthew.Ruppert@Medicine.ufl.edu

Patrick Tighe
Department of Anesthesiology
University of Florida
Gainesville, USA
ptighe@ufl.edu

Azra Bihorac
Department of Medicine
University of Florida
Gainesville, USA
abihorac@ufl.edu

Parisa Rashidi
Department of Biomedical
Engineering
University of Florida
Gainesville, USA
parisa.rashidi@ufl.edu

Abstract—Existing pain assessment methods in the intensive care unit rely on patient self-report or visual observation by nurses. Patient self-report is subjective and can suffer from poor recall. In the case of non-verbal patients, behavioral pain assessment methods provide limited granularity, are subjective, and put additional burden on already overworked staff. Previous studies have shown the feasibility of autonomous pain expression assessment by detecting Facial Action Units (AUs). However, previous approaches for detecting facial pain AUs are historically limited to controlled environments. In this study, for the first time, we collected and annotated a pain-related AU dataset, *Pain-ICU*, containing 55,085 images from critically ill adult patients. We evaluated the performance of OpenFace, an open-source facial behavior analysis tool, and the trained AU R-CNN model on our *Pain-ICU* dataset. Variables such as assisted breathing devices, environmental lighting, and patient orientation with respect to the camera make AU detection harder than with controlled settings. Although OpenFace has shown state-of-the-art results in general purpose AU detection tasks, it could not accurately detect AUs in our *Pain-ICU* dataset (F1-score 0.42). To address this problem, we trained the AU R-CNN model on our *Pain-ICU* dataset, resulting in a satisfactory average F1-score 0.77. In this study, we show the feasibility of detecting facial pain AUs in uncontrolled ICU settings.

Keywords— *Pain; Facial Action Units; Facial Landmarks; OpenFace; AU R-CNN*

I. INTRODUCTION

A widely accepted definition of pain is “an unpleasant sensory and emotional experience associated with actual or potential tissue damage” [1]. Untreated pain can potentially result in multiple deleterious complications, such as prolonged mechanical ventilation, longer Intensive Care Unit (ICU) stays, chronic pain after discharge, and increased mortality risk [2]. A 2013 study reports that the incidence of significant pain among patients in the ICU is 50% or higher [3]. About 17% of patients who were discharged from the ICU recollect

the experience of severe pain they have undergone even after 6 months [4]. The primary barrier to adequate treatment is under-assessment of pain in critically ill patients. Pain assessment is necessary in the ICU in order to manage opioid doses and for the overall recovery process. Characteristics of painful experiences are specific to each individual, making it a difficult variable to reliably quantify. ICU staff relies on self-reported pain by individual patients, which is the gold standard method for pain assessment. The Visual Analog Scale [5, 6] and Numeric Rating Scale [7] are among commonly used self-report pain scales, but nonetheless are biased and subjective to the individual. Many critically ill patients might not be able to self-report pain due to various reasons, including the use of ventilators, the influence of sedatives, or being in an altered mental state. Therefore, the ICU staff have to rely on observational and behavioral pain assessment tools for nonverbal patients, such as the Nonverbal Pain Scale (NVPS) and Behavioral Pain Scale (BPS) [8]. Even still, observer-based pain intensity measurement methods are subjective, error prone, and lack granularity.

The human face plays a prominent role in nonverbal communication [9, 10]. Patient facial expressions can help identify the presence of pain in the case of individuals who cannot self-report their pain status. NVPS and BPS include facial behavior to evaluate the pain in non-verbal patients. The Facial Action Coding System (FACS) is a facial anatomy-based action coding system that can capture instant changes in facial expression through processing of different facial AUs [11]. Facial AUs are defined based on the relative contraction or relaxation of individual or grouped facial muscles, therefore any facial expression can be represented as a combination of these facial AUs. Prkachin and Solomon identified certain facial action units associated with pain and developed the Prkachin Solomon Pain Index (PSPI) score based on the FACS [12]. The PSPI score considers action units AU-4 (Brow Lowerer), AU-6 (Cheek Raiser), AU-7 (Lid Tightener), AU-9 (Nose Wrinkler), AU-10 (Upper Lip Raiser), and AU-

43 (Eyes Closed), by accounting for both the presence of the AUs as well as their intensity on a scale of (0-5). The combination of AU presence and the corresponding intensity yields a 16-point pain scale. A major caveat of this approach is that the AUs must be coded manually by trained personnel, requiring time-consuming and costly training practices and making it a clinically unviable option.

An autonomous, real-time pain detection system has the potential to facilitate clinical workflow. The identification of AUs is an essential step in the development of an autonomous pain expression assessment system. A major factor hindering the development of an autonomous pain detection system is the lack of an annotated dataset in uncontrolled settings. Previously, researchers at McMaster University and the University of Northern British Columbia (UNBC) captured videos of 25 participants' face suffering from shoulder pain with varied diagnoses [13]. Each video frame was coded by certified FACS coders and the data was captured in a controlled laboratory setting. More recent datasets like BP4D-Spontaneous [14] and BP4D+ [15] are available with videos recorded from healthy volunteers, where facial AUs were manually annotated and pain was stimulated by the cold pressor task. To date, all existing datasets on pain expressions have been captured in controlled or semi-controlled settings. To our knowledge, this is the first effort to identify pain related AUs in a dynamic clinical setting.

In this study, we collected the *Pain-ICU* dataset, a facial pain AU dataset compiled from critically ill adult patients. Each image frame was annotated by three trained annotators. We first evaluated the performance of OpenFace [16], a leading open-source facial behavior analysis tool, to detect related AUs. We also used the AU R-CNN [17] model which showed state of the art performance on general purpose action unit (GP-AU) detection for BP4D-Spontaneous and DISFA [18] datasets. Although OpenFace has shown state of the art results on AU detection in other settings, in our dynamic ICU AU detection setting (ICU-AU), failed to generalize well. This result shows the need for developing customized tools for such settings. To address this issue, we trained the AU R-CNN model on the *Pain-ICU* dataset and outperformed OpenFace's performance. In this study, we show the requirements of a system trained on ICU data to detect facial action units for the objective of pain assessment in a clinical setting.

II. MATERIALS AND METHODS

A. Data Collection

All the data used in this study were collected in two surgical ICUs at the University of Florida Health Shands Hospital, Gainesville, Florida. The study was reviewed and approved by the University of Florida Institutional Review Board (UFIRB). All methods were performed, data were collected, and written informed consents were obtained by following all guidelines and regulations of the UFIRB. We complied with the guidelines to ensure correct application of deep learning predictive models within clinical setting [19]. Table I shows the patient cohort characteristics of this study. To collect video data from the ICU, we used a standalone

system with a camera focused on the patient's face. Fifteen-minutes of video was extracted within the proximity of patient-reported Defense and Veterans Pain Rating Scale (DVPRS) pain scores obtained from electronic health records. These videos were further processed to obtain individual frames at 1 frame per second (fps).

TABLE I. PATIENT CHARACTERISTICS TABLE

Participants (n = 10, number of images = 151,947)	
Age, median (IQR)	73.3 (54.7,74.5)
Female, number (%)	3 (33)
Race, number (%)	White, 9 (90) African American, 1 (10)
BMI, median (IQR)	24.4 (22.7,26.2)
Hospital stay in days, median (IQR)	19 (18,22)

Three annotators were trained on the FACS training manual. Annotators were evaluated on a separate training dataset before annotating ICU image frames. All of the annotations were performed individually on the assigned images. Only the images annotated by all three annotators were considered for evaluation. Furthermore, ground truth labels chosen where at least two out of three annotators agreed about the presence of a particular AU were used in our evaluations. In Table II, each AU is listed with its corresponding description.

TABLE II: LIST OF ALL ACTION UNITS WHICH ARE EITHER DETECTED BY OPENFACE TOOL OR AU R-CNN

AU	Description	OpenFace	AU R-CNN
AU25	Lips Part	✓	✓
AU26	Jaw Drop	✓	✓
AU43	Eyes Closed		✓

B. Methodology

Data preparation involved extraction of individual image frames from the videos, face detection in the image frames, face cropping, and annotation. We used FFmpeg [20] an opensource multimedia processing tool to extract individual frames from the videos. Following image extraction, frames were processed using the Facenet [21] tool to detect and crop patient faces from the input frames. Then, the extracted images were provided as input to the OpenFace and AU R-CNN models. All authors stand blinded on the annotation task; only annotators performed the AU annotation assignments.

OpenFace is a toolkit available for facial behavior analysis using computer vision algorithms [16]. The AU R-CNN model is a convolutional neural network for recognition of facial action units [17]. Although the objective of both models is the same (facial AU detection), they each have a distinct

approach to the problem. In the pipeline of AU identification, the OpenFace tool detects facial landmarks which are further used in facial alignment, feature fusion, and ultimately facial AU detection. OpenFace uses linear Support Vector Machines (SVM) kernel combined with Histograms of Gradients (HOGs) for action unit detection. Most approaches use facial landmarks to align the images to further detect facial AUs.

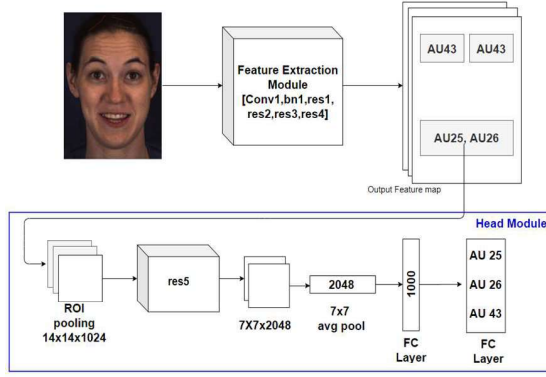


Fig. 1. Pipeline of AU R-CNN architecture showing the feature extraction module, ROI pooling layer and head module with sigmoid activation to detect the action units in the input face image.

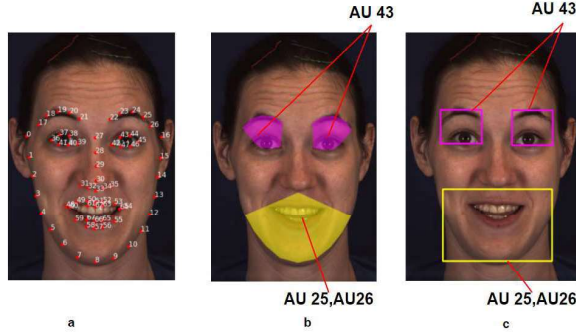


Fig 2: (a) 68 landmark locations obtained by using OpenFace landmark detection tool. (b) We obtain the AU masks i.e., regions where an AU can be localized for AUs 25, 26, and 46 using the landmarks. AU 25 and 26 share the same region of Interest. AU-43 can occur in two regions. (c) Minimum enclosing rectangular region of interest (ROI) box is obtained to provide as input to ROI pooling layer. Note: Although the ROI box are shown on face image here, these regions are obtained from the output feature map of Feature extraction module

Table II shows the AUs detected by OpenFace and trained AU R-CNN model. We evaluated the performance of these models on AU-25, AU-26, and AU-43. OpenFace implementation does not include detection of AU-43, so we report the performance of OpenFace on AU-25 and AU-26. Although OpenFace can detect other AUs, we were only interested in the AUs that are associated with pain expression. We considered AUs that are common between our *Pain-ICU* dataset and AUs OpenFace can detect for performance comparison of both models.

AU R-CNN proposed by Chen et al is an adaptive regional learning approach to detect the presence of AUs in

an image. The network architecture proposed shown in Fig. 1 consists of two modules, feature extraction module, and the head module. The feature extraction module consists of *res1*, *res2*, *res3*, *res4* blocks of ResNet-101 architecture [22]. The head module comprises of *ROI pooling layer*, *res5*, *average pool*, and *fully connected* layers. Feature extraction module takes an individual face image as input and output the feature map shown in Fig. 1. Corresponding AU mask regions shown in Fig. 2c are obtained from the output of feature extraction module and unrelated regions of the image are discarded. The ROI pooling layer introduced by Fast R-CNN [23] is used to convert the given rectangular region of interest to a feature map with a predefined fixed spatial extent (14X14). The last fully connected layer in head module is modified to have a size equal to the number of AUs being detected. The network uses the sigmoid cross entropy loss, as multiple AUs can be present in given region of interest (a multi-label problem).

Regions of Interest (ROIs) corresponding to AUs are obtained based on their location of occurrence on the face. Fig. 2 shows how ROIs are extracted from an input face image. We used OpenFace predicted landmark coordinates to obtain AU masks and the corresponding minimum enclosing rectangles shown in Fig 2a, Fig. 2b, and Fig. 2c, respectively. Two ROIs each corresponding individual eye regions were extracted for AU-43 (Eyes closed). Similarly, AU-25 (Lips part) and AU-26 (Jaw drop) both occur around the mouth region, therefore ROI around the mouth region is extracted. Although Fig. 2 shows the ROIs on the face image, ROIs are extracted from output feature map of feature extraction module. The minimum bounding rectangle coordinates and the feature map are usually 16x smaller than the input image resolution. Some of the AUs share a common region of interest similar to AU-25 and AU-26. Our study is limited to three AUs, so we only used three ROIs (2 ROIs for AU-43 and 1 ROI for both AU-25 and AU-26) corresponding to the AUs.

III. RESULTS

A. Experimental Setting

The *Pain-ICU* dataset contains 151,947 image frames from 10 ICU patients. We considered 55,085 images which are marked as clear by all three annotators. Most discarded images were discarded as no patient was visible in the image frame, or due to lack of lighting, patient face occluded by a blanket, or poor resolution. All remaining 55,085 images were evaluated for AU detection. We evaluated both OpenFace and AU R-CNN on the *Pain-ICU* dataset. We ran the OpenFace model on the entire dataset to evaluate its performance on AU detection and obtain the facial landmarks. Although OpenFace can detect multiple AUs, in this study we reported its performance on AU-25, and AU-26. OpenFace performance on the entire dataset is shown in Table III. AU R-CNN has achieved state of the art results on general purpose AU detection, but AU-25, AU-26 and AU-43 were not included in their published work. We trained the AU R-CNN model on our *Pain-ICU* dataset which consists of AU-25, AU-26, and AU-43. Data from 10 patients was randomly split into

three subsets containing 3, 3, and 4 patients, strictly ensuring same patient's data is exclusive to one of the subsets. The performance of trained AU R-CNN is reported in Table IV. The reported performance is the mean of three runs where every subset is considered as a test set and trained on the rest of the patients' data. AU R-CNN model was trained on two NVIDIA 2080 TI GPUs for 20 epochs with a mini-batch size of 12 images. We used RMSprop optimizer [20] as a gradient descent optimization algorithm with an initial learning rate 0.001 and sigmoid cross entropy as a loss function.

B. Evaluation Metrics

We used precision, recall, accuracy, and F1-score to evaluate the performance of both OpenFace and AU R-CNN models. Every AU is individually evaluated and reported. True Positive (TP) for an AU denotes if the AU is present in the image and is correctly identified as present by the model. True Negative (TN) denotes the AU is absent in the image frame and is correctly identified as not present in the image. False Positive (FP) denotes AU is incorrectly identified as present in the image when it is absent. False Negative (FN) denotes AU present in the image but identified as absent. In the results we reported the number of image frames in which a given AU is present in the support column of tables III – VII. AU R-CNN generates ROI level predictions, to obtain image level prediction from ROI level predictions we used “bit-wise OR” to merge the predictions. Once the image level prediction is obtained, we evaluate the model prediction against the ground truth labels.

TABLE III. OPENFACE PERFORMANCE ON ENTIRE ICU DATA REPORTED FOR EACH ACTION UNIT. PERFORMANCE IS EVALUATED AGAINST GROUND TRUTH ANNOTATION.

Action Unit	f1-score	precision	recall	accuracy	support
AU25	0.49	0.69	0.38	0.51	32583
AU26	0.34	0.61	0.24	0.58	23827
Average	0.42	0.65	0.31	0.55	

TABLE IV. TRAINED AU R-CNN PERFORMANCE REPORTED FOR EACH ACTION UNIT. AU R-CNN PERFORMANCE IS EVALUATED AGAINST GROUND TRUTH ANNOTATION. REPORTED METRICS ARE MEAN PERFORMANCE ON ALL THREE BATCHES.

Action Unit	f1-score	precision	recall	accuracy
AU 25	0.75	0.75	0.75	0.74
AU 26	0.71	0.87	0.62	0.80
AU 43	0.85	0.76	0.96	0.77
Mean	0.77	0.79	0.78	0.77

TABLE V. TRAINED AU R-CNN PERFORMANCE ON DATA FROM WHITE MALE PATIENTS FOR EACH ACTION UNIT.

Action Unit	f1-score	precision	recall	accuracy	support
AU 25	0.87	0.88	0.86	0.81	27074
AU 26	0.74	0.9	0.63	0.74	21331
AU 43	0.89	0.82	0.97	0.83	27594
Mean	0.83	0.87	0.82	0.79	

TABLE VI. TRAINED AU R-CNN PERFORMANCE ON DATA FROM WHITE FEMALE PATIENTS FOR EACH ACTION UNIT.

Action Unit	f1-score	precision	recall	accuracy	support
AU 25	0.24	0.2	0.34	0.57	1603
AU 26	0.28	0.8	0.19	0.95	545
AU 43	0.81	0.71	0.93	0.69	5463
Mean	0.44	0.57	0.49	0.74	

TABLE VII. TRAINED AU R-CNN PERFORMANCE ON DATA FROM AFRICAN AMERICAN PATIENTS FOR EACH ACTION UNIT.

Action Unit	f1-score	precision	recall	accuracy	support
AU 25	0.47	0.59	0.4	0.67	1869
AU 26	0.43	0.58	0.35	0.84	881
AU 43	0.67	0.5	1	0.51	2423
Mean	0.52	0.56	0.58	0.67	

IV. DISCUSSION

OpenFace can detect facial landmarks, head pose, eye-gaze, and facial action unit recognition. In this study, we were only interested in pain-related facial action unit detection, as these units play a key role in pain expression recognition. We reported the performance of OpenFace on ICU data in Table III with respect to AUs on all images we used in this study. Table III does not include AU43 as it is not present in AUs that can be detected by OpenFace. OpenFace struggled to capture AU-25 (Lips Part) and AU-26 (Jaw Drop) based on the F1-score. In the pipeline of OpenFace, detecting facial landmarks is one of the initial steps. Accurate detection of these facial landmarks plays a crucial role in the performance of the tool to detect AUs. OpenFace uses a combination of predicted landmarks and Histograms of the gradients to predict the AUs. Inaccurate prediction of these landmarks could result in poor performance of the model in prediction of action unit occurrence. The presence of medical devices

and inconsistent face orientation affected the prediction of landmarks. As both AU-25 (lips part) and AU-26 (Jaw drop) both depend on mouth region. Inaccurate prediction of landmarks near the mouth region where medical devices most likely present resulted in poor performance of OpenFace on our Pain-ICU dataset. Openface is trained on the datasets in which data is collected from the controlled lab environment. Controlled environment images have proper light intensity, close camera location with respect to face, and no occlusion of the face. Real world ICU environment cannot ensure the ideal conditions of a controlled environment. All the above factors effected the performance of OpenFace on the ICU data.

Table IV shows the performance of trained AU R-CNN model on the ICU data test set. Reported metrics are mean values of AU R-CNN performance on all the three batches. The model performed well in the case of all AUs AU-25, AU-26, and AU-43. This trained model showed better performance compared to OpenFace on AU-25 and AU-26. One reason for better detection of these AUs is due to their strong presence in the *Pain-ICU* dataset in terms of number of supporting images, which aided the network during training. The F1-score of AU-25, AU-26 and AU-43 exceeded the accuracy, which shows that model overestimated the presence of these AUs. This trained model showed how training the models can improve detection of AUs on ICU images. Although facial landmarks are used by both OpenFace and AU R-CNN, in the case of AU R-CNN landmarks are only used identify the region that should be provided as input to the head module, therefore highly accurate landmark detection was not essential for successful identification of AUs. AU R-CNN is not vulnerable to slight errors in the landmarks at least for the AUs we reported in this study. The bounding box extracted from Feature extraction module shown in Fig. 1 will likely contain the region of AU occurrence even in the case of mildly inaccurate prediction of landmarks.

Racial and gender bias in Artificial Intelligence (AI) is an important topic many researchers trying to address and eliminate. Buolamwini et al, have shown gender and racial bias in commercial AI facial analysis software in their work [24]. Although a small sample, we reported the trained AU R-CNN performance stratified by race and gender. Table V, table VI, and table VII shows performance of AU R-CNN on white male patients, white female patients, and African American male patients, respectively. Our *Pain-ICU* dataset has lower representation of female patients (3) and African American (1) patients. Our trained model is not successful in detecting AUs in females and African American patients especially in the case of AUs 25 and 26. Although we have strong presence of AUs in our Pain-ICU dataset, this presence is mainly based on data from white male patients, therefore trained AU R-CNN could not capture the AUs accurately on female and African American patients. To obtain an unbiased result, the dataset should have adequate sample size with sufficient representation of minority and underrepresented groups.

	mild(0-4)	moderate(5-6)	severe(7-10)
AU43	0.56	0.59	0.54
AU26	0.4	0.28	0.19
AU25	0.54	0.46	0.5

Fig 3: AU presence vs patient reported pain score. Image frames recorded within 1 hour from the collected pain timestamp are only considered.

The ultimate objective of our AU detection task is to determine patient's facial pain expression and pain score based on detected AUs. While previous studies have shown correlation between AUs and pain scores, no study has examined the association between AUs and pain score in uncontrolled ICU settings. ICU staff collected patient self-reported pain score (DVPRS) on a scale of 0 to 10, which can be further classified into mild (0-4), moderate (5-6), and high (7-10) categories. We ran mixed effects models to find association between AU presence and the pain score. None of the AUs we reported in this had shown statistically significant association with the patient reported pain score. In Fig. 3 we show the presence of AUs in our dataset against the patient reported pain score. Each cell shows normalized score number of frames in which AU is present to number of frames available for a given pain score category. It can be seen AUs we considered in study were sensitive to pain but not specific to pain. PSPI is a defined metric to obtain pain score from the facial AU presence. PSPI score is calculate based on action units AU-4, AU-6, AU-7, AU-9, AU-10, and AU-43. It should also be noted that the non-zero PSPI score does not necessarily mean the person is in pain. As an example, it can be seen patient reported pain score 0 has substantial presence of AU-43. several action units share presence in different facial expressions like pain and happiness (AU-6), fear (AU-4), disgust (AU-10/AU-9). As PSPI is not fool proof we also intend to use other ground truth label, patient reported pain score.

There are few limitations of our dataset. To avoid disrupting routine care, cameras are placed at a specific, unobtrusive location and at a distance in the ICU room, therefore some images might have lower resolution. Another limitation is that our dataset only has a strong presence of AUs 25, 26, and 43 while other AUs are limited or none. Also, patients in the ICU are more likely to be administered medication(s) which may affect the intensity of an individual's facial expression, resulting in low or no presence of other AUs. In the future, we intend to collect additional data on several other AUs associated with pain expression. A model which can incorporate multiple modalities such as activity, physiological signals, and electromyography data can make autonomous nonverbal pain assessment more feasible.

We acknowledge the growing concerns of facial recognition due to privacy issues. Our goal is to create a model which can improve patient outcomes and reduce workload on overburdened ICU staff, while preventing any potential abuse of privacy or misuse. We consciously chose not to include any patient-identifying information such as sample images in this work. All patient data are safely secured on a private server, limiting data access to specific study staff members, and following all local, state, and federal patient privacy rules and regulations. All analyses were performed locally on the server, and only the de-identified, end numerical results were retrieved from the server. In the future, we intend to perform all analyses in an online learning mode, to avoid storing any data for training, further enhancing the privacy and security features of such a system.

V. CONCLUSION

Autonomous pain assessment can be helpful in the ICU in terms of both cost and efficiency. Pain assessment through facial expression requires the detection of facial action units (AU). In this paper, we evaluated the performance of OpenFace and AU R-CNN models on actual ICU data in detecting facial action units. ICU data are distinctly different from controlled environment data, resulting in poor performance by the OpenFace model. The model struggled to identify the AUs, although they achieved good performance on the controlled environment datasets. Factors including the presence of medical assist devices on patient's faces, lighting in the ICU, image quality, and patient face orientation resulted in poor performance of these models. AU R-CNN model trained on the *Pain-ICU* dataset was able to achieve better performance compared to OpenFace. Although our trained model showed overall better performance, the performance was lower for detecting AUs in female and African American patients. The performance of trained model clearly states, model trained predominantly on white male patients is not adequate to perform accurate AU detection on general patient population. We conclude that to achieve accurate results in uncontrolled ICU settings, models need to be trained on ICU-specific data with strong presence of action units, and adequate presence of data from underrepresented groups to achieve the end goal of pain assessment.

Acknowledgements

This study is supported by National Science Foundation CAREER award 1750192, 1R01EB029699 and 1R21EB027344 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), R01GM-110240 from the National Institute of General Medical Science (NIH/NIGMS), and 1R01NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS).

REFERENCES

- [1] H. Merskey, "Pain terms: a list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy," *Pain*, vol. 6, pp. 249-252, 1979.
- [2] F. Tennant, "Complications of uncontrolled, persistent pain," *Pract Pain Manag*, vol. 4, no. 1, pp. 11-14, 2004.
- [3] A. De Jong *et al.*, "Decreasing severe pain and serious adverse events while moving intensive care unit patients: a prospective interventional study (the NURSE-DO project)," *Critical care*, vol. 17, no. 2, p. R74, 2013.
- [4] C. Granja *et al.*, "Understanding posttraumatic stress disorder-related symptoms after critical care: the early illness amnesia hypothesis," *Critical care medicine*, vol. 36, no. 10, pp. 2801-2809, 2008.
- [5] P. E. Bijur, W. Silver, and E. J. Gallagher, "Reliability of the visual analog scale for measurement of acute pain," *Academic emergency medicine*, vol. 8, no. 12, pp. 1153-1157, 2001.
- [6] H. Breivika, "Fifty years on the Visual Analogue Scale (VAS) for pain-intensity is still good for acute pain. But multidimensional assessment is needed for chronic pain," *Scandinavian journal of pain*, vol. 11, no. 1, pp. 150-152, 2016.
- [7] C. L. Von Baeyer, L. J. Spagrud, J. C. McCormick, E. Choo, K. Neville, and M. A. Connelly, "Three new datasets supporting use of the Numerical Rating Scale (NRS-11) for children's self-reports of pain intensity," *PAIN®*, vol. 143, no. 3, pp. 223-227, 2009.
- [8] M. Odhner, D. Wegman, N. Freeland, A. Steinmetz, and G. L. Ingersoll, "Assessing pain control in nonverbal critically ill adults," *Dimensions of Critical Care Nursing*, vol. 22, no. 6, pp. 260-267, 2003.
- [9] F. De la Torre and J. Cohn, "Guide to Visual Analysis of Humans: Looking at People, chapter Facial Expression Analysis," *Springer*, vol. 9, pp. 31-33, 2011.
- [10] P. Ekman, W. V. Friesen, M. O'Sullivan, and K. Scherer, "Relative importance of face, body, and speech in judgments of personality and affect," *Journal of personality and social psychology*, vol. 38, no. 2, p. 270, 1980.
- [11] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system: The manual on CD ROM," *A Human Face, Salt Lake City*, pp. 77-254, 2002.
- [12] K. M. Prkachin and P. E. Solomon, "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain*, vol. 139, no. 2, pp. 267-274, 2008.
- [13] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The UNBC-McMaster shoulder pain expression archive database," in *Face and Gesture 2011*, 2011: IEEE, pp. 57-64.
- [14] X. Zhang *et al.*, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692-706, 2014.
- [15] Z. Zhang *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3438-3446.
- [16] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018: IEEE, pp. 59-66.
- [17] C. Ma, L. Chen, and J. Yong, "AU R-CNN: Encoding expert prior knowledge into R-CNN for action unit detection," *Neurocomputing*, vol. 355, pp. 35-47, 2019.
- [18] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151-160, 2013.
- [19] W. Luo *et al.*, "Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view," *Journal of medical Internet research*, vol. 18, no. 12, p. e323, 2016.
- [20] "FFmpeg Developers. (2016). ffmpeg tool (Version beld324) [Software].", ed.

- [21] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815-823.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [23] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
- [24] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*, 2018, pp. 77-91.