

Automated Assessment of Children's Postoperative Pain Using Computer Vision

Karan Sikka, MS^a, Alex A. Ahmed, BS^a, Damaris Diaz, MD^b, Matthew S. Goodwin, PhD^c, Kenneth D. Craig, PhD^d, Marian S. Bartlett, PhD^{a,e}, Jeannie S. Huang, MD, MPH^{b,f}

abstract

BACKGROUND: Current pain assessment methods in youth are suboptimal and vulnerable to bias and underrecognition of clinical pain. Facial expressions are a sensitive, specific biomarker of the presence and severity of pain, and computer vision (CV) and machine-learning (ML) techniques enable reliable, valid measurement of pain-related facial expressions from video. We developed and evaluated a CVML approach to measure pain-related facial expressions for automated pain assessment in youth.

METHODS: A CVML-based model for assessment of pediatric postoperative pain was developed from videos of 50 neurotypical youth 5 to 18 years old in both endogenous/ongoing and exogenous/transient pain conditions after laparoscopic appendectomy. Model accuracy was assessed for self-reported pain ratings in children and time since surgery, and compared with by-proxy parent and nurse estimates of observed pain in youth.

RESULTS: Model detection of pain versus no-pain demonstrated good-to-excellent accuracy (Area under the receiver operating characteristic curve 0.84–0.94) in both ongoing and transient pain conditions. Model detection of pain severity demonstrated moderate-to-strong correlations ($r = 0.65$ – 0.86 within; $r = 0.47$ – 0.61 across subjects) for both pain conditions. The model performed equivalently to nurses but not as well as parents in detecting pain versus no-pain conditions, but performed equivalently to parents in estimating pain severity. Nurses were more likely than the model to underestimate youth self-reported pain ratings. Demographic factors did not affect model performance.

CONCLUSIONS: CVML pain assessment models derived from automatic facial expression measurements demonstrated good-to-excellent accuracy in binary pain classifications, strong correlations with patient self-reported pain ratings, and parent-equivalent estimation of children's pain levels over typical pain trajectories in youth after appendectomy.



WHAT'S KNOWN ON THIS SUBJECT: Clinical pain assessment methods in youth are vulnerable to underestimation bias and underrecognition. Facial expressions are sensitive, specific biomarkers of the presence and severity of pain. Computer vision-based pattern recognition enables measurement of pain-related facial expressions from video.

WHAT THIS STUDY ADDS: This study demonstrates initial validity for developing computer vision algorithms for automated pain assessment in children. The system developed and tested in this study could provide standardized, continuous, and valid patient monitoring that is potentially scalable.

^aInstitute for Neural Computation, and ^bDepartment of Pediatrics, University of California San Diego, San Diego, California; ^cDepartment of Health Sciences, Northeastern University, Boston, Massachusetts; ^dDepartment of Psychology, University of British Columbia, Vancouver, British Columbia, Canada; ^eEmotient, Inc., San Diego, California; and ^fRady Children's Hospital, San Diego, California

Mr Sikka performed the machine learning under the guidance of Dr Bartlett, drafted the initial manuscript, and reviewed and revised the manuscript; Mr Ahmed carried out a portion of the initial analyses and reviewed and revised the manuscript; Dr Diaz performed data collection, performed a portion of the initial analyses, and reviewed and revised the manuscript; Drs Craig and Goodwin reviewed all analyses, and critically reviewed and revised the manuscript; Drs Bartlett and Huang conceptualized and designed the study, reviewed all analyses, and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

www.pediatrics.org/cgi/doi/10.1542/peds.2015-0029

DOI: 10.1542/peds.2015-0029

Accepted for publication Mar 30, 2015

Address correspondence to Jeannie S. Huang, MD, MPH, Department of Pediatrics, University of California San Diego, 9500 Gilman Dr, MC 0984, La Jolla, CA 92093. E-mail: jshuang@ucsd.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2015 by the American Academy of Pediatrics

Pain is one of the most common surgical complications,¹ with undertreatment associated with adverse outcomes.^{2,3} Published guidelines mandate adequate postoperative pain control for all, including children, to ensure safety and efficacy in pain management, and maintain maximum physical function, psychological well-being, and quality living.^{4,5}

Pain assessment is required for effective delivery of pain control but relies heavily on self-report of pain. However, self-report measures require cognitive, linguistic, and social competencies not available to many vulnerable populations, including infants, young children, and persons with communicative/neurologic impairments. In addition, self-report is vulnerable to bias,^{6,7} and human-based self-report pain assessment methods are sporadically applied, thereby lacking the time-sensitivity to alter treatment plans promptly when pain changes.

In children, pain assessment by-proxy is common, yet underrecognition and underestimation of pain is pervasive even when performed by trained professionals and parents.⁸⁻¹² Observational scales focusing on nonverbal behaviors in response to pain have been developed,¹³ but place substantial demands on clinician time, have notable variability in pain cue definitions,^{14,15} and are subject to observer bias.¹⁵⁻²² Similar to pain self-report, observational scales often lack the ability to meaningfully alter pain interventions in a time-sensitive fashion. Taken together, these limitations indicate a need for the development of more automated, standardized, continuous, minimally biased, and scalable pain measures.

Facial expressions represent a sensitive and specific biomarker of the presence and severity of pain.^{23,24} They can be assessed by using the Facial Action Coding System (FACS),²⁵ which measures facial

expressions by using 46 anatomically based component movements known as facial action units (AUs). Manual coding of FACS captures facial information reliably and accurately, but typically requires 1 to 3 hours to code each minute of video.²⁶

Computer vision (CV) and machine-learning (ML) techniques can automatically code facial expressions of pain²⁷⁻²⁹ and reliably and validly measure pain-related FACS AUs from video.²⁹ The Computer Expression Recognition Toolbox (CERT)²⁹ measures FACS AUs in real time. An ML model based on CERT was able to distinguish genuine from faked pain in adults statistically significantly better than human observers.³⁰

In the current study, we examined the application of CERT to automatically analyze facial expressions indicative of pain in children after appendectomy. We hypothesized that ML on facial measures from CERT would provide accurate, standardized pain monitoring as compared with self-reported and by-proxy methods of pain assessment in children.

Two major forms of postsurgical pain have been identified^{31,32}: (1) pain at rest, or endogenous pain associated with disease and injury, including surgery; and (2) movement-evoked pain or exogenous pain brought on or aggravated by pain-evoking maneuvers (eg, movement, clinical examination, or physiotherapy). Analgesia varies in its impact on endogenous versus exogenous pain, and experts advocate assessment of both pain types in postsurgical trials.³² Both were examined in this study.

METHODS

Participants

Fifty youth, 5 to 18 years old, who had undergone laparoscopic appendectomy within the past 24 hours were recruited from a pediatric tertiary care center. Exclusion criteria

included regular opiate use within the past 6 months, documented mental or neurologic deficits preventing study protocol compliance, and any facial anomaly that might alter CV facial expression analysis. Parents provided written informed consent and youth gave written assent. The local institutional review board approved research protocols.

Study participants were on average 12 (5, 17) (median [minimum, maximum]) years old and 54% were boys. Thirty-five were Hispanic, 9 non-Hispanic white, 5 Asian, and 1 Native American. Participants were hospitalized for 3 (1, 9) hospital days.

Experimental Design and Data Collection

Data Collection

Data collection occurred over 3 study visits: (1) within 24 hours after appendectomy; (2) 1 calendar day after the first visit (median [25%, 75%]) time lapse of 20 (19, 21) hours between visits 1 and 2; and (3) at a follow-up visit (median [25%, 75%]) 21 [17, 27] days postoperatively). At every study visit, facial video recordings and self-reported pain ratings by the participant and by-proxy pain ratings by parent (all visits) and nurse (visits 1 and 2) were collected. Demographic and clinical data also were collected.

Pain Video Recordings

Video recordings captured both endogenous (hereafter labeled ongoing) and exogenous (transient) pain experiences. Participants faced a Canon VIXIA-HF-G10 (Melville, NY) video camera in an upright position, and video was recorded at 30 frames per second at 1440 × 1080 pixel resolution. First, facial activity was recorded for 5 minutes as a measure of ongoing pain. Then, video recordings of facial activity were collected as representative transient pain samples when manual pressure was exerted at the surgical site for 2, 10-second periods (typical of a clinical examination, with each

press averaging 1.5 to 2.0 inches in depth).

Participant Pain Ratings and Proxy Estimates

Study participants rated pain during ongoing and transient pain samples by using an 11-point 0-to-10 Numerical Rating Scale (NRS), a reliable and valid pediatric pain assessment instrument¹³ used universally at the hospital where this study was performed. Participants were asked to point at the number representing their pain level on a card. Parent (visits 1–3) and inpatient nurses (visits 1–2) simultaneously estimated participants' pain by using the NRS, without knowledge of participants' pain ratings. At each pain-rating session, parents remained in the room to remove possible participant anxiety related to parental absence and to standardize context.

Analysis

Video Analysis

Videos were analyzed by CERT²⁹ to measure facial AUs. Fourteen pain-related AUs were selected for analysis (Table 1, Fig 1). Three statistics (mean, 75th percentile, and 25th percentile) for each AU were computed across each pain event probe (ongoing and transient) and comprised the input to ML models. Details are provided in the Supplemental Information.

ML Models for Estimating Pain From Facial Expressions

Two types of models were developed to estimate pain from the facial AU measures: binary pain classification (presence or absence of clinically significant pain) and pain-intensity estimation (pain as a continuous measure). The ML methods were based on regression, and are detailed in the Supplemental Information. For binary pain classification, child NRS ratings ≥ 4 were defined as trials containing clinically significant pain based on a common interpretation of

TABLE 1 List of AUs and Other Facial Information Extracted From Video Clips by CERT

Action Unit	Description
4	Brow lower
6	Cheek raiser (orbit tighten)
7	Lid tightener
9	Nose wrinkler
10	Upper lip raiser
12	Lip corner puller
25	Lips part
26	Jaw drop
27	Mouth stretch
43	Eye closure
Smile	Smile detector
Yaw	Rotation of the head left and right (measured in degrees)
Pitch	Rotation up and down (measured in degrees)
Roll	Rotation in-plane, as in "tilt" (measured in degrees)

Data from the CERT smile detector were included, because it detects a combination of lip corner pulls (AU12) and cheek raise (AU6), and is a more robust detector because of being trained on 2 orders of magnitude more data.

ratings,^{33–35} whereas pain ratings of 0 were defined as trials with no pain. Consistent with previous studies,^{27,36} trials with self-ratings of 1 to 3 were excluded from the binary pain classification analysis. For pain-intensity estimation, all videos were used, with ratings across the full NRS scale of 0 to 10.

Time since surgery provided an alternate, objective ground truth for pain. The objective model for binary pain classification was trained to differentiate visit 1 from visit 3. The

objective model for pain-intensity estimation was trained to predict the number of days since surgery.

Evaluation of CVML Pain Estimation Models

Performance of the CVML models was evaluated by using cross-validation,³⁷ a method to test performance on data not used to develop the model. Basically, each model was developed multiple times, each time leaving out a predetermined number of participants from the dataset. Each model was then tested by using participants' data not included to develop the model. Performance was aggregated over the multiple tests. Details are provided in the Supplemental Information. Area under the receiver operating characteristic curve (AUC) and Cohen's κ were used to measure the performance of the binary pain classification models. Pearson correlations (both within and across subjects) were computed to evaluate pain-intensity estimation models, and z-tests were performed to test differences between models.

Comparison of CVML Models With Proxy Pain Estimates

Performance of the CVML models was compared with human observers' (parents and nurses) ability to estimate pain. Because inpatient nurse estimates were not available for visit 3 after patient discharge, and

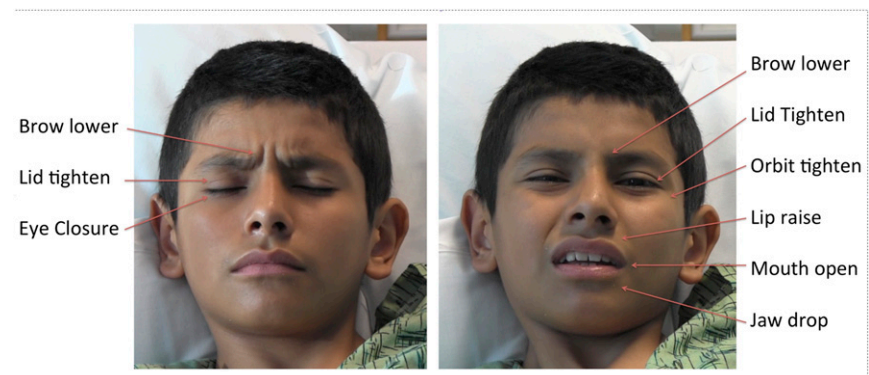


FIGURE 1

Examples of a child's facial expressions of pain from the study, illustrating many of the core facial actions observed in pain.

because most patients who received laparoscopic appendectomy are no longer in pain by the time of visit 3,^{38–40} nurse estimates for visit 3 were assigned a 0 rating. Median (minimum, maximum) pain ratings by children at the final visit were 0 (0, 0) for ongoing pain and 0 (0, 2) for transient pain.

For binary pain classification, nurse and parent estimates were compared with child ratings by using AUC and Cohen's κ . See Supplemental Information for details.

For pain-intensity estimation, Pearson correlations were computed between observer estimates and child ratings. Fisher z-tests were performed to compare overall correlations. Because parents and nurses were aware of elapsed time since surgery when assessing pain, the following 2 CVML pain-intensity estimation models were generated to compare the CVML performance with human observers: (1) CVML model trained only on facial AUs, and (2) CVML model trained on facial AUs and time since surgery. Pain-rating discrepancy (difference between observer estimate and child's self-rating) also was calculated as a measure of accuracy.

RESULTS

Pain Ratings

Pain-rating trajectories, Table 2, demonstrate known patterns of injury and healing associated with laparoscopic appendectomy^{38–40} (most severe immediately after surgery with pain resolution by follow-up).

Binary Classification of Clinically Significant Pain by CVML Models

Performance of the CVML binary pain classification models is shown in Table 3. AUC scores demonstrated good-to-excellent^{41,42} signal detection for both ongoing and transient pain, for the model trained to predict child ratings and for the

TABLE 2 NRS Child Pain Ratings, Machine Model Pain Level Estimates, and Proxy Estimates Over Time by Source and Condition

Source	Condition	Study Visit Number		
		1	2	3
Child	Ongoing	4.2 ± 0.33	2.7 ± 0.31	0
	Transient	4.86 ± 0.28	3.61 ± 0.22	0.08 ± 0.03
Machine	Ongoing	4.3 ± 0.05	2.3 ± 0.06	0.25 ± 0.06
	Transient	4.80 ± 0.10	3.0 ± 0.09	0.80 ± 0.15
Nurse	Ongoing	1.72 ± 0.26	0.98 ± 0.18	—
	Transient	3.84 ± 0.24	2.68 ± 0.22	—
Parent	Ongoing	3.92 ± 0.33	2.04 ± 0.25	0.12 ± 0.05
	Transient	5.56 ± 0.26	3.49 ± 0.22	0.37 ± 0.07

Data are expressed mean ± SE.

model trained with objective ground truth. Categorical agreement rates were fair to substantial ($\kappa = 0.36$ – 0.61) for the model trained with child ratings, and substantial for the model trained with objective ground truth ($\kappa = 0.70$ – 0.72). Inclusion of demographic data as independent variables did not alter binary pain classification model performance.

Estimation of Pain Intensity by CVML Models

Performance of the CVML pain-intensity estimation model is shown in Table 4. The pain-intensity estimation model is moderately to strongly⁴³ correlated ($r = 0.66$ – 0.72) with child self-reports of pain when compared within-subject. Overall (across-subjects) correlations between the pain-intensity estimation model and self-report pain ratings were moderate ($r = 0.46$ – 0.47 , $z = 4.4$ – 6.0 , $P < .0001$). Inclusion of demographic data as independent variables did not alter pain-intensity estimation model performances for either ongoing or transient pain conditions ($z < 0.3$, $P > .80$ for both). Within-subject correlations for the objective pain-intensity estimation models were strong ($r = 0.80$ – 0.86). The overall correlations for the objective models were moderate ($r = 0.55$ – 0.61), but higher than those for the pain-intensity estimation model trained using subjective ground truth of self-ratings ($z = 2.6$, $P = .01$) for transient pain.

Comparison of CVML Models With Nurse and Parent Estimates

CVML model performances are compared with nurses' and parents' ability to estimate children's pain in Tables 3 and 4. For pain versus no-pain categorization, Table 3, nurses attained only slight agreement with the children's self-ratings ($\kappa = 0.15$) for ongoing pain, whereas the binary pain classification model attained fair agreement ($\kappa = 0.36$). Agreements for transient pain were the same for nurses as for the binary pain classification model, and were in the substantial range ($\kappa = 0.61$ for both). Parents attained the highest agreement rates with their children ($\kappa = 0.50$ – 0.72). AUC was similar for nurses and the binary pain classification model for both ongoing and transient pain ($z = 0.18$ – 0.74 , $P > .5$). Parents had the highest signal detection rates for clinically significant pain in their children (AUC = 0.96), and parental AUC was higher than the CVML model AUC (ongoing: $z = 2.4$, $P = .02$; transient: $z = 2.1$, $P = .04$).

Pain-intensity estimation by CVML models are compared with nurses and parents in Table 4. For ongoing pain, the pain-intensity estimation model demonstrated a higher correlation with child self-ratings than nurses ($z = 2.0$, $P = .04$) and performed similar to parents ($z = 1.2$, $P = .22$). For transient pain, the pain-intensity estimation model performed similarly to nurses ($z = 1.7$, $P = .09$), but demonstrated a lower correlation than parents with

TABLE 3 Performance of BPC Models, and Comparison With Nurses and Parents for Binary Presence/Absence of Clinically Significant Pain.

Pain Condition	Metric ^a	BPC					
		Machine				Human	
		BPC	BPC + Demographics	BPC-Objective	BPC-Objective + Demographics	Nurse	Parent
Ongoing	AUC	0.84 ± 0.05	0.84 ± 0.06	0.91 ± 0.03	0.93 ± 0.03	0.86 ± 0.04	0.96 ± 0.02
	Cohen's κ	0.36 ± 0.10	0.34 ± 0.11	0.61 ± 0.07	0.70 ± 0.06	0.15 ± 0.06	0.50 ± 0.09
Transient	AUC	0.91 ± 0.03	0.91 ± 0.03	0.94 ± 0.02	0.94 ± 0.02	0.93 ± 0.02	0.96 ± 0.01
	Cohen's κ	0.61 ± 0.07	0.61 ± 0.07	0.72 ± 0.04	0.72 ± 0.04	0.61 ± 0.07	0.72 ± 0.06

BPC, Binary Pain Classification; +Demographics, demographics included as input to model, with face vector; Objective, time since surgery used as objective ground truth for pain; Ongoing, ongoing pain condition; Transient, transient pain stimulus condition.

^a Data are expressed as mean ± SE, computed over cross 10 validation partitions. See Supplemental Information.

child self-ratings ($z = 2.5, P = .01$). Comparing ongoing with transient pain conditions, nurses' correlations with child ratings were statistically significantly lower for ongoing than transient pain (overall $r = 0.53$ vs $0.74, z = 3.6, P < .001$), whereas pain-intensity estimation model estimates performed similarly for the 2 conditions ($z = 0.2, P > .8$), as did parent ratings ($z = 0.5, P > .6$).

We also examined discrepancies (differences) between child self-ratings and the model, nurse, and parent estimates, as listed in Table 4. The pain-intensity estimation model had the lowest mean discrepancy (0.0) followed by parents with a mean discrepancy of -0.3 for ongoing pain and +0.3 for transient pain, respectively. Nurses' estimates demonstrated the largest mean discrepancy of -1.4 from child ratings for ongoing pain (ie, on average, nurses estimated pain 1.4 points lower than children on the 11-point

NRS scale, $t[149] = 95.3, P < .0001$). The mean nurse estimation discrepancy for transient pain was -0.7, which was also statistically significantly below 0 ($t[296] = 109.7, P < .0001$). Nurse estimation discrepancy was larger for ongoing than transient pain ($t[443] = 43.7, P < .0001$).

DISCUSSION

Overall, CVML pain assessment models derived from automated facial expression measurements performed well in detecting clinically significant pain and in estimating pain severity for both ongoing/endogenous and transient/exogenous pain experiences in the postoperative setting. In addition, the CVML model performed at least as well as commonly used by-proxy pain assessment methods within subjects, suggesting efficacy in monitoring pediatric pain trajectories after surgery.

Computer-Based Detection of Pain and Estimation of Pain Severity

Strong signal detection rates were observed for assessment of the presence or absence of clinically significant pain. In addition, a positive linear relationship was observed between children's self-reports of pain severity and pain scores generated by our pain-intensity estimation models. Developed CVML models also effectively reflected recovery from surgery as measured by elapsed time. Taken together, CVML-based assessment of facial expressions has excellent concurrent criterion validity for pain detection and assessment. Our findings support the usefulness of automated facial expression measurement as a source for inferring the presence and severity of children's pain experiences. The strong correlations we observed between our models based on facial AUs alone and children's self-reported pain levels for

TABLE 4 Performance of PIE Models, and Comparison With Nurses and Parents for Estimating Pain-Intensity Levels

Pain Condition	Metric	PIE						
		Machine				Human		
		PIE	PIE + Demographics	PIE-Objective	PIE-Objective + Demographics	PIE + Time	Nurse	Parent
Ongoing	r (within) ^a	0.72 ± 0.07	0.71 ± 0.07	0.86 ± 0.03	0.86 ± 0.03	0.90 ± 0.02	0.69 ± 0.06	0.88 ± 0.02
	r (overall)	0.47	0.45	0.55	0.55	0.68	0.53	0.75
	Discrepancy ^b	0.0 ± 0.18	—	-0.01 ± 0.18	—	0.0 ± 0.15	-1.4 ± 0.18	-0.3 ± 0.14
Transient	r (within) ^a	0.66 ± 0.05	0.65 ± 0.05	0.80 ± 0.04	0.80 ± 0.04	0.82 ± 0.03	0.77 ± 0.04	0.78 ± 0.04
	r (overall)	0.46	0.45	0.61	0.59	0.67	0.74	0.77
	Discrepancy ^b	0.0 ± 0.15	—	0.01 ± 0.15	—	0.0 ± 0.13	-0.7 ± 0.11	0.3 ± 0.11

PIE, pain-intensity estimation; +Demographics, demographics included as input to model, with face vector; +Time, time included as input to model, with face vector; Objective, time since surgery used as objective ground truth for pain; Ongoing, ongoing pain condition; Transient, transient pain stimulus condition; r (within), mean Pearson correlation coefficient for within-subject correlations; r (overall), Pearson correlation coefficient over all test data.

^a Data are expressed as mean ± SE, computed over the 50 within-subject correlations.

^b Data are expressed as mean ± SE, computed over estimates.

both ongoing and transient pain were comparable to those previously demonstrated in research evaluating manual facial activity coding.⁴⁴

Our models demonstrated efficacy for estimating both ongoing clinical endogenous pain and transient exogenous pain: a distinction recognized by many professionals as important for understanding both pain mechanisms and delivery of clinical care.^{31,32,45,46} These forms of pain respond differently to various pain medications and should be individually assessed and managed.⁴⁷ Our models not only demonstrate efficacy in estimating and assessing pain levels in both case scenarios, but also offer a potential method to continuously monitor these pain experiences accurately and efficiently.

This study explored 2 forms of ground truth for the developed CVML pain models, a subjective ground truth consisting of self-ratings of pain, and an objective form of ground truth consisting of time since surgery. Overall, CVML models were better at estimating the objective form of ground truth. For pain scenarios with well-studied and recognized recovery trajectories, such as the current model of laparoscopic appendectomy,^{38,39} time is potentially a very useful form of ground truth for the study of pain and for development of tools for estimating pain. Future work will explore combining both forms of ground truth in the development of CVML models for estimating pain from the face.

Comparison With Proxy Pain Assessment by Nurses and Parents

The CVML automated assessment method was at least as successful at estimating children's self-ratings of pain severity as nurses for exogenous (transient) pain. For endogenous (ongoing) pain, the CVML scores performed better than nurses, both for detecting clinically significant pain and for estimating children's pain severity ratings. These findings are

similar to those reported in the literature demonstrating a discrepancy between children's pain ratings and nurses' perceptions of pain severity in children.^{8-10,48} Our findings demonstrated a tendency for nurses to underestimate pain relative to children's self-ratings, and this discrepancy was statistically significantly larger for endogenous (ongoing) pain than for the exogenous (transient) pain. The CVML approach presented here, on the other hand, demonstrated no discrepancy for both types of pain.

Automatic detection of clinical pain, using a system developed and tested in this study, could aid medical professionals by providing efficient, accurate, valid, and scalable patient monitoring. Current pain-monitoring systems rely heavily on self-report and human observer-based assessment, which require significant human resources and time, and are potentially biased. Major advantages of automated detection systems include the potential for continuous, long-term monitoring; less reliance on and need for human resources; and reduced bias. Using CVML systems to alert clinicians to instances of pain at the time they occur instead of during scheduled assessments could further enhance efficient, timely allocation of pain interventions appropriate to need. The system is capable of operating in real time.

Effect of Demographics on CVML-Based Pain Estimation

Inclusion of demographic data (gender, age, race/ethnicity) as independent variables did not alter the accuracy of CVML pain estimation, suggesting that facial pain expressions do not vary over the gender, ethnicities, or age range of youth studied. Our findings stand in contrast to research indicating that human observers' estimates of pain tend to vary with certain demographic characteristics, namely a person's gender, age, socioeconomic status, and racial attributes.¹⁶⁻¹⁸ Further studies are needed in infants and children younger than 5 years, as well as in those with developmental/cognitive delay.

Study Limitations and Opportunities for Future Research

We present an initial investigation of CVML-assisted measurement of facial expressions as a measure of children's pain in the postoperative setting. Although promising, the approach still warrants additional investigation with other forms of clinical pain and across the broad age range of children. CV analysis of facial expression requires an approximately frontal camera (Fig 2). The specifications of the CERT system tested here were $\pm 15^\circ$ of frontal in yaw, pitch, and roll. The newer commercial version of the software (Emotient Analytics, San Diego, CA)



FIGURE 2 Picture of the camera setup used for video recording at the hospital bed during study visits 1 and 2.

improves the head pose tolerance to $\pm 30^\circ$. In our study, children were evaluated in an upright/semi-upright position; however, cameras mounted above hospital beds are becoming more available and can provide frontal views of supine patients. CV also requires at least moderate lighting. In the current study, lighting consisted of standard ceiling lights with window shades closed to avoid strong side lighting. Moderate motion is not a problem, but rapid motion can blur at 30 frames per second; higher frame rates would reduce blur. Further work is needed to determine whether such a tool can be easily integrated into clinical workflow and thus add benefit to current clinical

pain assessment methods and ultimately treatment paradigms.

CONCLUSIONS

A novel ML-generated tool measuring facial expressions achieves concurrent validity for pain assessment as demonstrated by good-to-excellent accuracy in binary pain classification, strong correlations with patient self-reported pain ratings, and parent-equivalent estimation of children's pain levels.

ACKNOWLEDGMENTS

We acknowledge the contributions of the patients and families and

the nursing staff at Rady Children's Hospital San Diego who volunteered their participation to make this work possible.

ABBREVIATIONS

AU: action unit
 AUC: area under the receiver operating characteristic curve
 CERT: Computer Expression Recognition Toolbox
 CV: computer vision
 FACS: Facial Action Coding System
 ML: machine learning
 NRS: Numerical Rating Scale

FINANCIAL DISCLOSURE: Dr Bartlett is a founder, employee, and shareholder of Emotient, Inc. Mr Ahmed was a contractor at Emotient for 32 hours on a project unrelated to pain and unrelated to health care. The other authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: All phases of this study were supported by National Institutes of Health National Institute of Nursing Research grant R01 NR013500. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Funded by the National Institutes of Health (NIH).

POTENTIAL CONFLICT OF INTEREST: Dr Bartlett is a founder, employee, and shareholder of Emotient, Inc. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies. Mr Ahmed was a contractor at Emotient for 32 hours on a project unrelated to pain and unrelated to health care. The other authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

- Rashiq S, Dick BD. Post-surgical pain syndromes: a review for the non-pain specialist. *Can J Anaesth*. 2014;61(2):123–130
- Chorney JM, McGrath P, Finley GA. Pain as the neglected adverse event. *CMAJ*. 2010;182(7):732
- Upp J, Kent M, Tighe PJ. The evolution and practice of acute pain medicine. *Pain Med*. 2013;14(1):124–144
- American Society of Anesthesiologists Task Force on Acute Pain Management. Practice guidelines for acute pain management in the perioperative setting: an updated report by the American Society of Anesthesiologists Task Force on Acute Pain Management. *Anesthesiology*. 2012;116(2):248–273
- American Academy of Pediatrics. Committee on Psychosocial Aspects of Child and Family Health; Task Force on Pain in Infants, Children, and Adolescents. The assessment and management of acute pain in infants, children, and adolescents. *Pediatrics*. 2001;108(3):793–797
- Cohen LL, Lemanek K, Blount RL, et al. Evidence-based assessment of pediatric pain. *J Pediatr Psychol*. 2008;33(9):939–955; discussion 956–957
- Larochette AC, Chambers CT, Craig KD. Genuine, suppressed and faked facial expressions of pain in children. *Pain*. 2006;126(1-3):64–71
- Duignan M, Dunn V. Congruence of pain assessment between nurses and emergency department patients: a replication. *Int Emerg Nurs*. 2008;16(1):23–28
- Puntillo K, Neighbor M, O'Neil N, Nixon R. Accuracy of emergency nurses in assessment of patients' pain. *Pain Manag Nursing*. 2003;4(4):171–175
- Rajasagaram U, Taylor DM, Braitberg G, Pearsell JP, Capp BA. Paediatric pain assessment: differences between triage nurse, child and parent. *J Paediatr Child Health*. 2009;45(4):199–203
- Singer AJ, Gulla J, Thode HC Jr. Parents and practitioners are poor judges of young children's pain severity. *Acad Emerg Med*. 2002;9(6):609–612
- Zhou H, Roberts P, Horgan L. Association between self-report pain ratings of child and parent, child and nurse and parent and nurse dyads: meta-analysis. *J Adv Nurs*. 2008;63(4):334–342
- von Baeyer CL, Spagrud LJ, McCormick JC, Choo E, Neville K, Connelly MA. Three new datasets supporting use of the Numerical Rating Scale (NRS-11) for children's self-reports of pain intensity. *Pain*. 2009;143(3):223–227
- Chang J, Versloot J, Fashler SR, McCrystal KN, Craig KD. Pain assessment in children: validity of facial expression items in observational pain scales. *Clin J Pain*. 2015;31(3):189–197

15. Kappesser J, Williams AC, Prkachin KM. Testing two accounts of pain underestimation. *Pain*. 2006;124(1-2): 109–116
16. Schiavenato M, Craig KD. Pain assessment as a social transaction: beyond the “gold standard.” *Clin J Pain*. 2010;26(8):667–676
17. Tait RC, Chibnall JT. Racial/ethnic disparities in the assessment and treatment of pain: psychosocial perspectives. *Am Psychol*. 2014;69(2):131–141
18. Tait RC, Chibnall JT, Kalauokalani D. Provider judgments of patients in pain: seeking symptom certainty. *Pain Med*. 2009;10(1):11–34
19. De Ruddere L, Goubert L, Prkachin KM, Stevens MA, Van Ryckeghem DM, Crombez G. When you dislike patients, pain is taken less seriously. *Pain*. 2011; 152(10):2342–2347
20. De Ruddere L, Goubert L, Stevens MA, Deveugele M, Craig KD, Crombez G. Health care professionals’ reactions to patient pain: impact of knowledge about medical evidence and psychosocial influences. *J Pain*. 2014;15(3):262–270
21. De Ruddere L, Goubert L, Vervoort T, Kappesser J, Crombez G. Impact of being primed with social deception upon observer responses to others’ pain. *Pain*. 2013;154(2):221–226
22. Hadjistavropoulos T, Hadjistavropoulos HD, Craig KD. Appearance-based information about coping with pain: valid or biased? *Soc Sci Med*. 1995;40(4): 537–543
23. Craig KD, Prkachin KM, Grunau RVE. *The Facial Expression of Pain. Handbook of Pain Assessment*. 3rd ed. New York, NY: Guilford; 2011:117–133
24. Williams AC. Facial expression of pain: an evolutionary account. *Behav Brain Sci*. 2002;25(4):439–455; discussion 455–488
25. Ekman P, Friesen W. *Facial action coding system: a technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press; 1978
26. Bartlett MS, Hager JC, Ekman P, Sejnowski TJ. Measuring facial expressions by computer image analysis. *Psychophysiology*. 1999;36(2):253–263
27. Lucey P, Cohn JF, Matthews I, et al. Automatically detecting pain in video through facial action units. *IEEE Trans Syst Man Cybern B Cybern*. 2011;41(3): 664–674
28. Pantic M, Rothkrantz LJ. Expert system for automatic analysis of facial expressions. *Image Vis Comput*. 2000;18:881–905
29. Littlewort G, Whitehill J, Wu T, et al. The computer expression recognition toolbox (CERT). *IEEE Conference on Automatic Face and Gesture Recognition*. Volume 9. 2011; 298–305
30. Bartlett MS, Littlewort GC, Frank MG, Lee K. Automatic decoding of facial movements reveals deceptive pain expressions. *Curr Biol*. 2014;24(7):738–743
31. Bennett GJ. What is spontaneous pain and who has it? *J Pain*. 2012;13(10): 921–929
32. Srikantharajah S, Gilron I. Systematic review of movement-evoked pain versus pain at rest in postsurgical clinical trials and meta-analyses: a fundamental distinction requiring standardized measurement. *Pain*. 2011;152(8): 1734–1739
33. Birnie KA, McGrath PJ, Chambers CT. When does pain matter? Acknowledging the subjectivity of clinical significance. *Pain*. 2012;153(12):2311–2314
34. Hodgins MJ. Interpreting the meaning of pain severity scores. *Pain Res Manag*. 2002;7(4):192–198
35. Hoffman DL, Sadosky A, Dukes EM, Alvir J. How do changes in pain severity levels correspond to changes in health status and function in patients with painful diabetic peripheral neuropathy? *Pain*. 2010;149(2):194–201
36. Ashraf AB, Lucey S, Cohn JF, et al. The painful face: pain expression recognition using active appearance models. *Image Vis Comput*. 2009;27(12):1788–1796
37. Tukey JW. Bias and confidence in not-quite large samples. *Ann Math Stat*. 1958;29:614
38. Miyauchi Y, Sato M, Hattori K. Comparison of postoperative pain between single-incision and conventional laparoscopic appendectomy in children. *Asian J Endosc Surg*. 2014;7(3):237–240
39. Paya K, Fakhari M, Rauhofer U, Felberbauer FX, Rebhandl W, Horcher E. Open versus laparoscopic appendectomy in children: a comparison of complications. *JSLS*. 2000;4(2):121–124
40. Lejus C, Delile L, Plattner V, et al. Randomized, single-blinded trial of laparoscopic versus open appendectomy in children: effects on postoperative analgesia. *Anesthesiology*. 1996;84(4):801–806
41. El Khouli RH, Macura KJ, Barker PB, Habba MR, Jacobs MA, Bluemke DA. Relationship of temporal resolution to diagnostic performance for dynamic contrast enhanced MRI of the breast. *J Magn Reson Imaging*. 2009;30(5):999–1004
42. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8(4):283–298
43. Taylor R. Interpretation of the correlation coefficient: a basic review. *J Diag Med Sonogr*. 1990;1:35–39
44. Kunz M, Mylius V, Schepelmann K, Lautenbacher S. On the relationship between self-report and facial expression of pain. *J Pain*. 2004;5(7):368–376
45. Dennis SG, Melzack R. Self-mutilation after dorsal rhizotomy in rats: effects of prior pain and pattern of root lesions. *Exp Neurol*. 1979;65(2):412–421
46. Mogil JS, Craiger SE. What should we be measuring in behavioral studies of chronic pain in animals? *Pain*. 2004;112(1-2):12–15
47. Dahl JB, Erichsen CJ, Fuglsang-Frederiksen A, Kehlet H. Pain sensation and nociceptive reflex excitability in surgical patients and human volunteers. *Br J Anaesth*. 1992;69(2):117–121
48. Chambers CT, Reid GJ, Craig KD, McGrath PJ, Finley GA. Agreement between child and parent reports of pain. *Clin J Pain*. 1998;14(4):336–342