Measurement Of Acute Pain In The Pediatric Emergency Department Through Automatic Detection Of Behavioral Parameters: A Pilot Study

Letizia Bergamasco[1,2][0000-0001-6462-1699], Marco Gavelli[1][0000-0001-9187-7358], Carla Fadda[3][0000-0003-2898-2221], Emilia Parodi[4][0000-0002-4260-2689], Claudia Bondone[5][0000-0003-4120-2533] and Emanuele Castagno[5][0000-0001-7756-5075]

[1]Fondazione LINKS, Via Pier Carlo Boggio 61, 10138 Torino, Italy
{letizia.bergamasco,marco.gavelli}@linksfoundation.com
[2]Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[3]Scuola di Specializzazione in Pediatria, Università degli Studi di Torino, Piazza Polonia 94, 10126 Torino, Italy
carla.fadda@unito.it
[4]S.C. Pediatria e Neonatologia, A.O. Ordine Mauriziano di Torino, Via Magellano 1, 10128 Torino, Italy
eparodi@mauriziano.it
[5]S.C. Pediatria d'Urgenza, Ospedale Infantile Regina Margherita – A.O.U. Città della Salute e della Scienza di Torino, Via Zuretti 23, 10126 Torino, Italy
{ecastagno,cbondone}@cittadellasalute.to.it

Abstract. Acute pain is a frequent symptom in children who access the Emergency Department (ED). Its measurement through validated tools compatible with the time of triage is essential to develop the most appropriate pain-relieving strategy. The algometric scales that can be used in children in whom self-assessment is not possible are based on the evaluation of behavioral and physiological parameters. However, the actual use of algometric scales in the ED is scarce due to environmental factors, heterogeneity of the scales and lack of training, thus making automated pain assessment desirable. In this study, we propose a camera-based system to provide an objective and contactless pain assessment in children aged less than 3 years, through the automatic detection of behavioral parameters from video recordings. To investigate the feasibility of its usage in the ED environment, we collected video recordings of healthy children aged 3-36 months admitted to the ED with acute pain as the main or accompanying symptom, while pain was measured by a healthcare professional according to the Face, Legs, Activity, Cry, and Consolability (FLACC) pain scale. For the recorded videos, we compared the scores for the items Face (F), Legs (L) and Activity (A) given by the operator with the ones given by our system, analyzing the potentiality and limitations of our approach. By showing that automatic pain assessment in young children in the ED could integrate human evaluation to make it easier and faster, without substituting it, we provide the basis for further research in this field.

Keywords: automatic pain assessment · camera-based approach · children · pain · pediatric emergency department · algometric scale · FLACC · Google Mediapipe

1 Introduction

Acute pain is a frequent and feared symptom in childhood and is reported in up to 78% of admissions to the Pediatric Emergency Department (ED) [1]. Pain should be adequately considered, measured, and treated whenever it is reported by children or their caregivers, regardless of age, clinical situation, and social role [2]. Acute and repetitive pain experienced at early stages of life can lead to persistent structural and functional changes of the nociceptive system, helping to determine the final architecture of pain system [3]. Many studies have indeed confirmed that untreated pain produces both short and long-term physical and psychological negative effects: healing times are lengthened, complications increase, and long-term sequelae may develop [4-5].

Accurate assessment and measurement of pain are the cornerstones of pain management and are essential to provide timely adequate analgesic strategy. The measurement of pain in children under the age of 3 years, who cannot provide effective self-assessment, requires standardized, validated tools appropriate to their developmental level, the context, and their prior pain experiences. This can be mostly challenging in the ED setting, in particular on the very first evaluation in triage, where time is limited, anxiety is high, and children and their caregivers are unfamiliar with healthcare professionals and environment [6]. Available validated objective scales are based on the evaluation of both physiological and behavioral parameters. However, literature data suggest that the actual use of algometric scales in the Pediatric ED is limited. Major critical issues are related to environmental factors specific to triage, heterogeneity of scales used, and training deficiencies [7-9]. Among objective pain scales, the Face, Legs, Activity, Cry, and Consolability (FLACC) is based on the detection of behavioral parameters and has been validated for children less than 3 years also in the emergency setting [10].

Due to such reasons, the assessment of pain through objective scales in younger children could be improved by the development of automated machine-based systems aiming to monitor different pain indicators and providing a consistent, minimally biased evaluation of pain. In the past years, there has been an increasing interest in the use of technology for understanding human behavioral responses to pain based on the analysis of facial expressions, and of body or head movements, which are the most important indicators in patients with verbal communication inability [11-13]. Other studies have shown that machine-based systems can be used to detect and analyze physiological changes associated with pain, such as changes in skin color, increase in heart rate [14-15], and changes in the cerebral hemodynamic of specific brain's regions [16].

Several machine-based approaches have been introduced to analyze infants' body movements for the purpose of diagnosing a specific disease [17-18]. The development of a machine-based multimodal pain assessment tool that dynamically measures pain in infants has been also proposed, based on the analysis of different behavioral and physiological indicators [19]. Anyway, to our knowledge, no clinical trials have yet been conducted, and most of face detection algorithms are designed and trained for adult faces [19].

Our group has already developed and demonstrated a proof-of-concept computerized tool for the evaluation of pain in newborns, based on the analysis of facial expressions in video recordings [20]. Pain scores obtained from automated analysis have been compared to those assigned by trained healthcare professionals according to three objective neonatal pain scales: the Neonatal Facial Coding System, (NFCS), the Premature Infant Pain Profile (PIPP), and the Douleur Aiguë du Nouveau-né (DAN), showing that manual pain evaluation is challenging and often results in a large variability across scores between different operators, making automated assessment desirable [20].

The main aim of this pilot study was to develop an automated computerized tool for pain evaluation in children aged less than 3 years, using only video recordings acquired from an RGB camera without the aid of sensors on the skin. Second, we aimed to create a dataset and ad hoc registration setting that could be used to demonstrate the feasibility of the usage of such automatic system for pain assessment for children in this specific age in the ED environment. Finally, our goal was to compare the scores of the behavioral parameters assigned by the automatic system with those assigned by a healthcare operator to the items Face (F), Legs (L) and Activity (A) of the FLACC pain scale, analyzing the potentiality and limitations of our approach.

2 Materials and Methods

2.1 Data Collection

In this pilot study, we enrolled healthy children aged 3-36 months admitted to the Pediatric ED of our tertiary teaching Children's Hospital between April and September 2022 with acute pain as main or accompanying symptom. We excluded children with chronic disorders; those for whom face and limbs were not fully visible because of dressing, medications, or any medical device; and those admitted with high triage priority code. Pain was measured in all children by the same healthcare professional using the Italian validated version of the FLACC scale [10,21] (see Table 1), along with the assessment and recording of heart rate and oxygen saturation for 60 seconds using the pulse oximeter available in our ED. At the same time, a 60 second video of children's full figure was recorded with an RGB camera with a resolution of 1920x1080 pixels and a frame rate of 30 fps. The camera device was placed 100 cm far in front of the subject, in the same light conditions. Written informed consent was obtained from all parents of the involved children. The study protocol was approved by our Local Ethic Committee.

Overall, we have recruited 14 Caucasian children (7 males); the median age was 16 months (range 3-32 months). A total of 22 minutes of recording were acquired, with a mean length of 1.16 minutes per child. The acquisition of recordings for a sufficient time was possible for all children, with variable duration of recording fragments suitable for automatic analysis, due to the movements of the child.

Table 1. Face, Legs, Activity, Cry and Consolability (FLACC) scale (adapted from [10]).

Categories	Scoring		
	0	1	2
Face	No particular expression	Occasional grimace/frown, withdrawn or disinterested	Frequent/constant quivering chin, clenched jaw
Legs	Normal position or relaxed	Uneasy, restless, tense	Kicking or legs drawn up
Activity	Lying quietly, normal position, moves easily	Squirming, shifting back and forth, tense	Arched, rigid or jerking
Cry	No cry	Moans or whimpers, occasional complaint	Crying steadily, screams or sobs, frequent complaints
Consolability	Content and relaxed	Reassured by occasional touching, hugging or being talked to, distractible	Difficult to console or comfort

2.2 Pain Scale Implementation

Since the FLACC pain scale has been validated for children less than 3 years in the emergency setting, we have selected it as the reference pain scale for the output for our automated pain evaluation system. However, in our implementation of the FLACC, cry (C) and consolability (C) categories were excluded, since the audio signal acquired in the emergency department environment turned out to be too noisy to be properly processed and analyzed. Therefore, for this research study a partial version of the FLACC (that we will call pFLACC) has been used, including only face (F), legs (L) and activity (A). The pFLACC score related to an observation period was calculated as defined in Eq. (1):

$$pFLACC_{score} = F_{score} + L_{score} + A_{score} \quad (1)$$

where F_{score}, L_{score} and A_{score} indicate the scores computed for the single categories (Face, Legs and Activity, respectively) in the considered observation period. Each single-category score ranges between 0 and 2, thus resulting in a total $pFLACC_{score}$ ranging between 0 and 6.

The computation of these single-category scores was based on the analysis of facial expressions and body movements. Some face and body parameters have been identified, deriving from the way the FLACC score is traditionally computed, as described in Table 1. In particular, the identified parameters were 7: mouth opening, brow bulging, eye squeezing, legs outstretching, pedaling, body movement, and arms flailing.

For each parameter, an algorithm has been developed to quantify it using numerical values, as will be better explained in Section 2.3. Moreover, a suitable threshold has been defined for each parameter to discriminate the range of values of the parameter representing the "normal" condition, and the range of values representing the "non-normal" condition, related to the pain experience. Threshold values have been defined empirically and have been calibrated based on the available dataset of videos.

All the 7 parameters were continuously measured by the automatic system along all the observation period, at the end of which the $pFLACC_{score}$ was computed. In the case of the videos of our dataset, the observation period was 60 seconds, corresponding to the duration of the videos. The values of each parameter were compared with the threshold established for that parameter for the duration of the whole observation period; two time thresholds have also been defined empirically (1/3 and 2/3 of the observation time) to check for how long the parameters were in "normal" and "non-normal" range, and therefore assign value 0, 1 or 2 to the pFLACC items. The detailed description of the computation of single-category scores (F, L, A) is reported in the following paragraphs.

Face Score. The computation of F_{score} involved mouth opening, brow bulging and eye squeezing parameters. Such parameters were computed and compared to the relative thresholds along an observation period, as previously defined. A score ranging from 0 to 2 was assigned to F_{score}, according to the methodology illustrated in Table 2.

Table 2. Computation of the face score (F_{score}).

Status of involved parameters	**Interpretation**	**Assigned F_{score}**
Mouth opening, brow bulging and eye squeezing exceeding the corresponding thresholds for less than 1/3 of the observation period	Neutral expression	0
Mouth opening, brow bulging or eye squeezing exceeding the corresponding thresholds for more than 1/3 and less than 2/3 of the observation period	Occasional frown	1
Mouth opening, brow bulging or eye squeezing exceeding the corresponding thresholds for more than 2/3 of the observation period	Frequent frown	2

Legs Score. The computation of L_{score} involved legs outstretching and pedaling parameters. Similarly to F_{score}, such parameters were computed and compared to the relative thresholds along an observation period. A score ranging from 0 to 2 was assigned to L_{score}, according to the methodology illustrated in Table 3.

Table 3. Computation of the legs score (L_{score}).

Status of involved parameters	**Interpretation**	**Assigned L_{score}**
Legs outstretching and pedaling exceeding the corresponding thresholds for less than 1/3 of the observation period	Relaxed	0
Legs outstretching or pedaling exceeding the corresponding thresholds for more than 1/3 and less than 2/3 of the observation period	Restless	1
Legs outstretching or pedaling exceeding the corresponding thresholds for more than 2/3 of the observation period	Kicking or outstretching legs	2

Activity Score. The computation of A_{score} involved body movement and arms flailing parameters. Also in this case, similarly to F_{score} and L_{score}, the involved parameters were computed and compared to the relative thresholds along an observation period, resulting in a score ranging from 0 to 2. The adopted methodology is illustrated in Table 4.

Table 4. Computation of the activity score (A_{score}).

Status of involved parameters	**Interpretation**	**Assigned A_{score}**
Body movement and arms flailing exceeding the corresponding thresholds for less than 1/3 of the observation period	Normal position	0
Body movement or arms flailing exceeding the corresponding thresholds for more than 1/3 and less than 2/3 of the observation period	Squirming	1
Body movement or arms flailing exceeding the corresponding thresholds for more than 2/3 of the observation period	Jerking	2

2.3 Face and Body Parameters Computation

As mentioned in Section 2.2, there are 7 parameters involved in the computation of the $pFLACC_{score}$: mouth opening, brow bulging, eye squeezing, legs outstretching, pedaling, body movement, and arms flailing. We have calculated these parameters starting from the child’s face and body landmarks, that were detected and tracked in the video recording along the observation period. To do so, in this work we used Google Mediapipe Holistic (GMH), an open-source framework using machine learning techniques to detect and track face and pose landmarks in real time from the video recording of a person [22,23].

For each frame of a video, GMH provides 543 landmarks (33 pose landmarks, 21 hand landmarks for each hand, and 468 face landmarks) [23]. It also allows setting a minimum confidence value for the detection and tracking of these points; in our system, we set these values to 0.5, so that we kept as valid landmarks only the ones with confidence higher than 0.5. For a set of landmarks, we will call a frame where those landmarks are valid a "valid frame". For what concerns the format of coordinates of the landmarks, we chose to use the coordinates that GMH provides in the image reference frame. Each landmark is identified by a set of 3D coordinates (x, y, z); in our analysis we used only x and y coordinates, and not the z coordinate, being an estimation of the depth of the point made by GMH. Fig. 1 shows an example of a frame of a video in our collected dataset, where landmarks provided by GMH are visualized.

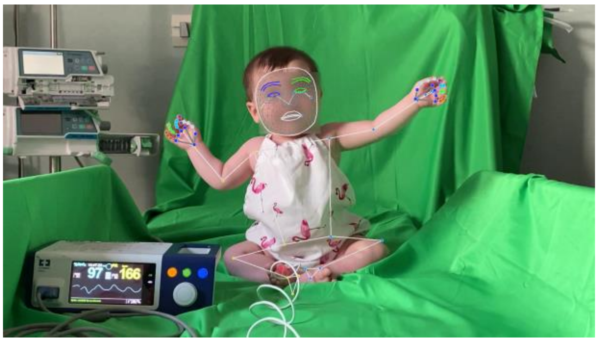

Fig. 1. Example of a frame of a video in our collected dataset, where landmarks provided by GMH are visualized. The image has been postprocessed to anonymize the child's face.

Among all the provided landmarks, we chose the ones that we found most appropriate to quantify the 7 identified parameters, as will be explained in the following paragraphs. In general, all the parameters were computed with a moving window mechanism, with a window length of 1 second. Basically, each video recording was read frame by frame in a simulated real-time fashion. At each new frame that was being analyzed, the current window was shifted by one frame. Some parameters could be calculated for each single valid frame (mouth opening, brow bulging, eye squeezing, legs outstretching), and were therefore averaged over the whole window to get a single value for the window; the other parameters (pedaling, body movement, arms flailing), instead, could be calculated only once for the whole window, by considering the variation of intermediate features along the window. The scheme of the moving window mechanism is illustrated in Fig. 2.

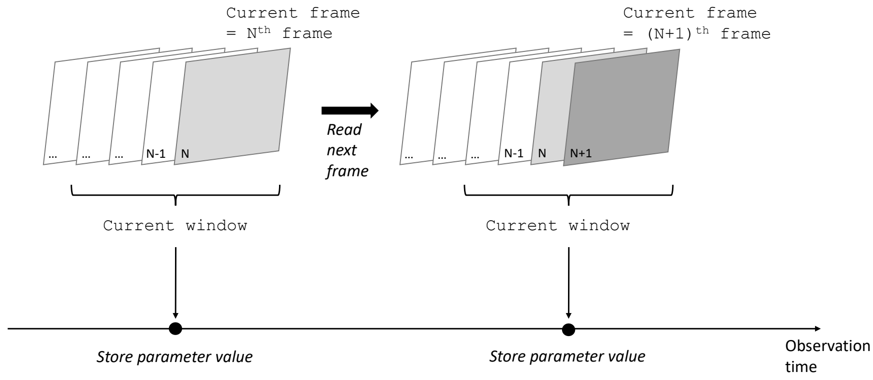

Fig. 2. Scheme of the moving window mechanism for the calculation of face and body parameters.

Mouth Opening. For each frame, the height of the mouth was computed as the height of the minimum bounding rectangle enclosing all the mouth landmarks. Similarly, the height of the whole face was computed as the height of the rectangle enclosing all the face oval landmarks. The mouth opening value for the considered frame was given by the ratio between the mouth height and the face height, converted to a percentage. The mouth opening value for a window was the average of the mouth opening values of the valid frames in the window.

Brow Bulging. For each frame, the distance between eyebrows medial borders was calculated and the width of the whole face was retrieved as the width of the rectangle enclosing all the face oval landmarks. The brow bulging value for the considered frame was given by the ratio between the distance between eyebrows medial borders and the face width, converted to a percentage. The brow bulging value for a window was the average of the brow bulging values of the valid frames in the window.

Eye Squeezing. For each frame, and for both the right and left part of the face, the distance between mid-eyebrow and mid lower eyelid was calculated. Then, the mean between these two distances was retrieved. The face height was computed in the same way as for the mouth opening parameter. The eye squeezing value for the considered frame was given by the ratio between the calculated mean distance and the face height, converted to a percentage. The eye squeezing value for a window was the average of the eye squeezing values of the valid frames in the window.

Legs Outstretching. For each frame, the angle formed by each leg was computed from the positions of the landmarks representing the extremities of the leg and the knee. The legs outstretching value for the considered frame was given by the maximum between the angles formed by each of the two legs. The legs outstretching value for a window was the average of the legs outstretching values of the valid frames in the window.

Pedaling. For each frame, the distance between hip and ankle for each leg was computed from the positions of the corresponding landmarks. In a window, the standard deviation of these distances was calculated for each leg separately. The pedaling value for a window was given by the maximum of these two standard deviation values.

Body Movement. For each frame, the center of the torso was calculated as the center of the quadrilateral represented by the four landmarks at the extremities of the torso. In a window, the standard deviation of the positions assumed by the center of the torso was computed. The pedaling value for a window was given by such standard deviation value.

Arms Flailing. For each frame, the position of the landmarks representing the left wrist and right wrist was considered. In a window, the standard deviation of the positions assumed by each of these two landmarks was computed. The arms flailing value for a window was given by the maximum between the two standard deviation values.

3 Results and Discussion

In Section 2, we have illustrated how it is possible to implement a partial version of the FLACC algometric scale to perform automatic pain assessment. In this Section we analyze and discuss the comparison between the pain scores given by the healthcare professional on our collected dataset and the corresponding pain scores given by our automatic system, to investigate the feasibility of its usage with children aged less than 3 years in the ED environment.

As we have shown in our previous work, where we developed a proof-of-concept tool for the automatic evaluation of pain in newborns [20], while comparing scores obtained from an automatic system and scores assigned by healthcare professionals, we should not aim to merely maximize the agreement between them, but we should rather analyze the causes of the differences between the two scores. This is because of the subjectivity and inter-operator variability that are present in pain evaluation made by healthcare professionals, which are motivating us in the investigation of automatic pain assessment.

Therefore, what we can analyze in our comparison are the differences between the values measured by the machine and the healthcare professional for each item, and the reasons behind them. Indeed, for the same observed limb movements and facial expressions, the analysis of what allows the human operator to distinguish pain from restlessness and anxiety is the basis for developing an efficient computer system.

In Table 5, we report the scores for each item (F, L, A) and the total pFLACC scores obtained for all the collected video recordings, as assigned by the healthcare professional and the automatic system. For a quantitative comparison, we also report cosine similarity values in Table 6.

Table 5. Scores assigned by the healthcare professional and the automatic system to all the video recordings collected in our dataset.

	Healthcare professional				**Automatic system**			
ID	**F**	**L**	**A**	**pFLACC**	**F_{score}**	**L_{score}**	**A_{score}**	**$pFLACC_{score}$**
1	0	0	0	0	0	0	1	1
2	0	0	1	1	0	0	1	1
3	2	1	1	4	1	0	0	1
4	2	0	1	3	2	1	0	3
5	2	0	1	3	2	0	0	2
6	0	0	0	0	0	0	1	1
7	1	0	0	1	0	0	0	0
8	0	0	0	0	0	0	1	1
9	2	1	1	4	1	0	0	1
10	2	1	1	4	0	0	0	0
11	1	0	1	2	0	0	1	1
12	0	1	0	1	0	1	0	1
13	0	0	0	0	1	0	0	1
14	0	0	0	0	1	0	1	2
15	1	0	0	1	0	1	0	1
16	0	0	0	0	0	0	1	1
17	0	0	0	0	0	0	1	1
18	0	0	0	0	0	0	1	1
19	0	0	0	0	0	0	1	1

Table 6. Cosine similarity results for the comparison of the scores assigned by the healthcare professional and the automatic system to all the video recordings collected in our dataset.

F	**L**	**A**	**pFLACC**
0.72	0.29	0.24	0.58

For what concerns the evaluation performed by the healthcare professional, from the comparison of the scores and the feedback given by the healthcare professional, a sort of bias has emerged among analyzed items: facial expression seems to be the most influent to allow healthcare professionals to distinguish restlessness from real pain, followed by activity and lastly by legs movements. In fact, the values assigned to L (legs) and A (activity) quantify movement in general, but do not define the mood in detail. It is therefore difficult to distinguish when this occurs because of pain or due to other reasons.

With respect to the automatic system, instead, the three items (F, L, A) are calculated in a completely independent way, and they have the same weight in the summation that is performed to get the pFLACC score. If there is a high value for the face item, this does not influence the computation of the legs and activity items.

On the technical side, the major limitation that we observed for the automatic system is that the robustness in the landmarks tracking performed by GMH may not be sufficient when movements occur. In fact, when the child turns to the side or moves the head, the landmarks tracking experiences a degraded performance or is even lost. This may lead to get parameters values that exceed the corresponding thresholds, and, consequently, F_{score}, L_{score} and A_{score} assume values greater than 0 even when the child is quiet.

Using adaptive thresholds instead of fixed thresholds for the face and body parameters could make the system more robust to the variation of facial proportions in children of different ages, and help mitigating the effects of the landmark tracking issues.

If we compare our study with the experience of Parodi et al [20] with newborns, it is worth noting that pain evaluation in video recordings of children aged less than 3 years is more challenging: newborns not only move less, but also have a different range of facial expressions and movements than infants and older children. Such observation could be considered to enhance the software performance and for the future development of the automatic system.

Of course, the automatic tool tested in our preliminary research is still not suitable to be used in routine practice, as it represents the first exploratory step within a larger project. Indeed, the analysis that we have performed has been limited by the small population, in particular the low number of infants of our dataset.

Anyway, our results represent the basis to develop a system in which human assessment could be integrated, standardized, and improved thanks to an automatic system. Pain assessment in the pediatric ED is an essential part of triage evaluation and is considered as the fifth vital sign [9]. Rapid and standardized assessment is crucial, but environmental and cultural factors may limit optimal performance on most occasions [7-9]. The standard practice to evaluate pain in children who are not able to make self-assessment is based on the observation made by caregivers and healthcare professionals through validated scales that sometimes fail to meet psychometric standards and requires continuous monitoring. Inter-operator variability is undoubtedly a limit of traditional pain measurement with validated scales. In clinical practice, a more objective approach for pain assessment is desirable to correctly recognize and treat pain.

Different approaches for automated recognition of pain expression have been proposed in the last 20 years; however, most of face detection algorithms have been designed and trained for adult faces and are poorly suitable for infants and young children [19,24-26]. Our research proposes an automatic system for objective and contactless pain assessment through the automatic detection of behavioral parameters from video recordings, showing the potentiality that machine-based pain evaluation has also in infants and children aged less than 3 years.

However, on one side the automatic system could reduce the variability in the use of behavioral pain scale in infants and young children, such as FLACC, and make it faster. On the other side, the expertise of trained triage healthcare professionals will continue to be irreplaceable, as it allows to understand and capture aspects that are not yet appreciable by the machine [20]. In fact, a trained observer can pick up nuances beyond the mere detection of specific movements and expressions, and caregivers can usually discern between face expressions of pain, discomfort, or restlessness of their children. In front of a facial grimace (or a frowning expression), the software calculates movements and expressions in an objective way, while a human operator not only analyzes the same parameters, but is also able to better contextualize them. Therefore, it is essential to properly combine both automatic and human pain assessment. Understanding those aspects paves the way for further development and improvement of our pain assessment system in the future.

4 Conclusion and Future Work

Our pilot study explores the possibility to provide automatic, objective and minimally biased assessment of pain in young children, supporting the evaluation made by healthcare professionals, that remains irreplaceable.

The results of our study suggest that the proposed automated pain assessment system is a promising tool that can potentially aid and improve the evaluation of healthcare professionals even in the emergency setting, providing the basis for further research in this field. Anyway, even when it will be validated for routine practice, automatic detection of behavioral parameters should be never intended as a substitute of the observation provided by healthcare professionals. Instead, it could integrate human evaluation and contribute to make it easier and faster.

The main strength of our study is that, for the first time, we have created the assumptions to collect a dataset for the development of an automatic pain detection system in children in the ED. Moreover, we have implemented the recording setting, adapting the one previously elaborated for newborns [20] to children aged less than 3 years. The collaboration between clinicians and engineers allowed us to create a multidisciplinary project for the development of this system.

Further challenges will be the substantial extension of the original recording dataset and the complementation of video processing with the analysis of audio information, thus implementing the full FLACC pain scale. Moreover, future research will include the improvement of the robustness of landmarks tracking in video recordings of children in movement conditions. Another challenge will be to improve the automatic pain assessment approach so to better encode the nuances that the human operator grasps, and integrate them with the detection of physiological parameters, in order to provide a more objective and standardized evaluation of pain.

References

1. Grant, P. S.: Analgesia delivery in the ED. The American journal of emergency medicine, 24(7), 806-809 (2006).
2. Raja, S. N., et al.: The revised IASP definition of pain: Concepts, challenges, and compromises. Pain, 161(9), 1976 (2020).
3. Walker, S. M.: Neonatal pain. Pediatric Anesthesia, 24(1), 39-48 (2014).

4. Birnie, K. A., et al.: Hospitalized children continue to report undertreated and preventable pain. Pain Research and Management, 19(4), 198-204 (2014).
5. Brattberg, G.: Do pain problems in young school children persist into early adulthood? A 13-year follow-up. European Journal of Pain, 8(3), 187-199 (2004).
6. Drendel, A. L., Brousseau, D. C., Gorelick, M. H.: Pain assessment for pediatric patients in the emergency department. Pediatrics, 117(5), 1511-1518 (2006).
7. Krauss, B. S., Calligaris, L., Green, S. M., Barbi, E.: Current concepts in management of pain in children in the emergency department. The Lancet, 387(10013), 83-92 (2016).
8. Marzona, F., Pedicini, S., Passone, E., Pusiol, A., Cogo, P.: Mandatory Pain Assessment in a Pediatric Emergency Department: Failure or Success?. The Clinical Journal of Pain, 35(10), 826-830 (2019).
9. Benini, F., et al.: Consensus on Pediatric Pain in the Emergency Room: the COPPER project, issued by 17 Italian scientific societies. Italian journal of pediatrics, 46, 1-3 (2020).
10. Voepel-Lewis, T., Shayevitz, J. R., Malviya, S.: The FLACC: a behavioral scale for scoring postoperative pain in young children. Pediatr Nurs, 23(3), 293-297 (1997).
11. Grunau, R. V., Craig, K. D.: Pain expression in neonates: facial action and cry. Pain, 28(3), 395-410 (1987).
12. Brahnam, S., Chuang, C. F., Shih, F. Y., Slack, M. R.: Machine recognition and representation of neonatal facial displays of acute pain. Artificial intelligence in medicine, 36(3), 211-222 (2006).
13. Brahnam, S., Nanni, L., Sexton, R.: Introduction to neonatal facial pain detection using common and advanced face classification techniques. Advanced Computational Intelligence Paradigms in Healthcare–1, 225-253 (2007).
14. Zamzami, G., Ruiz, G., Goldgof, D., Kasturi, R., Sun, Y., Ashmeade, T.: Pain assessment in infants: Towards spotting pain expression based on infants' facial strain. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (Vol. 5, pp. 1-5). IEEE (2015).
15. Yoo, S. K., Lee, C. K., Park, Y. J., Kim, N. H., Lee, B. C., Jeong, K. S.: Neural network based emotion estimation using heart rate variability and skin resistance. In Advances in Natural Computation: First International Conference, ICNC 2005, Changsha, China, August 27-29, 2005, Proceedings, Part I 1 (pp. 818-824). Springer Berlin Heidelberg (2005).
16. Ranger, M., Gélinas, C.: Innovating in pain assessment of the critically ill: exploring cerebral near-infrared spectroscopy as a bedside approach. Pain Management Nursing, 15(2), 519-529 (2014).
17. Rahmati, H., Aamo, O. M., Stavdahl, Ø., Dragon, R., Adde, L.: Video-based early cerebral palsy prediction using motion segmentation. In 2014 36th annual international conference of the IEEE engineering in medicine and biology society (pp. 3779-3783). IEEE (2014).
18. Stahl, A., Schellewald, C., Stavdahl, Ø., Aamo, O. M., Adde, L., Kirkerod, H.: An optical flow-based method to predict infantile cerebral palsy. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 20(4), 605-614 (2012).
19. Zamzmi, G., Goldgof, D., Kasturi, R., Sun, Y., Ashmeade, T.: Machine-based multimodal pain assessment tool for infants: a review. arXiv preprint arXiv:1607.00331 (2016).
20. Parodi, E., Melis, D., Boulard, L., Gavelli, M., Baccaglini, E.: Automated newborn pain assessment framework using computer vision techniques. In Proceedings of the International Conference on Bioinformatics Research and Applications 2017 (pp. 31-36) (2017).
21. Di Bari A, Destrebecq A, Osnaghi F, Terzoni F.: Traduzione e validazione in italiano della scala Revised FLACC per la valutazione del dolore nel bambino con grave ritardo mentale. Pain Nurs Mag 2013;4 (2013).

22. Lugaresi, C., et al.: Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172 (2019).
23. Grishchenko, I., Bazarevsky, V.: Mediapipe holistic. https://ai.googleblog.com/2020/12/mediapipe-holistic-simultaneous-face.html (2020). Last accessed 20 Mar 2023
24. Kovac, J., Peer, P., Solina, F.: Human skin color clustering for face detection (Vol. 2, pp. 144-148). IEEE (2003).
25. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Incremental face alignment in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1859-1866) (2014).
26. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE (2001).