Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from Gaussian of Local Descriptors

Received 21 November 2018
Revised 23 April 2019
Accepted 15 May 2019

Sheryl Brahnam
*Computer Information Systems, Missouri State University,
Springfield, Missouri, USA*
Loris Nanni
DEI, University of Padua, Padua, Italy
Shannon McMurtrey
Management Information Systems, Drury University, Springfield, Missouri, USA
Alessandra Lumini
DISI, Università di Bologna, Cesena, Italy
Rick Brattin
*Computer Information Systems, Missouri State University,
Springfield, Missouri, USA*
Melinda Slack
Medical Director of Neonatology, Mercy Hospital, Cherokee, Missouri, USA, and
Tonya Barrier
*Computer Information Systems, Missouri State University,
Springfield, Missouri, USA*

Publishers note: The publisher wishes to inform readers that the article "Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from Gaussian of Local Descriptors" was originally published by the previous publisher of *Applied Computing and Informatics* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Brahnam, S., Nanni, L., McMurtrey, S., Lumini, A., Brattin, R., Slack, M., Barrier, T. (2020), "Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from Gaussian of Local Descriptors", *Applied Computing and Informatics*. Vol. ahead-of-print No. ahead-of-print. https://10.1016/j.aci.2019.05.003. The original publication date for this paper was 17/05/2019.

The authors wish to acknowledge the cooperation and generous assistance of all the nurses in the neonatal unit at Mercy Hospital in Springfield Missouri. We also gratefully acknowledge the support of NVIDIA Corporation with their donation of the Titan Xp GPU used in this research.

Conflicts of interest: None.

Applied Computing and
Informatics
Vol. 19 No. 1/2, 2023
pp. 122-143
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1016/j.aci.2019.05.003

Abstract

Diagnosing pain in neonates is difficult but critical. Although approximately thirty manual pain instruments have been developed for neonatal pain diagnosis, most are complex, multifactorial, and geared toward research. The goals of this work are twofold: 1) to develop a new video dataset for automatic neonatal pain detection called iCOPEvid (infant Classification Of Pain Expressions videos), and 2) to present a classification system that sets a challenging comparison performance on this dataset. The iCOPEvid dataset contains 234 videos of 49 neonates experiencing a set of noxious stimuli, a period of rest, and an acute pain stimulus. From these videos 20 s segments are extracted and grouped into two classes: pain (49) and nopain (185), with the nopain video segments handpicked to produce a highly challenging dataset. An ensemble of twelve global and local descriptors with a Bag-of-Features approach is utilized to improve the performance of some new descriptors based on Gaussian of Local Descriptors (GOLD). The basic classifier used in the ensembles is the Support Vector Machine, and decisions are combined by sum rule. These results are compared with standard methods, some deep learning approaches, and 185 human assessments. Our best machine learning methods are shown to outperform the human judges.

Keywords Bag-of-features, Texture descriptors, Machine learning, Support vector machine, Neonatal pain detection, Computer vision

Paper type Original Article

1. Introduction

Erroneous beliefs about the nociceptive physiology of neonates has greatly hindered research into neonatal pain. For many years, it was assumed that pain pathways were not fully developed in neonates [1]. Not only did this assumption inhibit research in this area until the mid-1980s, but it also justified inhumane surgeries performed on neonates without full anesthesia or postoperative pain interventions [2]. Not until studies revealed that nociceptive systems are fully developed in neonates did the scientific community take neonatal pain seriously [2]. Today it is widely recognized that infants experience acute physiologic, behavioral, metabolic, cortical, and hormonal responses to pain. Given that pain inhibitory pathways are not fully developed in infants, neonates may even experience greater levels of pain than do older children and adults [3].

More recent studies show that if left untreated, pain in neonates can have long-term effects on neurologic and behavioral development [4–8], especially in preterm infants [9–11]. Repeated unexpected acute pain experiences have been associated with slower weight gain [12], alterations in cerebral white matter and subcortical gray matter [13], and changes in the brain mechanisms involved in blood and oxygen supply [14,15]. These changes in the brain of neonates can result in learning and developmental disabilities as well as in lower IQ (for a review of the literature, see [16]). For all these reasons, it is critical that pain in neonates be detected and treated as early as possible.

Diagnosing pain in neonates is difficult, however. The clinical definition of pain assumes that the person experiencing pain has the ability to describe the location, duration, quality, and intensity of his or her pain experiences. Unable to speak, neonates must rely exclusively on the proxy judgments of others [17]. As a result, assessing pain in neonates is highly subjective and often results in misdiagnosis with the unwanted consequences of unnecessary interventions and failures to treat. Moreover, there is no universal method of evaluation [18] or consensus on the best proxy modalities (physiologic, behavioral, metabolic, or cortical) for assessing infant pain [19]. Recent research even suggests a discordance among, as well as within, specific modalities [20,21].

Most current pain instruments for neonatal pain assessment make use of two modalities: physiological indicators of pain (such as changes in heart and respiratory rates, blood pressure, vagal tone, and palmer sweating) [22] and behavioral indicators (such as body movement, crying, and facial expressions) [23]. The most popular validated scales are NIPS (Neonatal Infant Pain Scale) [24], PIPP (Premature Infant Pain Profile) [25], specifically, PIPP-R (Revised) [26], and DAN (Douleur Aiguë du Nouveau-né) [27]. Use of such scales is not without controversy, however [18]. Although approximately thirty neonatal pain instruments have been developed, most are complex, multifactorial, and geared towards research, with few having validated clinical utility. Complaints concern the complexity of the instruments (involving so many measurements that pain scores are delayed) and the inability of most instruments to provide a reliable indication of pain over time, which is clinically important, for example, for assessing chronic and repeated acute pain and the effectiveness and necessity of additional treatment. Studies have also shown problems with observer bias based on professional attitudes towards the instruments, desensitization due to overexposure to neonatal pain in the clinic, context, culture, gender, and perceived attractiveness of the neonate [28–31]. One way of overcoming the complexity of pain diagnosis and of reducing delays in treatment is to develop machine systems capable of automatically detecting pain in neonates. Not only would such systems help counter observer bias, but they would also offer the advantage of providing close monitoring of neonates over extended periods of time.

Research in automatic pain recognition began to take off about a decade ago with the research of 1) Brahnam et al. [32–36], who explored classifier systems to detect the facial expressions of pain in 204 static images of neonates experiencing stressful stimuli, 2) Barajas-Montiel and Reyes-Garía [37], who explored classifying cry states in 1623 samples, and 3) Pal et al. [38], who combined facial features with cry features from 100 samples to recognize different emotional states, including pain. In the last ten years, research in automatic pain detection has continued to focus on these two behavioral indicators of neonatal pain: facial expressions [32–36,39–46] and infant cries [38,47]. Little research to date has made use of the various physiological measures to detect pain [37,48–50], and none that we are aware of have involved neonates. The general consensus in the literature is that physiological measures alone are highly unreliable primarily because responses vary greatly among individual infants, but it is also difficult to distinguish physiological changes associated with pain from those associated with other distressful emotions, such as fear and anxiety [51] (for a recent review of the literature on physiological indices of acute pain in neonates, see [52]).

In this paper we focus on the automatic detection of pain in videos of neonates. Most research to date uses the iCOPE (infant Classification Of Pain Expressions) benchmark dataset [33–35], which contains 204 photographs of 26 neonates experiencing different noxious stimuli, rest, and an acute pain stimulus. This database was developed to detect pain experiences but was expanded by Gholami et al. [40,41] to classify pain intensity as well. The iCOPE dataset has been used to evaluate state-of-the-art single classifiers, such as Support Vector Machines (SVM) [33,34], Relevance Vector Machines [40,41], neural networks [32,34], ensembles fusing different texture descriptors, such as the Local Binary Pattern (LBP) and its variants [36,39], and handcrafted features combined with deep features [46]. Research using this dataset is limited in value, however, since the dataset fails to capture the dynamic patterns of facial expressions, which most likely provide important information for discriminating pain.

Recently, a couple of researchers have worked on motion-based pain detection. Zamzami et al. [42], for example, used optical flow, a well-known method of motion estimation, to calculate optical strain magnitudes in the facial tissues of ten infants video recorded during an acute procedure. SVM and K-nearest-neighbors (KNN) classifiers were used to classify segmented expressions into pain and nopain categories with 96% accuracy. In Sikka et al. [53], a method based on FACS (Facial Action Coding System) was used to estimate pain levels in the facial expressions of 50 youth experiencing both acute and chronic pain. An SVM produced an AUC of 0.84 for acute pain detection and 0.94 for chronic pain detection. A major drawback in these studies involves the datasets: the infant dataset contains a small number of samples of a small group of infants and the contrasting stimuli to pain may not have been sufficiently challenging; moreover, the larger child dataset contains no videos of neonates.

The main contributions of this study are two-fold. Our first goal was to develop a challenging database of neonatal pain video, which we call iCOPEvid, that would be made available to other researchers. This dataset contains video of 49 neonates experiencing a set of noxious stimuli, periods of rest, and an acute pain stimulus. Raw video images were segmented into 20 s of pain (49) and nopain (185), with the nopain video segments handpicked to produce a highly challenging dataset (we selected video segments that resemble pain expressions but are not stimulated by a pain stimulus). Our second goal was to develop a system for classifying the iCOPEvid segments into the two categories of pain and nopain and compare it with other state-of-the-art classifier systems, with the aim of producing a challenging classification rate for future comparisons.

The general outline of the classification approach proposed in this paper is as follows:

- The neonate's face is detected using Discriminative Response Map Fitting (DRMF), proposed in [54], which is based on a discriminative regression approach for the Constrained Local Models (CLMs) framework;
- The image is enhanced using [55], which is based on both an intra-layer optimization and inter-layer aggregation;
- Descriptors are used to represent each frame: the whole image is divided into four equal regions and from each region and the whole image a set of features is extracted and trained using an SVM;
- SVMs trained on these descriptors are fused;
- The scores of different frames of each video are then combined for classifying the video segments.

For improving the texture descriptors based on Bag-of-Features (BoF), we build an ensemble of codebooks for the well-known Bag-of-Words (BoW) method [56]. These codebooks are obtained using various strategies: different sizes for the patch, different PCA (Principle Component Analysis) projections [57] for reducing the size of the extracted feature vectors, and different patch shapes. As a result, a variety of global texton vocabularies are created. For each vocabulary, a different SVM is trained.

Each PCA projection retains different training subsets for building different projection matrices. This technique provides a method for building an ensemble of classifiers by varying the projection matrix. In addition, we proposed some variants of the Gaussian of Local Descriptors (GOLD) [58] for extracting features from an image, which is an improvement of BoW [56]. Our idea is to consider the covariance matrix obtained using the standard GOLD approach as an image that can be represented using standard texture descriptors. Different local descriptors, codebook generation methods, and subwindow configurations are combined to form an ensemble for detecting neonatal pain.

The remainder of this document is organized as follows. In Section 2 we provide a detailed description and justification of the study design, including details on how the raw data were collected, processed, selected, and organized. In Section 3 we describe twelve powerful texture descriptors. GOLD is described in Section 4, and BoF in Section 5. Our ensemble approach is presented in Section 6, and results comparing these approaches as well as several Convolutional Neural Network (CNN) approaches are presented in Section 7. In Section 8, we present 185 human assessments on the same dataset and describe this study in detail. Results in Sections 7 and 8 show that our best ensemble, which combines the best feature set with the GOLD approach, outperforms all other methods, including those of the human judges. We conclude in Section 9 with suggestions for future research. The iCOPEvid dataset is freely available and can be obtained by emailing the first author. All MATLAB source code used in our experiments is included in the dataset.

2. iCOPEvid pain video dataset

The objective of the iCOPEvid dataset was to obtain a representative yet highly challenging set of video sequences of neonatal facial expressions for evaluating automatic neonatal pain detection systems. In the medical literature, a number of noxious stimuli are typically compared in studies on neonatal pain: a pain-inducing stimulus (pinprick or puncture of a lancet) and various noxious stimuli, such as friction on the external lateral surface of the heel [59,60] and disturbances such as exposure to bright light [31] and diaper change [61]. These noxious stimuli are also compared with facial expressions in a resting state.

Following the designs of these studies, the first still image iCOPE database [32,34,35,62] collected a series of facial photographs of 26 infants (13 boys and 13 girls) responding to four stimuli: 1) the puncture of a heel lance, 2) friction on the external lateral surface of the heel, 3) transport from one crib to another, and 4) an air stimulus to provoke an eye squeeze (a facial characteristic of neonatal expressions of pain). After each stimulus, the infants were given a period of rest and photographs were collected during this period. Infants could be crying, awake, or asleep during the resting period. In addition, all infants were swaddled to obtain unobstructed images of each infant's face.

The goal of the iCOPEvid database was to obtain raw video of infants as they naturally appear in a hospital neonatal nursery, where some infants are swaddled and some are not and where flailing fists and legs often occlude facial expressions. The stimuli introduced in this study are also those commonly experienced by neonates in the nursery and include a state mandatory blood test requiring puncture of a heel lance (pain stimulus), friction on the external lateral surface of the heel administered by a cotton swab dabbed in alcohol in preparation of the heel lance, and recordings during and after two additional movement stimuli: transport from one crib to another and occasional physical disturbances. In addition, raw video was collected from neonates left undisturbed for a period of time (at rest) between stimuli administration. As with the iCOPE study design, videos were collected during the resting periods, and infants could be crying, awake, or asleep during these breaks between stimuli.

2.1 Subjects

This study complied with the protocols and ethical directives for research involving human subjects at Mercy Hospital in Springfield Missouri, USA. Informed consent was obtained from a parent, usually the mother in consultation with the father. Parents were recruited in the neonatal unit of Mercy Hospital shortly after delivery. Only mothers who had experienced uncomplicated deliveries were approached. Video images were collected from 49 neonates (26 boys and 23 girls) ranging in age from 34 to 70 h. Forty-one infants were Caucasian, one African American, one Korean, two Hispanic, and four interracial. Five males had been circumcised the day before and one male 43 h before the videos were taken. The last feeding time before video sessions ranged from 20 min to approximately 4 h. All infants were in good health.

2.2 Apparatus

All High Definition (HD) video was captured using a Sony HDR-XR520V HDD Camcorder following a white balance under ambient lighting conditions in a room separated from other newborns. The resolution of the video sequences, recorded as MP4 files, is 1920 × 1080 with a frame rate of 29.97 FPS.

2.3 Procedure

The facial expressions of the neonates were captured in one session. All stimuli were administered by an attending nurse. A total of 16 neonates were swaddled and 33 were allowed to move freely during and following stimulus administrations so that the infants were often obstructing their faces with their hands and feet or by twisting their bodies.

Following the requirements of standard medical procedures, videos of four stimuli and three resting periods were taken in the following sequence:

(1) Movement 1 Stimulus: Transport from one crib to another (Rest/Cry): After being transported from one crib to another, 1 m of video was taken. The state of the neonate was noted as either crying or resting.

(2) Resting 1: The infant was allowed to rest undisturbed for 1 m and video was taken of the infant; the infant's state during this period varied.

(3) Movement 2 Stimulus: The second movement stimulus was applied by periodically physically disturbing the neonate for 1 m. Video was taken during the disturbance and for 1 m following the application of the stimulus.

(4) Resting 2: The infant was allowed to rest undisturbed for 1 m and video was taken of the infant; the infant's state during this period varied.

(5) Friction Stimulus: The neonate received vigorous friction on the external lateral surface of the heel with cotton wool soaked in 70% alcohol for one full minute, and 1 m of video was taken as soon as the stimulus was applied.

(6) Resting 3: The infant was allowed to rest undisturbed for 1 m and video was taken of the infant; the infant's state during this period varied.

(7) Pain Stimulus: The external lateral surface of the heel was punctured for blood collection, and at least 1 m of video was taken, starting immediately after the introduction of the lancet and while the skin of the heel was squeezed for blood samples.

2.4 Image classes

A total of 234 video segments were extracted from the raw video footage and appropriately labeled for the iCOPEvid dataset. The first 20 s of video immediately following administration of the four noxious stimuli were extracted to form the pain/nopain categories, along with one to four 20 s resting segments that presented, where possible, facial expressions that varied and that ideally shared characteristics similar to pain expressions (e.g. yawning and crying). As illustrated in Figure 1, the resting segments were hand-selected by two of the investigators to provide difficult expressions to classify. The pain category contains 49 video segments (one for each infant) and the nopain category contains 185 video segments (several from each infant).

3. Evaluated texture descriptors

In this section, we describe twelve of the texture descriptors tested in this paper. Most are LBP-based [63] approaches, which depend on an analysis of patterns within local pixel neighborhoods; descriptors extracted from these neighborhoods are then aggregated to form a global representation for the whole image or for larger regions of an image [64]. The LBP variants outlined here include Local Ternary Patterns (LTP) [65], Local Phase Quantization (LPQ) [66], Binarized Statistical Image Features (BSIF) [67], Local Binary Pattern Histogram Fourier (LHF) [68], Rotation Invariant Co-occurrence among adjacent LBPs (RICLBP) [69], Extended Local Binary Patterns (ELBP) [70], and Local Configuration Pattern (LCP) [71].

In addition to LBP-based descriptors, we tested Histogram of Oriented Gradients (HOG) [72], Heterogeneous Auto-Similarities of Characteristics (HASC) [73], Morphological Features (MORPHO) [74], GIST [75], and the SIFT-like descriptor Laplacian Features (LF) [76].

Because the descriptors presented in this section have been extensively described in the literature and are familiar to most researchers in image/video classification, we only very briefly describe them in Table 1, where we also provide the parameter settings used for each descriptor. The two descriptors based on Bag-of-Words (BoW), Gaussian of Local Descriptors (GOLD) [58], and Bag-of-Features (BOF) [77], which our approach is based on, are extensively discussed in sections 4 and 6, respectively. We compare our approach with CNN-Based descriptors at the end of the experimental section. Thus, in this study, we compare and evaluate the main types of texture descriptors as outlined in Liu et al.'s [64] recent review of the literature on texture descriptors.

4. Gaussian of Local descriptors (GOLD)

In this paper we not only test the original Gaussian of Local Descriptors (GOLD) features [58] in our ensembles, but we also treat the covariance matrix C generated in the process (see Section 4.4 below) as an image from which we extract standard texture descriptors. GOLD [58] improves the BoW [77] approach, which is designed to handle viewpoint, illumination, occlusions, and inter-class variability. A bag of words is a sparse histogram over a vocabulary in a group of documents. In computer vision, a bag of *visual* words of features is a histogram over a vocabulary of local image features. The BoW descriptor extracts local features to generate a codebook that is then used to encode local features into codes that form a global image representation. The codebook generation step is based on vector quantization, typically through k-means clustering. Once BoW features are extracted, they are sent to a classifier for testing and training.

Basing the final step in BoW on vector quantization is not ideal because it binds the dataset characteristics too tightly to the feature representation, which creates a hidden dependency that is reflected in many specializations of BoW [79,80]. This has led researchers to investigate building codebooks that are more independent [79,80]. One way to avoid hidden dependencies between training set characteristics and the feature representation is simply to eliminate the quantization step. In [81] this is accomplished by modeling each set of vectors by a probability density function (pdf) that is compared with a kernel defined over the pdfs. Choosing a pdf with zero-mean Gaussian distribution, as proposed by [81], is equivalent to using average pooling, which is what was done in a method proposed by [82]. This insight led

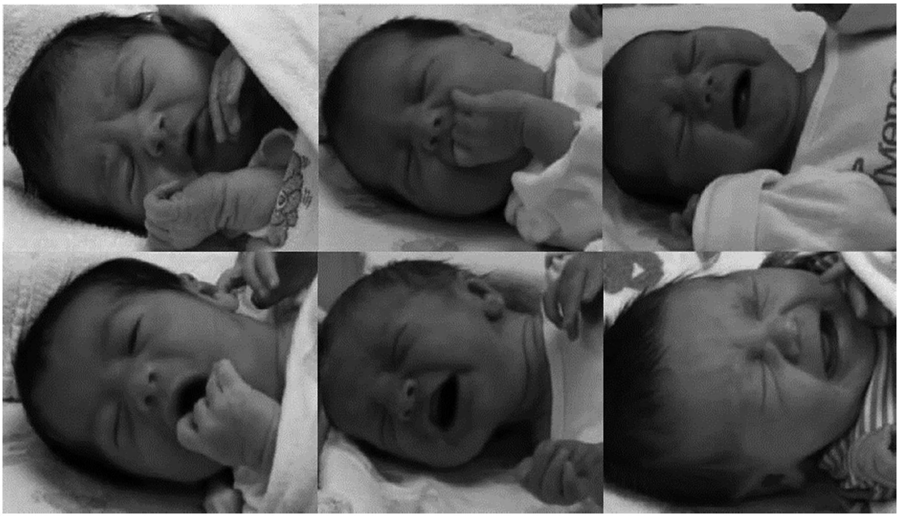

Figure 1.
Illustration of the challenge the iCOPEvid dataset presents for distinguishing pain (top) from resting (bottom). These still images were taken from video sequences of six different neonates (not swaddled).

Acronym	Brief Description of Descriptor and Parameter Settings	Source
LTP	Local Ternary Patterns is a three-value variant of LBP that addresses LBP's high sensitivity to noise in the near-uniform regions of LBP [78]). We used two (*radius, neighboring points*) configurations: (1, 8) and (2, 16). We tested both LTP with uniform bins (LTP-u) and LTP with rotation invariant uniform bins (LTP-r). Source code is available at: http://www.ee.oulu.fi/mvg/page/lbp_matlab	[65]
LPQ	Local Phase Quantization is blur robust variant of LBP. Different LPQ descriptors were evaluated in our experiments. Two were selected for our final system, both extracted by varying the radius parameter R($R = 3$ and $R = 5$) and each descriptor was used to train a separate SVM classifier. Source code for LPQ is available at http://www.ee.oulu.fi/mvg/download/lpq/)	[66]
GIST	Calculates the energy of a bank of Gabor-like filters, which are evaluated at eight orientations and four different scales. The square output of each filter is averaged on a 4 × 4 grid. GIST is evaluated at eight orientations and four scales	[75]
HOG	Histogram of Oriented Gradients calculates intensity gradients pixel to pixel. It then selects a corresponding histogram bin for each pixel based on the gradient direction. In our experiments we used a 2 × 2 version of HOG extracted on a regular grid at steps of 8 pixels. To form a more powerful descriptor, HOG features are stacked together considering sets of 2 × 2 neighbors	[72]
LF	Laplacian Features are a SIFT-like descriptor extracted using windows of different sizes (8 × 8 and 16 × 16). The multifractal spectrum (MFS) extracts the power-law behavior of the local feature distributions over the scale. Application of a multi-scale representation of the multi-fractal spectra, under a wavelet tight frame system, improves its robustness to changes in scale	[76]
BSIF	Binarized Statistical Image Features assigns an *n*-bit label to each pixel in an image using a set of *n* linear filters. Using Independent Component Analysis, the set of filters is estimated by maximizing the statistical independence of the filter responses of a set of patches from natural images	[67]
LHF	Local Binary Pattern Histogram Fourier is based on uniform Local Binary Patterns and is rotation invariant. The discrete Fourier transform is used to extract features that are invariant to rotation starting from the histogram rows of the uniform patterns	[68]
RICLBP	Rotation Invariant Co-occurrence among adjacent LBPs is an LBP variant that considers the spatial relations (the co-occurrence) among LBP patterns. RIC-LBP is rotation invariant for angles that are multiples of 45°. We computed RIC-LBP with three different configurations (LBP radius, displacement among LBPs): $(1, 2)$, $(2, 4)$ and $(4, 8)$	[69]
ELBP	Extended Local Binary Patterns is a variant of LBP that considers both intensity-based and difference-based descriptors. Intensity-based descriptors exploit the central pixel intensity or the intensities in the neighborhood. Difference-based descriptors consider the pixel that exploits the radial distance. ELBP contains 848 features. In this paper, we considered two neighborhoods (radius, pixels): $(1, 8)$ and $(2, 16)$	[70]
LCP	Local Configuration Pattern extracts two different levels of information: (1) local structural information (using LBP) and (2) microscopic configuration (MiC) information. MiC estimates the optimal weights to linearly reconstruct the central pixel intensity. This calculation exploits the intensities of the neighboring pixels and minimizes the reconstruction error. Rotation Invariance is achieved via the Fourier transform	[71]
MORPHO	Strandmark Morphological Features is a set of measures extracted from a segmented version of the image, including the aspect ratio, number of objects, area, perimeter, eccentricity, and other measures	[74]
HASC	Heterogeneous Auto-Similarities of Characteristics is simultaneously able to encode linear and nonlinear relations	[73]

Table 1. Summary of the basic texture descriptors evaluated in this study.

to the method proposed in [58], where a reference pdf that is a multivariate Gaussian distribution with mean and full variance is employed to obtain the new vector representation of GOLD, which uses the dot product to approximate a distance between distributions. GOLD has recently been shown to obtain state-of-the-art performance on many publicly available datasets [58].

The GOLD approach is a four-step process [58]:

Step 1. Feature extraction: dense SIFT descriptors are extracted on a regular grid of the input facial image;

Step 2. Spatial Pyramid Decomposition: the facial image is decomposed into subregions via multilevel recursive image decomposition. Features are softly assigned to regions based on a local weighting approach;

Step 3. Parametric probability density estimation: each region is represented as a multivariate Gaussian distribution of the extracted local descriptors through inferring local mean and covariance;

Step 4. Projection on the tangent Euclidean space: the covariance matrix is projected on the tangent Euclidean space and concatenated to the mean. This produces the final region descriptor.

Each of these steps, along with some variations applied in this work, are detailed more fully in sections 4.2–4.5 below.

4.1 Details step 1: feature extraction

In the original description of GOLD [58], feature extraction is performed by calculating SIFT descriptors (we use the function vl_phow from the vl_feat library [68]) or their color variations at four scales, defined by setting the width of the spatial bins to $\{4, 6, 8, 10\}$ pixels over a regular grid spaced by three pixels. Of course, features could be described using other methods; in this work, however, we use the features proposed in the original GOLD paper.

4.2 Details step 2: spatial pyramid decomposition

Spatial pyramid decomposition is accomplished by incrementally dividing a given image into increasingly smaller subregions, starting with level zero, where the decomposition is the entire image, followed by level one, where the image is subdivided into four quadrants, etc. A soft assignment of descriptors to these regions is performed that makes use of a weighting strategy that gives a weight to each descriptor that is based on its distance from the region's center. Given a region R, centered in (c_x, c_y) and with dimensions $R_w \times R_h$, and a local descriptor $D \in \mathfrak{R}^n$ computed at (d_x, d_y), its weighting function is computed as:

$$w(D, R) = \left(1 - \frac{d_x - c_x}{R_w}\right) \cdot \left(1 - \frac{d_y - c_y}{R_h}\right) \tag{1}$$

The function $w(D, R)$ belongs to the range [0,1] and is designed such that the descriptors extracted from the center of the region have the maximum weight, while descriptors placed on the borders between two regions are equally considered for both regions. In this work one level of decomposition is used.

4.3 Details step 3: parametric probability density estimation

Parametric probability density estimation uses the set of weighted descriptors belonging to each region to infer the parameters of a multivariate Gaussian distribution. If $D = \{D_1, \ldots, D_N\}$ is the set of weighed local descriptors of a region, $\mu \in \mathfrak{R}^n$ the mean vector, and $C \in \mathfrak{R}^{\mathrm{n}\times\mathrm{n}}$ the covariance matrix of a multivariate Gaussian distribution N belonging to a region, then:

$$N(\mathrm{D};\mu,\,\mathrm{C}) = \frac{1}{|2\mu C|^{2}} e^{-0.5(D-\mu)^{T}C^{-1}(D-\mu)} \tag{2}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} D_i \tag{3}$$

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(D_i-\mu)(D_i-\mu)^{T} \tag{4}$$

In this work we show that the covariance matrix *C* can be treated as an image so that standard texture descriptors can be used to describe it. Each of these feature vectors is fed into an SVM with a radial basis function kernel [83].

4.4 Details step 4: projection on the tangent Euclidean space

The parameters of μ and *C* provide a good representation of a region but are extremely high in dimensionality. To obtain a descriptor that can be fed into a classifier, the covariance matrix *C* is mapped into a point in the Euclidean space and concatenated to the mean μ so that the final region descriptor is a fixed length descriptor appropriate for linear classifiers based on the dot product.

Projection is performed in two steps: first, the covariance matrix *C* is projected on a Euclidean space tangent to the Riemannian manifold, at a specific tangency matrix *P*; second, the orthonormal coordinates of the projected vector are extracted. In [58] the best choice for *P* was determined to be the identity matrix *I* since the neighborhood relation between the points in the new space remain unchanged wherever the projection point P is located. Therefore, the projection formula of the covariance matrix *C* into a vector point *c* simply applies the vector operator to the standard matrix logarithm thus:

$$c = \nu ec(\log(I^{\frac{1}{2}}CI^{\frac{1}{2}})) \tag{5}$$

where log is the matrix logarithm operator and *vec* is the vector operator on the tangent space at identity, which for a symmetric matrix C is defined as $\nu ec(M) = [m_{1,1},\ \sqrt{2}\ m_{1,2},\ \sqrt{2}\ m_{1,3}, \ldots, m_{2,2},\ \sqrt{2}\ m_{2,3}, \ldots, m_{n,n}]$.

The final GOLD descriptor is size $(n^2 + 3n)/2$. This feature vector is fed into an SVM with a histogram kernel.

5. Bag-of-Features (BOF)

BOF [77], inspired by BoW [77], learns a visual vocabulary from extracted features and represents images by the frequencies of the extracted *visual words*. BOF feature extraction is basically a four-step process.

Step 1. Data organization and partitioning: the purpose of this step is to setup image category sets by first organizing a given image dataset into categories, and second by partitioning the dataset into training and testing sets.

Step 2. Feature extraction: in this step features (such as SIFT) are extracted from representative images in each category. There are many ways to extract features: sample features along a grid structure as with GOLD above, or extract features using interest point detectors or salient regions.

Step 3. Dictionary/vocabulary building: in this step a vocabulary of visual words is constructed using k-means clustering on the features extracted in Step 2 from the training set. The resulting k clusters are mutually exclusive, compact, and separated by similar characteristics. Each cluster center becomes a visual word, and the collection of centers (centroids) becomes the dictionary or codebook of visual words.

Step 4. Vector quantization: in this step an image can be quantified and abstractly represented using the dictionary by applying the following processes:

i. Extract features from the given image as in Step 2;

ii. For each feature, compute (using some distance measure) its nearest neighbor in the dictionary produced in Step 3;

iii. Take the set of nearest neighbor labels and build a histogram of size k. The histogram length corresponds to the number of visual words, and the bins are incremented based on the proximity of the descriptor to a particular cluster center. The histogram is the BOF descriptor.

In this work a simple BOF approach is used for codebook assignation: each descriptor is assigned to one codebook according to the minimum distance criterion, and a dense sample keypoint extraction is performed. SVM is used as the classifier. Since several feature descriptors are tested, several SVMs are trained, with classifier results combined using the sum rule (where the final score is the sum of the scores of a pool of classifiers that belong to an ensemble). Before fusion, the scores of each classifier are normalized to mean 0 and standard deviation 1.

To improve performance, an ensemble of codebooks is built using the following strategies:

- A different set of visual words is created for each class of the dataset, where for each NIMG image (images are clustered in groups, and the number of images, NIMG, in each group is 50 due to computational issues) visual words are built by clustering a local descriptor with k-means (the number k randomly selected between 10 and 40); for each descriptor the final vocabulary (codebook) is obtained by concatenating the visual words over all classes (this is done because k-means is an unstable clustering approach);
- A set of 50,000 patches are randomly extracted from the training set (considering the different classes), and these patches are then used to construct a PCA matrix (one projection matrix for each descriptor). This step is done three times: the first two times by retaining 99.9% of the variance and the third time by retaining 98% of the variance; a different codebook is created from each of these matrices and used to train a separate SVM;
- Each image is divided into overlapping patches with the size specified as a percentage (ps) of the original image taken at fixed steps $st = \min(ps \times l, ps \times h)/2$, where $l \times h$ is the size of the original image. We tested different values of ps. In our final version, we set $ps = 12\%$ and $ps = 8\%$;
- Different shapes of the extracted patches are used, both square and circular.

6. Proposed ensemble approach

In our proposed approach, a different SVM is trained for each descriptor. We used two different kernels in our experiments: 1) a histogram for BOF and the original GOLD algorithm, and 2) the radial basis function for all the other descriptors (including the image variants of GOLD).

In this section, we explain the five steps involved in our best performing approach. In outline form, they are the following:

- *STEP 1: FACE DETECTION.* For face detection we use Discriminative Response Map Fitting (DRMF), proposed in [54], which is a discriminative regression-based approach for the Constrained Local Models (CLMs) framework. The advantage of the DRMF is that it can be represented by a small set of parameters that can be efficiently used for reconstructing unseen response maps. The training procedure for the DRMF method can be broken down into two steps. In the first step, a dictionary is trained for the response map approximation so that it can be used for extracting the relevant features for learning the fitting of the update model. The second step involves iteratively learning the fitting update model, which is achieved by a modified boosting procedure.
- *STEP 2: ENHANCEMENT.* This step enhances the contrast of each image using the method proposed in [55], which is based on both an intra-layer optimization and inter-layer aggregation. A 2-D histogram $h(k, k+1)$is extracted by counting pairs of adjacent pixels with gray-levels k and $k+1$ using a tree-like layered structure. At each layer, a constrained optimization problem is formulated for enhancement and solved to obtain a difference vector. The difference vectors at all layers are then aggregated into a single unified difference vector.
- *STEP 3: DESCRIPTORS.* The whole image is divided into four equal regions; and, from each region, a set of features is extracted; moreover, a set of features is extracted from the whole image. From each set of features, a different SVM is trained, then the five SVMs are combined by sum rule.
- *STEP 4: GOLD DESCRIPTORS.* The descriptors described in Section 4 are extracted.
- *STEP 5: FUSION.* Fusion by weighted sum rule of the SVMs trained in steps 3–5. Notice that before each fusion the scores of each classifier are normalized to mean 0 and standard deviation 1.

7. Experimental results

We use the Area Under the receiver operating characteristic Curve (AUC) [84], which is a plot of the sensitivity vs. false positives ($1-$ specificity), as the performance indicator for all tests (including the evaluation of pain videos by human subjects in Section 8), where AUC is expressed as a percentage of the area of the unit square (i.e., as a value in the [0,100] range).

The testing protocol adopted was a 10-fold cross-validation (since we have 49 neonates, each fold contains the video of five neonates with one-fold containing four). All the frames of a given video segment belonged either to the training set or to the testing set.

The first experiment, reported in Tables 2 and 3, was aimed at comparing the descriptors listed in Section 3. We report the performance using two different testing protocols:

- Images: the AUC is calculated considering each frame as a different still image;
- Video: the AUC is calculated considering each video as a *single pattern*, where the score of each video is the *average of the scores* of the frames that belong to that video.

In Table 2 the method named *FUSg* is the fusion by sum rule among *RICLBP*, *LCP*, *BSIF*, *MORPHO*, and *LPQ*. We tested all the texture descriptors mentioned in Section 3, but in order to avoid a huge table cluttered with unnecessary details regarding those descriptors that performed poorly, we have reported only the highest performing approaches here.

The cells of the row labeled *Image* contain two values: the first is the AUC where features are extracted from the whole image only (labeled *WholeIm*), and the second is the AUC considering features extracted from the four subwindows (as detailed in Section 6 and labeled *SubIm*). Clearly, the best performance is obtained when considering the subwindows; for this reason, in the row labeled *Video*, only the performance obtained considering the subwindows is reported.

Due to the high computation time of BOF, it was tested using only four features. In Table 3 we compare several BOF variants for demonstrating the value of combining different codebooks:

- *Whole*: these are features extracted from the whole image, and both the values of *ps* and only one PCA projection (98% of the variance) are used for building the codebooks. Since only square patches are considered, the number of codebooks is two.
- *Itera*: these are features extracted as described with *Whole*, but all three PCA projections described in Section 5 are used for building the codebooks. Since only square patches are considered, the number of codebooks is six.
- *Circular*: these are features extracted as described with *Itera,* but both square and circular patches are used. To reduce computation time, when the circular patches are extracted, only one PCA projection (98% of the variance) is used for building the codebooks. Since both *ps* values are used, the number of codebooks is eight.
- *Sub*: these are features extracted as described with *Circular,* but the codebooks are built from the four subwindows as well as from the full image; thus, the number of codebooks is forty-five.
- *Sub_easy*: these are features extracted as described with *Whole*, but the codebooks are built from the four subwindows as well as from the full image; thus, the number of codebooks is ten.

		Descriptors							
Protocol	Type	LPQ	RICLBP	LCP	BSIF	ELBP	HASC	MORPHO	FUSg
Image	WholeIm	58.6	56.9	56.3	59.0	51.9	56.6	55.8	**62.5**
	SubIm	62.8	63.0	58.2	62.9	53.9	61.0	62.1	**67.4**
Video	SubIm	68.9	**76.1**	72.4	70.5	65.8	64.2	70.1	74.9

Table 2.
Performance (AUC) of texture descriptors.

	Descriptors					
Protocol	BOF Type	HF	LBP	LPQ	HOG	FUSbof
Image	Whole	58.9	57.9	62.0	61.9	65.6
	Itera	61.9	61.3	65.1	65.0	68.6
	Circular	62.9	62.4	66.1	65.6	69.3
	Sub	68.9	67.7	71.2	71.0	73.5
	Sub_easy	66.7	65.6	69.3	68.5	70.1
Video	Sub	73.8	74.1	75.8	75.0	76.4

Table 3.
Performance (AUC) of the BOF approaches.

Clearly BOF outperforms the global approaches, as does combining different codebooks. The fusion among all four descriptors, labeled *FUSbof*, which is based on different features, also improves performance.

In the last row we report the performance obtained by Sub using the Video testing protocol. Obviously, combing the four different descriptors by sum rule, i.e. FUSbof, improves performance.

In Table 4 we report the performance of GOLD and its variants. TB is the performance obtained by SVM trained using the texture descriptors extracted from the covariance matrix (see Section 4.4). TB_SUM is the sum rule among the different TBs. The fusions $a \times \mathrm{A} + b \times \mathrm{B}$ are the weighted sum rule between A, with weight a, and B, with weight b. In the last row of Table 3, we combine the approaches based on GOLD and the approaches based on TB.

Although the improvement is not impressive, it is clear that the idea of extracting texture features from the covariance matrix works (see [85]) and that further tests should be performed for optimizing the results.

The novel ensemble proposed here is given by the sum rule fusion among FUSg, FUSbof and (2 × GOLD + TB_SUM) to obtain an AUC (video classification rate) of 0.80. Note: before fusion, the scores of each method are normalized to mean 0 and standard deviation 1.

In an attempt to improve our best ensemble method, we used a powerful feature selection approach, Maximum Relevance Minimum Redundancy [86], to select a subset of all the features extracted by each texture descriptor. Table 5 reports the performance of 2 × GOLD + TB_SUM by varying the percentage of retained features and shows that feature selection did not significantly improve the performance; however, it did reduce the number of features that could be used to obtain a competitive performance.

In addition to the above methods, in Table 6 we show our results using the following transfer learning approaches using both the Image and Video protocols:

- VGGface_e; based on [87], which combines LBP and HOG features with deep learning-based features extracted from two pretrained CNNs (VGG face [88] and MBPCNN [89]);

		Image	Video
GOLD	69.1	78.4	
TB	LPQ	63.9	72.7
	RICLBP	66.1	75.1
	HOG	65.9	74.1
	HASC	66.1	74.3
	BSIF[1]	67.5	76.3
	SUM	67.6	75.5
3 × GOLD + TB_SUM		69.8	78.9
2 × GOLD + TB_SUM		70.0	**79.0**

[1] Five different BSIF are trained and then combined by sum rule, we use 5 filters of different sizes: 3, 5, 7, 9, 11.

Table 4. Performance of GOLD and fusion with other descriptors.

Table 5. Performance as a function of percentage of retained descriptors using Maximum Relevance Minimum Redundancy [86] on the image protocol.

25%	50%	75%	100%
65.1	68.2	70.7	70.0

this approach was used in [32] to assess neonatal pain using the iCOPE still image dataset. We combine LBP and HOG features with VGGface_e to test them on the iCOPEvid dataset;

- Dense: Densely Connected Convolutional Network (Densenet) [90] trained on the ImageNet [91] database and tuned on our training set;
- ResNet: This net is the same as ResNet-34 [92] except that it has fewer layers and 50% fewer filters per layer. ResNet has obtained the state-of-the-art classification performance on the LFW face benchmark [93] and was trained from scratch on 3 million faces. We used ResNet as a feature extractor for feeding into an SVM. Resnet is available at https://github.com/davisking/dlib-models.

Adding the CNN to our best performing ensemble (the fusion by sum rule among FUSg, FUSbof and 2 × GOLD + TB_SUM) did not significantly improve results. Our GOLD-based ensemble, therefore, should provide researchers with a challenging performance rate for future comparisons.

8. Human judgments

To compare our system with human evaluations, judgments were collected from college students enrolled in upper-level courses at a large Midwestern university. A total of 185 students (56.8% female) participated in the study to completion. The average participant was 24.8 years of age (SD = 7.3), with 82.7% single and never married. While only 12.4% were parents, most participants (71.9%) indicated some history caring for infants (age: 0–1) of their own, of family members (including siblings), or of friends. The average length of time caring for infants was 3.3 years (SD = 4.2). None of the participants had received formal neonatal pain assessment training. However, previous studies have indicated that clinician judgments of pain from facial expressions are as reliable as judgments from those who are not clinicians [94].

Ratings were collected online through a website developed specifically for this study. Participants were presented with all 234 videos in the iCOPEvid dataset. All videos were 20 s in length and did not include sound. Each video was randomly presented to participants, and they were asked to use their best judgment to rate the level of pain experienced by the neonate in each video. Participants were unaware of the actual pain or nopain condition experienced by the neonate. Participants indicated pain levels with a slide control anchored by "Absolutely No Pain" and "Extreme Pain." Responses were converted to a numeric value ranging from 0 to 100. Due to the large amount of time required to complete the survey, participants were allowed to exit the website and resume later at the point where they had left off. A total of 39,865 pain assessments were collected in this manner.

The AUC of human subjects, with a 95% confidence interval, is lower = 0.665/ upper = 0.677, a performance that is significantly lower than our best ensemble. Human subjects found it most difficult to discern pain versus nopain in the presence of the friction

Table 6. Performance using transfer learning with three pretrained CNNs and fusion with our best performing method.

		Image	Video
Transfer learning approaches	VGGface_e	66.7	74.3
	Dense	65.1	73.0
	ResNet	58.52	66.2
FUSg + FUSbof + 2 × GOLD + TB_SUM + VGGface_e		70.3	**79.8**
FUSg + FUSbof + 2 × GOLD + TB_SUM + Dense		70.2	78.5

stimulus. Statistical analysis using a Z-score of subject performance across individual stimuli confirm these differences. A Z-score indicates significant differences between two AUC values calculated from independent samples [95]. The critical ratio Z is defined as:

$$Z = \frac{|AUC_1 - AUC_2|}{\sqrt{SE^2_{AUC1} + SE^2_{AUC2}}} \tag{6}$$

where SE_{AUC1} refers to the standard error for AUC_1 and SE_{AUC2} refers to the standard error for AUC_2. The value of Z is compared to a critical value CV_Z at the preferred level of significance referenced in tables of the normal distribution. A value of Z greater than CV_Z is considered evidence that the AUCs are statistically different. Alternatively, one may interpret the p-value for the statistic. If the p-value is less than the desired level of significance (e.g., $p < .05$) then one may reasonably assume that the two AUCs are different.

Human subject AUCs for each stimuli measured individually against pain are: rest (0.699), movement (0.693), and friction (0.600). We find that the AUC for *friction* stimuli is statistically different from the AUCs for *rest* ($Z = 15.54$, $p < .001$) and *movement* ($Z = 13.44$, $p < .001$). This supports the poor results of the human subjects in their assessment of the *friction* stimulus. Removing the friction videos from the sample set significantly improves the overall AUC rating of human subjects from 0.671 to 0.697 ($Z = 4.64$, $p < .001$). An improvement is obtained as well as in the AUC rating of our best ensemble, which went from 0.800 to 0.815 (the CNN-based method increased from 66.7 to 67.5 in the Image set and from 74.3 to 75.5 in the Video set). A possible explanation for these differences is that some neonates in our study experienced pain when receiving the *friction* stimulus.

In Figures 2 and 3, we show some images that proved very difficult to classify, not only by machine but also by the human judgers.

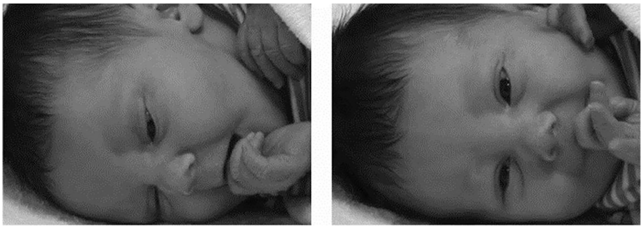

Figure 2.
Difficult pain frames; human rating of this subject (not just these images) 8.1% correct; best machine rating for this subject 13.8% correct.

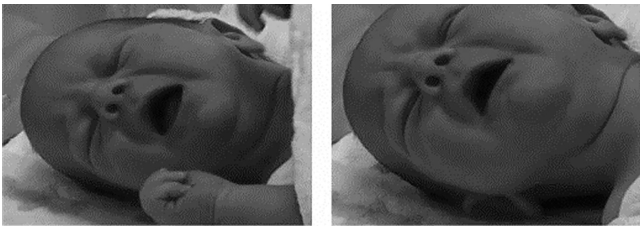

Figure 3.
Difficult nopain frames; human rating of this subject (not just these images) 35.2% correct; best machine rating for this subject 43.1% correct.

9. Conclusion

The aim of this work was 1) to develop a challenging video dataset for the detection of pain in videos of neonates, and 2) to develop a powerful classification system that would provide a challenging performance rate for future comparisons. For the video dataset, we collected 234 videos of 49 neonates experiencing a set of noxious stimuli, a period of rest, and an acute pain stimulus, and divided the videos into two classes: pain and nopain. We purposefully extracted a set of challenging videos for the nopain class.

We experimentally developed our classification system by comparing a number of SVMs trained on twelve different state-of-the-art texture descriptors, along with an ensemble of local approaches based on Bag-of-Features (BOF) and descriptors based on Gaussian of Local Descriptors (GOLD). A set of features were extracted from the covariance matrix generated in the GOLD process with the assumption that the different features extracted from different image patches are taken from a multivariate Gaussian distribution. An ensemble of global descriptors and BOF was then used to improve the performance of the new descriptors based on GOLD.

For improving the performance with respect to some standard approaches (which we also report), we proposed a set of features that could be extracted from the covariance matrix used in GOLD. To improve the texture descriptors based on BOF, we built an ensemble of codebooks for BoW using different strategies. The results of our best ensemble proved superior not only to the standard approaches but also to some deep learning approaches. Our best ensemble was also shown to outperform 185 human judges.

Although adding CNN to our best ensemble failed to improve results, additional studies using CNN are needed. In the future, we plan on exploring more CNN-based approaches. For example, to overcome the problem of a small dataset, we want to collect different datasets of infant faces and train them using a CNN network; we will then explore fine-tuning our neonatal pain dataset using that network.

Similarly, although the improvement using the covariance matrix was not impressive, it is clear that the idea of extracting texture features from the covariance matrix works (see [85]) and that further tests should be performed for optimizing results. Finally, we plan on extracting features from the entire set of images to see if this will enhance performance.

References

[1] N.L. Schechter, C.B. Berde, M. Yaster, Pain in infants, children and adolescents: an overview, in: N.L. Schechter, C.B. Berde, M. Yaster (Eds.), Pain in Infants, Children and Adolescents, William & Wilkins, Baltimore, 1993, pp. 3–10.

[2] J. Baños, G. Ruiz, E. Guardiola, Analysis of articles on neonatal pain published from 1965 to 1999, Pain Res. Manage. 6 (2001) 45–50.

[3] F. Schwaller, M. Fitzgerald, The consequences of pain in early life: injuryinduced plasticity in developing pain pathways, Eur. J. Neurosci. 39 (2014) 344–352.

[4] K.J. Anand, J.V. Aranda, C.B. Berde, S. Buckman, E.V. Capparelli, W. Carlo, P. Hummel, C.C. Johnston, J. Lantos, V. Tutag-Lehr, A.M. Lynn, L.G. Maxwell, T.F. Oberlander, T.N. Raju, S.G. Soriano, A. Taddio, G.A. Walco, Summary proceedings from the neonatal pain-control group, Pediatrics 117 (2006) S9–S22.

[5] R. Grunau, Long-term effects of pain in children, in: P. McGrath, B. Stevens, S. Walker, S. Zempsky (Eds.), Oxford Textbook of Paediatric Pain, Oxford University Press, 2013.

[6] R.E. Grunau, M.F. Whitfield, J. Petrie-Thomas, A.R. Synnes, I.L. Cepeda, A. Keidar, M. Rogers, M. MacKay, P. Hubber-Richard, D. Johannesen, Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants, Pain 143 (2009) 138–146.

[7] F. Denk, S.B. McMahon, I. Tracey, Pain vulnerability: a neurobiological perspective, Nat. Neurosci. 17 (2014) 192–200.

[8] S. Brummelte, R.E. Grunau, V. Chau, K.J. Poskitt, R. Brant, J. Vinall, A. Gover, A.R. Synnes, S.P. Miller, Procedural pain and brain development in premature newborns, Ann. Neurol. 71 (2012) 385–396.

[9] R.V.E. Frunau, Long-term consequences of pain in human neonates, in: J.K.S. Anand, B.J. Stevens, P.J. McGrath (Eds.), Pain in neonates, Elsevier, Amsterdam, 2007, pp. 55–76.

[10] M. Ranger, C.M.Y. Chau, A. Garg, T.S. Woodward, M.F. Beg, B. Bjornson, K. Poskitt, K. Fitzpatrick, A.R. Synnes, S.P. Miller, R.E. Grunau, Neonatal painrelated stress predicts cortical thickness at age 7 years in children born very preterm, PLoS ONE 8 (2013).

[11] B.O. Valeri, L. Holsti, M.B.M. Linhares, Neonatal pain and developmental outcomes in children born preterm: a systematic review, Clin. J. Pain 31 (2015) 355–362.

[12] J. Vinall, S.P. Miller, V. Chau, S. Brummelte, A.R. Synnes, R.E. Grunau, Neonatal pain in relation to postnatal growth in infants born very preterm, Pain 153 (2012) 1374–1381.

[13] R.E. Grunau, Long-term consequences of pain in human neonates, in: K.J.S. Anand, B.J. Stevens, P.J. McGrath (Eds.), Pain in neonates: 2nd revised and enlarged edition, Elsevier, New York, 2000, pp. 55–76.

[14] A.T. Bhutta, K.J.S. Anand, Vulnerability of the developing brain, Clin. Perinatol. 29 (2002) 357–372.

[15] K.J.S. Anand, F.M. Scalzo, Can adverse neonatal experiences alter brain development and subsequent behavior?, Neonatology 77 (2000) 69–82.

[16] T. Field, Preterm newborn pain research review, Infant Behav. Dev. 49 (2017) 141–150.

[17] D. Wong, C. Baker, Pain in children: comparison of assessment scales, Pediatr Nurs 14 (1988) 9–17.

[18] N. Witt, S. Coynor, C. Edwards, H. Bradshaw, A guide to pain assessment and management in the neonate, Curr. Emergency HospitalMed. Rep. 4 (2016) 1–010.

[19] I.A. Rouzan, An analysis of research and clinical practice in neonatal pain management, J. Am. Acad. Nurse Pract. 13 (2001) 57–60.

[20] R. Slater, A. Cantarella, L. Franck, J. Meek, M. Fitzgerald, How well do clinical pain assessment tools reflect pain in infants?, PLoS Med 5 (2008).

[21] R.P. Riddell, D.B. Flora, S.A. Stevens, B. Stevens, L.L. Cohen, S. Greenberg, H. Garfield, Variability in infant acute pain responding meaningfully obscured by averaging pain responses, Pain 154 (2013) 714–721.

[22] S. Coffman, Y. Alvarez, M. Pyngolil, R. Petit, C. Hall, M. Smyth, Nursing assessment and management of pain in critically ill children, Heart Lung 26 (1997) 221–228.

[23] P.A. McGrath, Pain in Children: Nature, Assessment, and Treatment, Guildford Press, New York, 1989.

[24] J. Lawrence, D. Alcock, P. McGrath, J. Kay, S.B. MacMurray, C. Dulberg, The development of a tool to assess neonatal pain, Neural Netw. 12 (1993) 59–66.

[25] B. Stevens, C. Johnston, P. Petryshen, A. Taddio, Premature infant pain profile: development and initial validation, Clin. J. Pain 12 (1996) 13–22.

[26] S. Gibbins, B.J. Stevens, J. Yamada, K. Dionne, M. Campbell-Yeo, G. Lee, K. Caddell, C. Johnston, A. Taddio, Validation of the premature infant pain profilerevised (PIPP-R), Early Human Dev. 90 (2014) 189–193.

[27] R. Carbajal, A. Paupe, E. Hoenn, R. Lenclen, M. Olivier Martin, DAN: une echelle comportementale d'evaluation de la douleur aigue du nouveau-ne, Archives Pédiatrie 4 (1997) 623–628.

[28] C. Miller, S.E. Newton, Pain perception and expression: the influence of gender, personal self-efficacy, and lifespan socialization, Pain Manage. Nurs. 7 (2006) 148–152.

[29] R.R.P. Riddell, M.A. Badali, K.D. Craig, Parental judgments of infant pain: importance of perceived cognitive abilities, behavioural cues and contextual cues, Pain Res. Manage. 9 (2004) 73–80.

[30] K.M. Prkachin, P. Solomon, T. Hwang, S.R. Mercer, Does experience influence judgments of pain behaviour? Evidence from relatives of pain patients and therapists, Pain Res Manage. 6 (2001) 105–112.

[31] R. Xavier Balda, R. Guinsburg, M.F.B. de Almeida, C.P. de Araujo, M.H. Miyoshi, B.I. Kopelman, The recognition of facial expression of pain in full-term newborns by parents and health professionals, Arch. Pediatr. Adolesc. Med. 154 (2000) 1009–1016.

[32] S. Brahnam, C.-F. Chuang, R. Sexton, F.Y. Shih, M.R. Slack, Machine assessment of neonatal facial expressions of acute pain, Decis. Support Syst. 43 (2007) 1247–1254.

[33] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.R. Slack, Svm classification of neonatal facial images of pain, in: WILF 2005 6th International Workshop on Fuzzy Logic and Applications, Crema, Italy, 2005.

[34] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.R. Slack, Machine recognition and representation of neonate facial displays of acute pain, Int. J. Artif. Intell. Med. (AIIM) 36 (2006) 211–222.

[35] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.R. Slack, Svm classification of neonatal facial images of pain, in: I. Bloch, A. Petrosino, A.G.B. Tettamanzi (Eds.), Fuzzy Logic and Applications (revised selected papers from the 6th International Workshop,WILF 2005, Crema, Italy, September 15-17, 2005), 2006, pp. 121–128.

[36] S. Brahnam, L. Nanni, R. Sexton, Introduction to neonatal facial pain detection using common and advanced face classification techniques, in: J. Lakhmi (Ed.), Computational Intelligence In Healthcare, Springer-Verlag, New York, 2007.

[37] S.E. Barajas-Montiel, C.A. Reyes-García, Fuzzy Support Vector Machines for Automatic Infant Cry Recognition, in: D.-S. Huang, K. Li, G.W. Irwin (Eds.) ICIC 2006, 2006, pp. 876–881.

[38] P. Pal, A.N. Iyer, R.E. Yantorno, Emotion Detection From Infant Facial Expressions And Cries, in: IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006.

[39] L. Nanni, S. Brahnam, A. Lumini, A local approach based on a Local Binary Patterns variant texture descriptor for classifying pain states, Expert Syst. Appl. 37 (2010) 7888–7894.

[40] B. Gholami, W.M. Haddad, A.R. Tannenbaum, Agitation and pain assessment using digital imaging, in: IEEE Engineering in Medicine and Biology Society, 2009, pp. 2176–2179.

[41] B. Gholami, W.M. Haddad, A.R. Tannenbaum, Relevance vector machine learning for neonate pain intensity assessment using digital imaging, IEEE Trans. Biomed. Eng. 57 (2010).

[42] G. Zamzami, G. Ruiz, D. Goldgof, R. Kasturi, Y. Sun, T. Ashmeade, Pain assessment in infants: Towards spotting pain expression based on infants' facial strain, in: International Conference and Workshops on Automatic Face and Gesture Recognition, 2015.

[43] E. Fotiadou, S. Zinger, W.E. Tjon a Ten, S. Bambang Oetomo, P.H.N. De With, Video-based discomfort detection for infants using a Constrained Local Model, IS&T/SPIE Imaging International Society for Optics and Photonics, 2016.

[44] C. Li, S. Zinger, W.E. Tjon a Ten, S. Zinger, P.H.N. De With, Video-based discomfort detection for infants using a Constrained Local Model, International Conference on Systems, Signals and Image Processing, IWSSIP, 2016.

[45] M.N. Mansor, M.N. Rejab, A computational model of the infant pain impressions with Gaussian and nearest mean classifier, in: IEEE International Conference on Control System, Computing and Engineering (ICCSCE), IEEE, 2013, pp. 249–253.

[46] L. Celona, L. Manoni, Neonatal facial pain assessment combining hand-crafted and deep features, British Machine Vision Conference (BMVC) (2015).

[47] R.R. Vempada, B. Siva Ayuappa Kumar, K.S. Rao, Characterization of infant cries using spectral and prosodic features, in: National Conference on Communications (NCC), Kharagpur, 2012, pp. 1–5.

[48] T.A. Olugbade, N. Bianchi-Berthouze, N. Marquardt, A.C. Williams, Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain, 2015 IEEE International Conference on Affective Computing and Intelligent Interaction (ACII) Xian, 2015.

[49] Y. Chu, X. Zhao, J. Han, Y. Su, Physiological signal-based method for measurement of pain intensity, Front. Neurosci. (2017).

[50] M. Kachele, P. Thiam, M. Amirian, F. Schwenker, G. Palm, Methods for personcentered continuous pain intensity assessment from bio-physiological channels, IEEE J. Sel. Top. Signal Process. 10 (2016) 854–864.

[51] L. Van Cleve, L. Johnson, P. Pothier, Pain responses of hospitalized infants and children to venipuncture and intravenous cannulation, J. Pediatr. Nurs. 11 (1996) 161–168.

[52] J.A. Waxman, R.R.P. Riddell, P. Tablon, L.A. Schmidt, A. Pinhasov, Development of cardiovascular indices of acute pain responding in infants: a systematic review, Pain Res. Manage. 2016 (2016) 15.

[53] K. Sikka, A.A. Ahmad, D. Diaz, M.S. Goodwin, K.D. Craig, M.S. Bartlett, J.S. Huang, Automated assessment of children's postoperative pain using computer vision, Pediatrics 136 (2015) e124–e131.

[54] A. Asthana, S. Zafeiriou, S. Cheng, M. Pantic, Robust discriminative response map fitting with constrained local models, CVPR (2013) 3444–3451.

[55] C. Lee, C. Lee, C.-S. Kim, Contrast enhancement based on layered difference representation of 2d histograms, IEEE Trans. Image Process. 22 (2013) 5372–5384.

[56] J. Sivic, A. Zisserman, Video google: a text retrieval approach to object matching in videos, in: IEEE International Conference on Computer Vision (ICCV '03), Washington, DC, 2003, pp. 1470–1477.

[57] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991) 71–86.

[58] G. Serra, C. Grana, M. Manfredi, R. Cucchiara, Gold: Gaussians of local descriptors for image representation, Comput. Vis. Image Underst. 134 (2015) 22–32.

[59] C.E. Izard, R.R. Huebner, D. Risser, G.C. McGinnes, L.M. Dougherty, The young infant's ability to produce discrete emotion expressions, Dev. Psychol. 16 (1980) 418–426.

[60] R.V.E. Grunau, C.C. Johnston, K.D. Craig, Neonatal facial and cry responses to invasive and non-invasive procedure, Pain 42 (1990) 295–305.

[61] F. Warnock, D. Sandrin, Comprehensive description of newborn distress behavior in response to acute pain (newborn male circumcision), Pain 107 (2004) 242–255.

[62] S. Brahnam, L. Nanni, R. Sexton, Neonatal facial pain detection using NNSOA and LSVM, in: The 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'08) Las Vegas, NV, 2008, pp. 352–357.

[63] S. Brahnam, L.C. Jain, A. Lumini, L. Nanni, Local Binary Patterns: New Variants and Applications, Springer-Verlag, Berlin, 2014.

[64] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, M. Pietikäinen, From bow to CNN: Two decades of texture representation for texture classification, arXiv, 1801.10324v (2018) 1-28.

[65] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Analysis and Modelling of Faces and Gestures, LNCS 4778 (2007) 168–182.

[66] V. Ojansivu, J. Heikkila, Blur insensitive texture classification using local phase quantization, ICISP (2008) 236–243.

[67] J. Kannala, E. Rahtu, Bsif: Binarized statistical image features, in: 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba, Japan, 2012, pp. 1363–1366.

[68] T. Ahonen, M. Pietikäinen, Soft histograms for local binary patterns, in: Finnish Signal Processing Symposium (FINSIG 2007), Oulu, Finland, 2007.

[69] R. Nosaka, K. Fukui, HEp-2 cell classification using rotation invariant cooccurrence among local binary patterns, Pattern Recognit. Bioinf. 47 (2014) 2428–2436.

[70] L. Liu, L. Zhao, Y. Long, G. Kuang, P. Fieguth, Extended local binary patterns for texture classification, Image Vis. Comput. 30 (2012) 86–99.

[71] Y. Guo, G. Zhao, M. Pietikainen, Texture classification using a linear configuration model based descriptor, British Machine Vision Conference (2011) 1–10.

[72] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 9th European Conference on Computer Vision, San Diego, CA, 2005.

[73] M. San Biagio, M. Crocco, M. Cristani, S. Martelli, V. Murino, Heterogeneous auto-similarities of characteristics (hasc): Exploiting relational information for classification, in: IEEE Computer Vision (ICCV13), Sydney, Australia, 2013, pp. 809–816.

[74] P. Strandmark, J. Ulén, F. Kahl, HEp-2 Staining Pattern Classification, International Conference on Pattern Recognition (ICPR2012), 2012.

[75] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vision 42 (2001) 145–175.

[76] Y. Xu, S. Huang, H. Ji, C. Fermüller, Scale-space texture description on SIFT-like textons, Comput. Vis. Image Underst. 116 (2012) 999–1013.

[77] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, Visual categorization with bags of keypoints, in: ECCV International Workshop on Statistical Learning in Computer Vision, 2004, pp. 59–74.

[78] T. Ojala, M. Pietikainen, T. Maeenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.

[79] T. Tuytelaars, C. Schmid, Vector quantizing feature space with a regular lattice, IEEE International Conference on Computer Vision (2007) 1–8.

[80] A. Arandjelović, A. Zisserman, All about VLAD, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2013, pp. 1578–1585.

[81] J. Farquhar, S. Szedmak, H. Meng, J. Shawe-Taylor, Improving "bag-of keypoints" Image Categorisation: Generative models and pdf-kernels, University of Southampton, 2005.

[82] J. Carreira, R. Caseiro, J. Batista, C. Sminchisescu, Semantic segmentation with second-order pooling, European Conference on Computer Vision (2012) 430–443.

[83] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[84] T. Fawcett, ROC Graphs: Notes and Practical Considerations for Researchers, HP Laboratories, Palo Alto, USA, 2004.

[85] L. Nanni, M. Paci, S. Brahnam, S. Ghidoni, An ensemble of visual features for gaussians of local descriptors and non-binary coding for texture descriptors, Expert Syst. Appl. 82 (2017) 27–39.

[86] M. Radovic, M. Ghalwash, N. Filipovic, Z. Obradovic, Minimum redundancy maximum relevance feature selection approach for temporal gene expression data, BMC Bioinf. 18 (2017), 9 9.

[87] L. Celona, L. Manoni, Neonatal facial pain assessment combining hand-crafted and deep features, in: B. Sebastiano, G.M. Farinella, L. Marco, G. Gallo (Eds.), New Trends in Image Analysis and Processing – ICIAP 2017, Springer, Cham, 2017.

[88] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep, face recognition, in, British Machine Vision Conference (2015).

[89] G. Levi, T. Hassner, Emotion recognition in the wild via convolutional neuralnetworks and mapped binary patterns, in: Proceedings of ACM International Conference on Multimodal Interaction (ICMI), 2015.

[90] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, CVPR 1 (2017) 3.

[91] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Adv Neural Inf Process Syst, Curran Associates Inc, Red Hook, NY, 2012, pp. 1097–1105.

[92] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, 2016, pp. 770–778.

[93] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, University of Massachusetts, Amherst, 2007.

[94] S. Grossman, V. Shielder, K. Swedeen, J. Mucenski, Correlation of patient and caregiver ratings of cancer pain, J. Pain Symptom Manage. 6 (1991) 53–57.

[95] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve, Radiology 143 (1982) 29–36.

Corresponding author
Sheryl Brahnam can be contacted at: sbrahnam@missouristate.edu