

Machine assessment of neonatal facial expressions of acute pain

Sheryl Brahnam^{a,*}, Chao-Fa Chuang^{b,1}, Randall S. Sexton^{a,2}, Frank Y. Shih^{b,1}

^a Computer Information Systems, Missouri State University, 901 South National, Springfield, Missouri 65804, USA

^b Computer Vision Laboratory, College of Computing Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA

Available online 20 March 2006

Abstract

We propose that a machine assessment system of neonatal expressions of pain be developed to assist clinicians in diagnosing pain. The facial expressions of 26 neonates (age 18–72 h) were photographed experiencing the acute pain of a heel lance and three nonpain stressors. Four algorithms were evaluated on out-of-sample observations: PCA, LDA, SVMs and NNSOA. NNSOA provided the best classification rate of pain versus nonpain (90.20%), followed by SVM with linear kernel (82.35%). We believe these results indicate a high potential for developing a decision support system for diagnosing neonatal pain from images of neonatal facial displays.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Neonate pain recognition; Medical face classification; Support vector machines; Linear discriminant analysis; Principal component analysis; Neural network simultaneous optimization algorithm

1. Introduction

Prior to the mid-1980s, neonates rarely received anaesthesia when undergoing surgeries and other potentially pain inducing procedures, as most anaesthesiologists at that time assumed newborns did not have the cortical development to experience pain [53]. In the last two decades, mounting evidence to the contrary has dislodged this assumption (see Ref. [63] for a recent review of the literature). There has been a dramatic shift in attitude so that today the health care system is deeply committed to developing better infant

pain management protocols and pain assessment instruments [32].

There are many reasons pain needs to be diagnosed in newborns. Aside from the fact that pain is a major indicator of medical conditions [31] and that the quality of patient care depends on the quality of pain management [7,19], recent research indicates that untreated pain in infants may result in central nervous system changes that slow development [55]. There are also studies that suggest frequent pain has a negative impact on parent–child bonding [28] and produces other significant long-term effects [2,14].

Clinically, pain is defined as a subjective experience and the most reliable method for assessing pain is through self report [34]. Most adults are capable of verbally describing the location, duration and intensity of their pain experiences. Nonverbal self reporting methods, such as the Wong-Baker FACES Pain Rating Scale [65] and the Faces Pain Scale (FPS) [4], have been devised that allow young children and others with

* Corresponding author. Tel.: +1 417 836 4932; fax: +1 417 836 6907.

E-mail addresses: sbrahnam@missouristate.edu (S. Brahnam), xccl235@njit.edu (C.-F. Chuang), rss000f@missouristate.edu (R.S. Sexton), shih@njit.edu (F.Y. Shih).

¹ Tel.: +1 973 596 5654; fax: +1 973 596 5777.

² Tel.: +1 417 836 6453; fax: +1 417 836 6907.

limited language skills to indicate the intensity of their pain experiences by pointing to one of several abstractly drawn faces expressive of increasing discomfort. Patients, however, that are incapable of communicating pain either verbally or pictorially must rely entirely on proxy judgments.

In evaluating neonatal pain, health professionals currently draw on both physiological and behavioral information. Physiological changes associated with pain include pupil dilation, a change in skin color, an increase in heart rate, respiratory rate and blood pressure, and a decrease in vagal tone and palmar sweating [9,52]. Experts caution health professionals against relying solely on physiological measures when assessing pain because the physiological parameters associated with pain are often indistinguishable from responses to many nonnoxious stimuli [52], especially those that provoke fear and anxiety [60].

Neonatal behavioral responses to pain include gross body movement, crying, changes in sleeping and eating patterns, and facial expressions [31,52]. Facial expressions, in particular, are considered the gold standard in pain assessment [11] because they are the most specific and frequent indicators of pain [15,16,51]. As such, most pain instruments developed for infants, toddlers and older children, including COMFORT [1], CRIES [25], FLACC (Face, Legs, Activity, Cry, Consolability) [33], MIPS (Modified Infant Pain Scale) [6], CFACS (Child Facial Coding System) [13] and NFCS (Neonatal Facial Coding System) [39] rely in whole or in part on facial displays. The facial characteristics associated with pain in infants include prominent forehead, eye squeeze, naso-labial furrow, taut tongue and an angular opening of the mouth [15]. Unlike facial behaviors, body movement and crying are less specific indicators of pain, as they are associated with other states, such as hunger, fright and discomfort [26]. Furthermore, neonates do not always respond to pain by crying and moving [24,43].

Even though the facial characteristics of infant expressions of pain have been studied extensively (see, for instance, [15]), there are serious problems with pain assessment instruments that use facial displays. The primary problem is that these tools depend on the knowledge, reliability and objectivity of those making the assessments, and research shows that health professionals are oftentimes biased, tend to underrate pain intensity [42] and fail to take into consideration a significant portion of the information available in infant facial displays [41]. Whether it is because health professionals become desensitized through constant exposure to suffering [66] or some other cause, there

is evidence that the greater the clinical experience of the health professional the more likely he or she is to underestimate patient pain [42,66]. A repeated refrain in the reference literature is that new assessment instruments need to be developed that take into account observational bias, desensitization and inattention to relevant facial details [21,50,66].

The objective of this study is to bypass these observational problems by developing a machine classification system to diagnose neonatal facial expressions of pain. Since assessment of pain by machine is based on pixel states, a machine classification system of pain will remain objective and will exploit the full spectrum of information available in a neonate's facial expressions. Furthermore, it will be capable of monitoring a neonate's facial expressions when the patient is left unattended and will not degrade over time.

Applying face recognition techniques to medical problems has only recently been attempted. Gunaratne and Sato [17] used a mesh-based approach to estimate asymmetries in facial actions to determine the presence of facial motion dysfunction for patients with Bell's palsy. Dai et al. [12] proposed a method for observing the facial expressions of patients in hospital beds, but the facial images used in their study were not of actual patients but rather of subjects responding to verbal cues suggestive of medical procedures and conditions.

The only other study that has explored machine classification of faces for medical purposes is [5], which this study extends. In this initial study, the facial expressions of 26 neonates between the ages of 18 h and 3 days old were photographed experiencing the pain of a heel lance and a variety of stressors, including transport from one crib to another, an air stimulus on the nose and friction on the external lateral surface of the heel. Three face classification techniques, principal component analysis (PCA), linear discriminant analysis (LDA) and support vector machines (SVMs), were used to classify the faces. An SVM with a polynomial kernel of degree 3 produced the best overall recognition rates of pain versus nonpain (88.00%), pain versus rest (94.62%), pain versus cry (80.00%), pain versus air puff (83.33%) and pain versus friction (93.00%).

As the intention of the initial study was to examine classification differences between pain expressions and a variety of similar nonpain expressions, the images used in the experiments were divided into training and testing sets based on facial expression categories, not subjects. As a result, the training and testing sets contained multiple samples of each subject in each expression category. While it is true that ideally, as is the case with speech recognition software, samples of

individual subjects would be used to personalize the classifier, in a clinical setting, it is more realistic to assume that the classifier would be trained previously on one set of subjects and then applied out of the box to future newborns.

This study extends [5] by examining PCA, LDA and SVM prediction accuracy using the more realistic evaluation protocol of requiring that subjects in the testing sets not be included in the training sets. Furthermore, only two categories of facial expressions are examined: pain and nonpain, a choice that is discussed further in Section 5. Along with PCA, LDA and SVMs, this study also compares the performance of the neural network simultaneous optimization algorithm (NNSOA). NNSOA uses a modified genetic algorithm to search simultaneously for a parsimonious network and for a global solution in a supervised multilayer feedforward neural network [47]. This algorithm has been shown to be successful in finding NN solutions that generalized well for real-world examples [45–47]. While PCA, LDA and SVMs have a proven track record in face classification [8,40,67], this is the first time NNSOA has been applied to a face recognition problem.

The basic concepts of PCA, LDA, SVMs and NNSOA are outlined in Section 2, and the study design is described in Section 3. The procedures used in the classification experiments are presented in Section 4. The evaluation protocol is discussed further in Section 5 and the experimental results are reported in Section 6. We conclude the paper, in Section 7, by noting some of the contributions and limitations of this study and by offering directions for future research.

2. Basic concepts of PCA, LDA, SVMs and NNSOA

The basic concepts behind PCA, LDA, SVMs and NNSOA are presented in this section. PCA and LDA have successfully been used to classify faces according to identity [54,56,57], gender [23,36,59], age [58], race [27,37] and facial expression [10,29,38]. Both PCA and LDA are linear classifiers. Linear classifiers represent images as a sum of linear combinations (sum of coefficients \times base). PCA seeks a projection that best represents the data and LDA seeks a projection that best separates the data. Because PCA and LDA are simple and well understood, they are typically used as algorithmic benchmarks.

Only recently have SVMs been applied to face recognition problems [20,35,40]. They typically outperform PCA and LDA [18,20,35]. Of particular interest to this study are the experiments of [35], where SVMs

outperformed human test subjects given the same face classification task. Although this is the first study to apply NNSOA to a face recognition task, NNSOA has successfully been applied to a number of medical problems, including diagnosing breast lumps, diabetes among Pima Indians and heart disease prediction [44].

2.1. PCA

The central idea behind PCA is to find an orthonormal set of axes pointing in the direction of maximum covariance in the data. In terms of facial images, the idea is to find the orthonormal basis vectors, or the eigenvectors, of the covariance matrix of a set of images, with each image treated as a single point in a high dimensional space. Since each image contributes to each of the eigenvectors, the eigenvectors resemble ghostlike faces when displayed. For this reason, they are oftentimes referred to in the literature as *eigenfaces* [56] and the new coordinate system is referred to as the *face space* [56].

Individual images can be projected onto the face space and represented exactly as weighted combinations of the eigenface components. The resulting vector of weights that describe each face is used both in face classification and in data compression.

Compression is achieved by reconstructing images using only those few eigenfaces that account for the most variability [49]. Since the eigenfaces are ordered, with each one accounting for a different amount of variation among the faces, images can be reconstructed using only the first few eigenfaces [49]. Because PCA results in a dramatic reduction of dimensionality and maps the most significant variations in a dataset, it is typically used to represent faces when performing other classification procedures.

Classification is performed by projecting a new image onto the face space and comparing the resulting weight vector to the weight vectors of a given class (see [56] for details). An outline of PCA face classification is provided below. Although technically PCA is not trained, the term *training* is used when describing PCA face recognition classifiers [56].

2.2. Outline of PCA face classification

| PCA training | PCA testing |
|---|--|
| Using a set of training feature vectors: | Using a set of testing feature vectors: |
| 1. Compute the average feature vector, Ψ . | 1. Subtract Ψ from the feature vectors to obtain Γ . |
| 2. Subtract Ψ from the feature | 2. Obtain the weight vectors, or |

- | | |
|---|---|
| <p>vectors to obtain Γ, the mean adjusted dataset.</p> <ol style="list-style-type: none"> 3. Derive eigenfaces using Γ. 4. Obtain the weight vectors, or eigenvalues, for each Γ_k by projecting it onto the resulting face space. 5. Reduce dimensionality by retaining only the most significant eigenvalues. 6. Obtain the class vectors, Ω, by averaging the eigenvalues of each Γ_k belonging to each classes. | <p>the eigenvalues, for each Γ_k by projecting Γ_k onto the face space derived using the training set.</p> <ol style="list-style-type: none"> 3. Reduce dimensionality as was done with the training set. 4. Classify each Γ_k based on its distance from Ω, using a distance metric. |
|---|---|
-

2.3. LDA

While PCA is optimal for reconstructing images from a low dimensional space, it is not optimal for discrimination. PCA yields projection directions that maximize the total scatter across all classes. LDA, in contrast, is a supervised learning procedure that projects the images onto a subspace that maximizes the between-class scatter and minimizes the within-class scatter of the projected data. A classical technique in pattern recognition, LDA, is an example of a *class specific method* in that it shapes the scatter in order to make it more reliable for classification [3]. There has been a tendency to prefer LDA to PCA because LDA deals directly with discrimination between classes, whereas PCA aims at faithfully representing the data.

2.4. SVMs

SVMs, introduced in [62], are learning systems that separate sets of input pattern vectors into two classes with an optimal separating hyperplane. The set of vectors is said to be optimally separated by the hyperplane if it is separated without error and the distance between the closest vectors to the hyperplane is maximal. SVMs produce the pattern classifier (1) by applying a variety of kernel functions (linear, polynomial, radial basis function (RBF) and so on) as the possible sets of approximating functions, (2) by optimizing the dual quadratic programming problem and (3) by using structural risk minimization as the inductive principle, as opposed to classical statistical algorithms that maximize the absolute value of an error or of an error squared.

Different types of SVM classifiers are used depending upon the type of input patterns: a linear maximal margin classifier is used for linearly

separable data, a linear soft margin classifier is used for linearly nonseparable, or overlapping, classes, and a nonlinear classifier is used for classes that are overlapped as well as separated by nonlinear hyperplanes. All three classifiers are outlined below. It should be noted, however, that the linearly separable case is rare in real world problems and was not explored in the experiments performed in this study.

2.5. Outline of SVM

Suppose, there is a set of training data, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ where $\mathbf{x}_i \in \mathbf{R}^n$ and $i = 1, 2, \dots, k$. Each \mathbf{x}_i , belonging as it does to one of two classes, has a corresponding value y_i , where $y_i \in \{-1, 1\}$.

2.5.1. Linear maximal margin classifier

The goal is to build the hyperplane that maximizes the minimum distance between the two classes. This hyperplane is called the optimal separating hyperplane (OSH). OSH has the form:

$$f(\mathbf{x}) = \sum_{i=1}^k \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b, \quad (1)$$

where α_i and b are the solutions of a quadratic programming problem.

The unseen test data \mathbf{x}_t can be classified by simply computing (2).

$$f(\mathbf{x}) = \text{sign}(w_0 \cdot \mathbf{x}_t + b_0) \quad (2)$$

Examining (2), it can be seen that the hyperplane is determined by all the training data, \mathbf{x}_i , that have the corresponding attributes of $\alpha_i > 0$. We call this kind of training data *support vectors*. Thus, the optimal separating hyperplane is not determined by the training data per se but rather by the support vectors.

2.5.2. Linear soft margin classifier

The objective in this case is to separate the two classes of training data with a minimal number of errors. To accomplish this, some non-negative slack variables, ξ_i , $i = 1, 2, \dots, k$, are introduced to the system. The penalty, or regularization parameter, C , is also introduced to control the cost of errors. The computation of the linear soft margin classifier is the same as the linear maximal margin classifier. Thus, we can obtain OSH using Eqs. (1) and (2).

2.5.3. Nonlinear classifier

In this case, kernel functions, such as the polynomial or RBF, are used to transform the input space to a feature space of higher dimensionality. In the feature space, a linear separating hyperplane is sought that separates the input vectors into two classes. In this case, the hyperplane and decision rule for the nonlinear training pattern is Eq. (3):

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^K \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (3)$$

where α_i and b are the solutions of a quadratic programming problem and $K(\mathbf{x}_i, \mathbf{x})$ is a kernel function.

2.6. NNSOA

NNSOA is a global search procedure that searches from one population of NN solutions to another, focusing on the area that provides the current best solution, while continuously sampling the total parameter space. NNSOA is a slight modification of a genetic algorithm used in previous NN studies (see [45]). NNSOA takes advantage of the GA's ability to simultaneously search multiple points (or solutions) at one time, unlike gradient search techniques, such as backpropagation, that are able to search for only one solution at a time. As explained below, what makes NNSOA unique is the addition of a penalty in the objective function. This penalty enables NNSOA to eliminate unneeded weights in the NN architecture. Thus, NNSOA is able to produce solutions that generalize better than solutions found using gradient search techniques.

Because NNSOA uses a genetic algorithm for the search procedure, it is not limited to differentiable functions, as is the case with gradient search techniques. Thus, NNSOA can have objective functions that add a penalty for the number of nonzero weights in a solution. NNSOA is able to eliminate unneeded weights in a solution by intermittently exchanging solution weights with hard zeros and then evaluating if that substitution helped or hindered the network's ability to predict using normal GA operations.

Backpropagation does not have the ability to zero out weights in a solution. Therefore, in a backpropagation network, or in any gradient search technique used for searching for optimal weights in a NN solution, the search must find a solution that has values for these unneeded weights that will in effect zero each other out when producing estimates. This works well for training data but is likely to introduced additional errors in the

estimates when these solutions are applied to out-of-sample data. A solution found by the NNSOA completely eliminates this possibility of additional error because the unneeded weights are set to a hard zero.

Not only does the NNSOA find better solutions by eliminating unneeded weights, but it also finds appropriate NN architectures by searching for the correct number of hidden nodes in a solution. This is done by starting a network with only one hidden node. After a user specified number of generations (MAXHID), the best solution out of the population of solutions is saved, and an additional hidden node is added to the architecture and trained for another MAXHID generation. The previous best solution is included in this additional training by replacing one of the randomly initialized solutions with the best solution found in the previous architecture. Since adding an additional node to the architecture increases the number of weights in the solutions equal to the number of inputs plus one, the additional weights for this best solution are set to hard zeros. The process of adding an additional hidden node after every MAXHID generation continues until the current best solution is worse than the previous architecture's best solution. At this point, the number of hidden nodes is set to the number of hidden nodes in the previous architecture, and the NN continues with the training process for a user defined number of generations. Once the MAXGEN number of generations has been reached, training is complete.

The operations used in NNSOA are outlined below.

2.7. Outline of NNSOA

2.7.1. Initialization

A population of 12 solutions is created by drawing random real values from a uniform distribution $[-1, 1]$ for input weights. The output weights are determined by ordinary least squares (OLS).

2.7.2. Evaluation

Each member of the current population is evaluated by an objective function based on its sum-of-squared error (SSE) value in order to assign each solution a probability for being redrawn in the next generation. To search for a parsimonious solution, a penalty value is added to the SSE for each nonzero weight (or active connection). The penalty for keeping an additional weight varies during the search and is equal to the current value of the root mean squared error (RMSE). This means that the penalty for keeping additional weights is high at the beginning of the training process

when errors are high. As the optimization process gets closer to the final solution, errors decrease and the penalty value becomes smaller. Based on the objective function, each of the 12 solutions in the population is evaluated. The probability of being drawn in the next generation is calculated by dividing the distance of the current solution's objective value from the worst objective value in the generation by the sum of all distances in the current generation.

2.7.3. Reproduction

A mating pool of 12 solutions is created by selecting solutions from the current population based on their assigned probability. This is done by selecting a random number in the range of 0 and 1 and comparing it to the cumulative probability of the current solution. When it is found that the random value is less than the current solution's cumulative probability, the current string is drawn for the next generation. This is repeated until the entire new generation is drawn. It should be noted that a given solution can be drawn more than once or not at all, depending on its assigned probability.

2.7.4. Crossover

Once reproduction occurs, providing a combination of solutions from the previous generation, the 12 solutions are then randomly paired so that six pairs are produced. A point is randomly selected for each pair. New solutions are produced by switching the weights above the randomly generated point. In this fashion, 12 new solutions are generated for the next generation.

2.7.5. Mutation

For each weight in a population of solutions, a random number is drawn; if the random value is less than 0.05, the weight is replaced by a value randomly drawn from the entire weight space. By doing this, the entire weight space is globally searched, thus enhancing the algorithm's ability to find global solutions.

2.7.6. Mutation 2

For each weight in the population of solutions, a random number is drawn; if the random value is less than 0.05, the weight is replaced by a hard zero. As a result of doing this, unneeded weights are identified as the search continues for the optimum solution. After this operator is performed, this new generation of 12 solutions begins again with evaluation and the cycle continues until it reaches 70% of the maximum set of generations.

2.7.7. Convergence enhancement

Once 70% of the maximum set of generations has been completed, the best solution replaces all the strings in the current generation. The weights of these 12 identical solutions are then modified by adding a small random value to each weight. These random values decrease to an arbitrarily small number as the number of generations increase to its set maximum number.

2.7.8. Termination

The algorithm terminates on a user specified number of generations.

3. Study design

One of the most important considerations in the study design was the choice of stimuli used to provoke facial displays in neonates. The objective was to obtain a representative set of images for evaluating machine classification of neonatal facial displays of pain.

Most early research in neonatal pain assessment examined differences in neonatal facial responses to only two noxious stimuli: a pain inducing stimulus (pin prick or puncture of a lancet) and friction on the external lateral surface of the heel [16,22]. As can be seen in Fig. 1, friction can produce expressions of distress that are similar to expressions of pain. Contemporary research tends to include more stressors, such as exposure to bright light [66] and diaper change [64]. These stressors are designed to provoke facial expressions that have features in common with the facial displays of pain. Exposure to bright light, for instance, often results in eye squeeze, a facial characteristic of pain.

This study follows contemporary research by including four noxious 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. Since classifiers easily discriminate changes in lighting, we introduced an air stimulus on the nose to provoke eye squeeze. The third stressor was introduced as a result of a logistic consideration. In order to get a clear photograph of the neonate's face, the infant had to be transported from the regulation hospital crib, which is surrounded by a fixed Plexiglas shield, to a crib with removable sides. We discovered that transporting the neonates supplied a stressor that often triggered a crying expression that was not in response to pain.

3.1. Subjects

This study complied with the protocols and ethical directives for research involving human subjects at St.

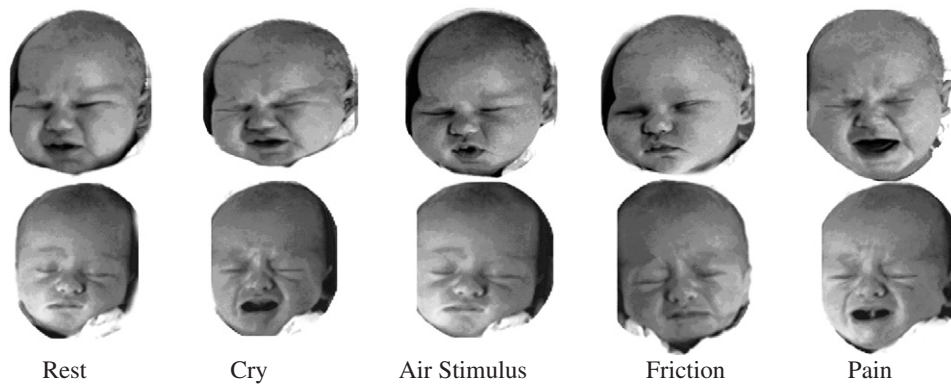


Fig. 1. Examples of the five facial expressions in the dataset.

John's Health System, Inc. Informed consent was obtained from a parent, usually the mother in consultation with the father. Most parents were recruited in the neonatal unit of a St. John's Hospital sometime after delivery. Only mothers who had experienced uncomplicated deliveries were approached.

A total of 204 color photographs were taken of 26 Caucasian neonates (13 boys and 13 girls) ranging in age from 18h to 3days old. Six males had been circumcised the day before the photographs were taken, and the last feeding time before the photography session ranged from 45min to 5h. All infants were in good health.

3.2. Apparatus

All photographs were taken using a Nikon D100 digital camera under ambient lighting conditions in a room separated from other newborns.

3.3. Procedure

The facial expressions of the newborns were photographed in one session. All stimuli were administered by an attending nurse. Following the requirements of standard medical procedures, photographs of the four stimuli were taken in the following sequence:

1. Transport from one crib to another (rest/cry): after being transported from one crib to another, the neonate was swaddled and a series of photographs was taken over the course of 1 min. The state of the neonate was noted as either crying or resting for each photograph taken in the series.
2. Air stimulus: after resting for at least one additional minute, the neonate's nose was exposed to a puff of air emitted from a squeezable plastic camera lens

cleaner. A series of pictures of the neonate's face was taken immediately after the air puff contacted the infant's face.

3. Friction: after resting for at least 1 min, the neonate received friction on the external lateral surface of the heel with cotton wool soaked in 70% alcohol for 10 to 15 s. The face of the neonate was repeatedly photographed during the friction rubbings.
4. Pain: after resting for at least 1 min, the external lateral surface of the heel was punctured for blood collection. Several continuous photographs of the neonate's face were taken, starting immediately after introduction of the lancet and while the skin of the heel was squeezed for blood samples.

Note: The 1min rest period between stimuli exposures follows the example of [66].

Of the 204 facial photographs, 67 are rest, 18 are cry, 23 are air stimulus, 36 are friction and 60 are pain. Fig. 1 provides two example sets, with backgrounds removed, of the five neonatal facial expressions of rest, cry, air puff, friction and pain.

4. Experimental procedures

As illustrated in Fig. 2, the experimental procedures used in this study can be divided into the following stages: preprocessing, feature extraction and classification.

In the preprocessing stage, the original images were cropped, rotated, and scaled. Eyes were aligned roughly along the same axis. The original 204 images, size 3008×2000 pixels, were also reduced to 100×120 pixels.

In the feature extraction stage, facial features were centered within an ellipse and color information was discarded. The rows within the ellipse were

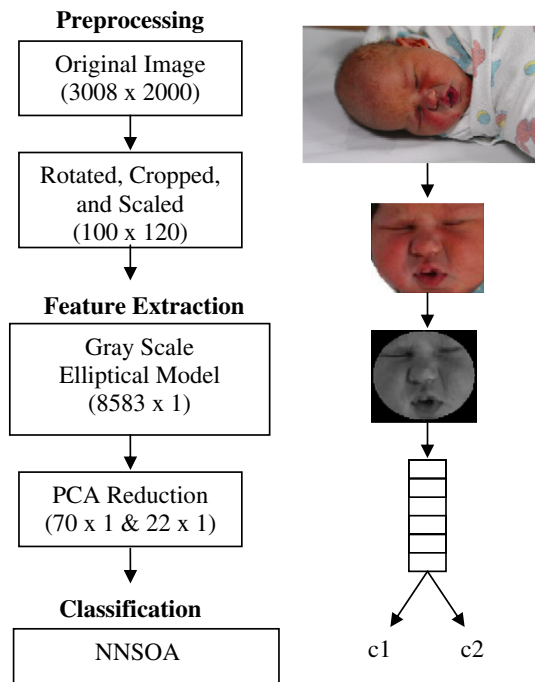


Fig. 2. The experimental procedure.

concatenated to form a feature vector of dimension 8583 with entries ranging in value between 0 and 255. PCA was then used to reduce the dimensionality of the feature vectors further (see Section 2.1 for an explanation). The first 70 principle components resulted in the best classification scores, except in the case of NNSOA, which needed on average only the first 22 principle components to perform optimally.

Finally, in the classification stage, PCA, LDA, SVMs and NNOSA were used to classify the feature vectors following the evaluation protocol described in the next section.

5. Evaluation protocol

The evaluation protocol used in this study classified images into two categories: pain and nonpain. The set of nonpain images was obtained by combining the rest, cry, air puff and friction images into one category of 144 images. The set of pain images consisted of the remaining 60 images. In addition, a total of 26 experiments were performed by each classifier. For each of the 26 subjects, the set of facial images for that subject formed the testing set and the facial images of the remaining 25 subjects formed the training set. The classification scores for each experiment were

computed by averaging the number of correct classifications made.

This study did not attempt to evaluate pain intensity because of the problems involved in measuring pain intensity levels in neonates. As mentioned in the introduction, the major issue with neonatal pain assessment is the absence of self-report and the resulting difficulty in identifying the presence of pain. Our binary approach to neonatal assessment does not differ from that taken in developing a number of pain assessment instruments that focus on neonates and young infants (see, for instance, [48,61]).

6. Experimental results

This section describes experimental results using PCA, LDA, SVMs and NNSOA. Tables 1 and 2 present the classification scores and 95% confidence interval for each of the classifiers. Examining these tables, we see that NNSOA has the highest classification rate of 90.20% accuracy, with a 95% confidence interval of $\pm 4.16\%$.

Comparing NNSOA to the other classification methods, we can see from Table 2 that a statistical difference exists in the performance of NNSOA as compared to LDA and SVMs with RBF and polynomial degree 4 kernels. No statistical difference in performance exists between NNSOA and SVMs with linear and polynomial degrees of 2 and 3 kernels. However, examining Table 2 further, we see that NNSOA has the least standard deviation; therefore, it is the most stable method of classification explored in this study. A more detailed discussion of method parameters and the classification results of each of the classifiers are provided below.

In general, we believe that the results of these experiments indicate a high potential for developing a decision support system for diagnosing neonatal pain from images of neonatal facial displays. However, Table 1, which presents the classification results for the 26 experiments performed by each method, presents a perplexing problem. It is clear by examining the results in this table that certain subjects are easy to classify while others are more difficult. For instance, all four methods correctly classified the image sets associated with subjects 2, 21 and 25, whereas subjects 1, 9 and 26 proved difficult for most methods. Thus far we have not been able to determine the set of facial and image characteristics that obstructed and facilitated the machine classifications.

Table 1

Average method classification scores and individual experiment scores

| Exp. no. | Method | | | | | | | |
|----------|------------------------|------------------------------|------------------------------|------------------------------|---------------------|---------|---------------------------|---------------------------|
| | SVM with linear kernel | SVM with polynomial degree=2 | SVM with polynomial degree=3 | SVM with polynomial degree=4 | SVM with RBF kernel | NNSOA | PCA L ₁ metric | LDA L ₁ metric |
| 1 | 77.78% | 44.44% | 77.78% | 55.56% | 55.56% | 77.78% | 88.89% | 66.67% |
| 2 | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100% | 100.00% |
| 3 | 87.50% | 75.00% | 75.00% | 75.00% | 75.00% | 100.00% | 62.50% | 75.00% |
| 4 | 60.00% | 100.00% | 100.00% | 100.00% | 60.00% | 80.00% | 60.00% | 60.00% |
| 5 | 75.00% | 83.33% | 75.00% | 75.00% | 75.00% | 75.00% | 66.67% | 75.00% |
| 6 | 85.71% | 57.14% | 85.71% | 57.14% | 85.71% | 85.71% | 85.71% | 85.71% |
| 7 | 88.89% | 88.89% | 77.78% | 77.78% | 77.78% | 77.78% | 77.78% | 77.78% |
| 8 | 88.89% | 66.67% | 77.78% | 66.67% | 66.67% | 88.89% | 77.78% | 77.78% |
| 9 | 50.00% | 66.67% | 33.33% | 66.67% | 83.33% | 66.67% | 16.67% | 33.33% |
| 10 | 70.00% | 80.00% | 80.00% | 70.00% | 90.00% | 80.00% | 60.00% | 80.00% |
| 11 | 100.00% | 100.00% | 100.00% | 87.50% | 50.00% | 100.00% | 100.00% | 100.00% |
| 12 | 75.00% | 87.50% | 87.50% | 87.50% | 87.50% | 100% | 75.00% | 62.50% |
| 13 | 60.00% | 60.00% | 90.00% | 60.00% | 60.00% | 90.00% | 80.00% | 70.00% |
| 14 | 90.91% | 100.00% | 81.82% | 72.73% | 72.73% | 100% | 81.82% | 81.82% |
| 15 | 83.33% | 83.33% | 83.33% | 83.33% | 83.33% | 83.33% | 66.67% | 66.67% |
| 16 | 83.33% | 83.33% | 83.33% | 83.33% | 66.67% | 91.67% | 91.67% | 91.67% |
| 17 | 100.00% | 88.89% | 77.78% | 66.67% | 55.56% | 100.00% | 100.00% | 100.00% |
| 18 | 90.00% | 80.00% | 80.00% | 70.00% | 60.00% | 100.00% | 100.00% | 100.00% |
| 19 | 85.71% | 85.71% | 85.71% | 85.71% | 85.71% | 85.71% | 100% | 100.00% |
| 20 | 75.00% | 83.33% | 66.67% | 50.00% | 58.33% | 92.31% | 91.67% | 58.33% |
| 21 | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| 22 | 87.50% | 75.00% | 75.00% | 62.50% | 62.50% | 100.00% | 62.50% | 62.50% |
| 23 | 50.00% | 83.33% | 100.00% | 100.00% | 66.67% | 80.00% | 66.67% | 33.33% |
| 24 | 100.00% | 85.71% | 85.71% | 42.86% | 57.14% | 100.00% | 100.00% | 85.71% |
| 25 | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| 26 | 100.00% | 50.00% | 50.00% | 50.00% | 50.00% | 100.00% | 66.67% | 66.67% |
| Average | 82.35% | 79.90% | 80.39% | 72.06% | 70.10% | 90.20% | 80.39% | 76.96% |

6.1. PCA and LDA

PCA and LDA used the sum of absolute differences, or L₁, distance metric in all experiments. Referring to Table 1, the average classification score for PCA was 80.39% and for LDA 76.96%. However, we see from Table 2 that there is no statistical difference in performance between PCA and LDA. This is not unexpected, as it has been shown that in general LDA

outperforms PCA only when large and representative training data sets are given [30].

All PCA and LDA experiments were processed in the MATLAB environment under the Windows XP operating system using a Pentium 4—2.80 GHz processor.

6.2. SVM Results

SVMs with five different kernels (linear, RBF, polynomial degree 2, polynomial degree 3 and polynomial degree 4) performed the 26 experiments defined by the evaluation protocol. The regularization parameter, C , for the SVMs was determined using a grid search. Since the recognition rates in our experiments were not significantly different in terms of different values for C , we adopted the regularization parameter $C=1$. The bandwidth parameter, $\gamma=1.2$, in SVM using RBF kernels was also optimized using a grid search.

Referring to Table 1, an SVM with linear kernel provided the best recognition rate of 82.35%. However, examining Table 2, there is no statistical significant

Table 2

Standard deviations of experiment scores and 95% confidence intervals using t distribution ($\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$)

| Method | 95% confidence interval | Standard deviation |
|----------------------------------|-------------------------|--------------------|
| NNSOA | 90.20%±4.16% | 10.30% |
| SVM with linear kernel | 82.35%±6.20% | 15.34% |
| SVM with polynomial degree=2 | 79.90%±6.36% | 15.74% |
| SVM with polynomial degree=3 | 80.39%±6.23% | 15.41% |
| SVM with polynomial degree=4 | 72.06%±7.03% | 17.41% |
| SVM with RBF kernel | 70.10%±6.32% | 15.64% |
| PCA with L ₁ distance | 80.39%±7.98% | 19.75% |
| LDA with L ₁ distance | 76.96%±7.81% | 19.34% |

between the classification rates of the various SVM methods.

All SVM experiments were processed in the same MATLAB environment used for the PCA and LDA experiments. SVM was implemented using the OSU SVM Classifier MATLAB Toolbox developed by Ohio State University.

6.3. NNSOA results

Following the experimental protocol described in Section 5, a total of 26 separate NNs were trained and tested. MAXHID and MAXGEN were set to 100 and 5000, respectively. The objective function used in the experiments is shown in Eq. (4).

$$\text{Min} \left\{ E = \sum (O_i - \hat{O}_i)^2 + C \sqrt{\frac{\sum_{i=1}^N (O_i - \hat{O}_i)^2}{N}} \right\} \quad (4)$$

where N is the number of observations in the data set, O is the observed value of the dependent variable, \hat{O} is the NN estimate and C is the number of nonzero weights in the network.

For these experiments, the output weights were found by using ordinary least squares, i.e., by regressing the outputs from the hidden nodes onto the real outputs. In this way, only values for the input weights were searched.

The average number of hidden nodes from the 26 networks was 4.11. Since there were 70 inputs+1 bias, each additional hidden node added 71 inputs to the solution, making the average number of weights 291.81 (71×4.11). However, because NNSOA eliminates weights in a solution by zeroing out the weights that are not useful for prediction (see Section 2.4 for details), the average number of weights across the 26 NNs was much smaller. On average, only 23.58 weights were found to be nonzero. The reduction of architecture to approximately 24 weights per network helped NNSOA generalize to out-of-sample data.

An added advantage of eliminating unneeded weights is the identification of relevant variables. The input variables that had all zero weight connections were not used in producing estimates. As a result, the average number of inputs that were actually used in NNSOA prediction across the 26 networks was reduced from the 70 to an average of only 21.88. The parsimonious architecture of the NNSOA networks and reduction in input space may account in

part for the superior performance of NNSOA as compared with the results reported above for PCA, LDA and SVM.

All NNSOA experiments were conducted on a 1.5GHz machine, using the Windows XP operating system. The core code of the NNSOA program was written in FORTRAN, with Visual Basic used for the interface.

7. Conclusion

Pain is a major indicator of medical problems and a major source of discomfort. Although the vast majority of patients are capable of describing their pain experiences, other patients, such as neonates, must rely entirely on proxy judgments. At present, the medical assessment of neonatal pain takes into consideration a number of physiological and behavioral factors, with neonatal facial expressions playing a central role in pain assessment. Studies demonstrate, however, that clinicians are not entirely impartial in their judgments, vary according to their level of clinical experience and oftentimes fail to exploit the full spectrum of information available in an infant's facial expressions.

In this paper, we propose that a machine assessment system of neonatal expressions of pain be developed to assist health professionals in diagnosing pain. We report a study designed to investigate machine classification of neonatal facial displays. The facial expressions of 26 neonates between the ages of 18h and 3 days old were photographed experiencing the pain of a heel lance and three nonpain inducing stressors: transport from one crib to another, an air stimulus on the nose and friction on the surface of the heel. The neonatal facial images were divided into two categories: (1) pain, which included all the heel lance images, and (2) nonpain, which combined the images of the nonpain inducing stressors. Two advanced classification algorithms, SVM and NNSOA, and two baseline face recognition algorithms, PCA and LDA, performed 26 classification experiments, one for each subject. The facial images of 25 subjects formed the training set, and the images of the remaining subject formed the testing set. The classification rates of NNSOA and the SVMs were promising. NNSOA provided the best classification rate of 90.20% accuracy. SVM with a linear kernel provided the best SVM classification rate of 82.35% accuracy. PCA and LDA using an L_1 distance metric produced an average classification rate of 80.39% and 76.96%, respectively.

We believe that this study makes a number of contributions aside from proposing a machine

assessment system of neonatal expressions of pain. First, we are one of the very first to apply face recognition technology to a medical task. Second, even though machine classification of emotion has long been an area of active investigation, we are unaware of research outside our own endeavors that includes facial expressions of actual pain states. Third, we are also one of the first to perform a machine face classification study that includes infant faces and we are the first to explore machine classification of neonatal faces.

There are a number of limitations in our study that also deserve comment. First, only reactions to acute pain experiences were included in the dataset. This study does not address chronic pain. Chronic pain is produced by diseased conditions and is more persistent. Although chronic pain is not as commonly experienced by children as adults [31], it is thought to have more serious long-term psychological and neurological consequences for children [2,31] and needs to be addressed in future studies. Second, this study uses two-dimensional still photographs and does not consider the dynamic and multidimensional nature of facial expressions. Third, this study does not speculate on the practicality of implementing a face recognition decision support system within a hospital setting.

In terms of future research possibilities, we are working on another study that will include video images of approximately 500 neonates. Video will allow us to investigate the dynamic and multidimensional nature of facial displays. The dataset will include neonatal facial displays provoked by additional stimuli; for example, the temperature change of removing a heel warmer applied to increase blood flow and the repeated deep pain of the heel squeezes that follow heel lancing. Another promising research possibility would be to compare machine assessment of neonatal pain, using facial displays as input, to one or several of the pain assessment instruments currently in use.

To conclude, we believe the results of this study indicate a high potential for developing a decision support system for diagnosing neonatal pain from neonatal facial displays. Given the potential benefits such a system could provide paediatric clinicians in the care of neonates, this is a research area that warrants further investigation.

Acknowledgements

The authors gratefully acknowledge the partial funding of this project by Missouri State University faculty grant #1015-22-2181.

References

- [1] B. Ambuel, K.W. Hamlett, C.M. Marx, J.L. Blumer, Assessing distress in pediatric intensive care environments: the comfort scale, *Journal of Pediatric Psychology* 17 (1992) 95–109.
- [2] K.J.S. Anand, R.E. Grunau, T. Oberlander, Developmental character and long-term consequences of pain in infants and children, *Child and Adolescent Psychiatric Clinics of North America* 6 (1997) 703–724.
- [3] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (7) (1997) 711–720.
- [4] D. Bieri, R.A. Reeve, G.D. Champion, L. Addicoat, J.B. Ziegler, The face pain scale for the self assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties, *Pain* 41 (1990) 139–150.
- [5] S. Brahnam, C.-F. Chuang, F.Y. Shih, M.R. Slack, Machine recognition and representation of neonate facial displays of acute pain, *Journal of Artificial Intelligence in Medicine* (in press).
- [6] M. Buchholz, H.W. Karl, M. Pomietto, A. Lynn, Pain scores in infants: a modified infant pain scale versus visual analogue, *Journal of Pain and Symptom Management* 15 (2) (1998) 117–124.
- [7] C.R. Chapman, K.L. Syrjala, Measurement of pain, in: J.J. Bonica, J.D. Loeser, C.R. Chapman, W.F. Fordyce (Eds.), *The Management of Pain*, vol. 1, Lea and Febiger, Philadelphia, 1990, pp. 580–594.
- [8] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, *Proceedings of IEEE* 83 (1995) 705–740.
- [9] S. Coffman, Y. Alvarez, M. Pyngolil, R. Petit, C. Hall, M. Smyth, Nursing assessment and management of pain in critically ill children, *Heart and Lung* 26 (3) (1997) 221–228.
- [10] G.W. Cottrell, J. Metcalfe, Empath: face, emotion, and gender recognition using holons, in: D. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, Morgan and Kaufman, San Mateo, CA, 1991, pp. 564–571.
- [11] K.D. Craig, The facial display of pain in infants and children, in: G.A. Finley, P.J. McGrath (Eds.), *Measurement of Pain in Infants and Children*, Pain Research and Management, vol. 10 IASP Press, Seattle, 1998, pp. 103–121.
- [12] Y. Dai, Y. Shibata, T. Ishii, K. Hashimoto, K. Katamachi, K. Noguchi, N. Kakizaki, D. Cai, An associate memory model of facial expressions and its application in facial expression recognition of patients on bed, *Proceedings of IEEE International Conference on Multimedia and Expo*, 2001, pp. 772–775.
- [13] C.A. Gilbert, C.M. Lilley, K.D. Craig, P.J. McGrath, C.A. Court, S.M. Bennett, C.J. Montgomery, Postoperative pain expression in preschool children: validation of the child facial coding system, *Clinical Journal of Pain* 15 (3) (1999) 192–200.
- [14] 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.
- [15] R.E. Grunau, R.V.E. Grunau, K.D. Craig, Pain expression in neonates: facial action and cry, *Pain* 28 (3) (1987) 395–410.
- [16] R.V.E. Grunau, C.C. Johnston, K.D. Craig, Neonatal facial and cry responses to invasive and non-invasive procedure, *Pain* 42 (3) (1990) 295–305.

- [17] P. Gunaratne, Y. Sato, Estimation of asymmetry in facial actions for the analysis of motion dysfunction due to paralysis, *International Journal of Image and Graphics* 3 (4) (2003) 639–652.
- [18] G. Guo, S.Z. Li, K.L. Chan, Support vector machines for face recognition, *Image and Vision Computing* 19 (2001) 631–638.
- [19] D. Harrison, C. Evans, L. Johnston, P. Loughnan, Bedside assessment of heel lance pain in the hospitalized infant, *Journal of Obstetric, Gynecologic, and Neonatal Nursing* 31 (5) (2002) 551–557.
- [20] B. Heisele, P. Ho, T. Poggio, Face recognition with support vector machines: global versus component-based approach, *Proceedings of The Eighth IEEE International Conference on Computer Vision*, Vancouver, 2001, pp. 688–694.
- [21] M.S. Hultgren, Assessment of postoperative pain in critically ill infants, *Progress in Cardiovascular Nursing* 5 (3) (1990) 104–112.
- [22] C.E. Izard, R.R. Huebner, D. Risser, G.C. McGinnes, L.M. Dougherty, The young infant's ability to produce discrete emotion expressions, *Developmental Psychology* 16 (1980) 418–426.
- [23] A. Jain, J. Huang, Integrating independent components and linear discriminant analysis for gender classification, *Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.
- [24] C.C. Johnston, M.E. Strada, Acute pain response in infants: a multidimensional description, *Pain* 24 (1986) 373–382.
- [25] S.W. Krechel, J. Bilder, Cries: a new neonatal postoperative pain measurement score: initial testing of validity and reliability, *Paediatric Anaesthesia* 5 (1995) 53–61.
- [26] B.M. Lester, A biosocial model of infant crying, in: L. Lipsill (Ed.), *Advances in Infancy Research*, Ablex, New York, 1984, pp. 167–212.
- [27] X. Lu, A.K. Jain, Ethnicity identification from face images, *Proceedings of SPIE: Biometric Technology for Human Identification*, 2004, pp. 114–123.
- [28] R.E. Marshall, F.L. Porter, A.U. Rogers, J.A. Moore, B. Anderson, S.B. Boxerman, Circumcision 2: effects upon mother–infant interaction, *Early Human Development* 7 (1982) 367–374.
- [29] A.M. Martinez, R. Benavente, The AR Face Database, 1998.
- [30] A.M. Martinez, A.C. Kak, PCA versus LDAs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2) (2001) 228–233.
- [31] P.A. McGrath, *Pain in Children: Nature, Assessment, and Treatment*, Guildford Press, New York, 1989.
- [32] P.J. McGrath, A.M. Unruh, Neonatal pain in a social context, 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. 237–250.
- [33] S.I. Merkel, T. Voepel-Lewis, J.R. Shayevitz, S. Malviya, The Flacc: a behavioral scale for scoring postoperative pain in young children, *Pediatric Nursing* 23 (1997) 293–297.
- [34] H. Merskey, D.G. Albe-Fessard, J.J. Bonica, A. Carmon, R. Dubner, F.W.L. Kerr, U. Lindblom, J.M. Mumford, P.W. Nathan, W. Noordenbos, C.A. Pagni, R.A. Sternbach, S.S. Sunderland, Pain terms: a list with definitions and notes on usage recommended by the IASP subcommittee on taxonomy, *Pain* 6 (1979) 249–252.
- [35] B. Moghaddam, M.-H. Yang, Gender classification with support vector machines, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2000, pp. 306–311.
- [36] A.J. O'Toole, K.A. Deffenbacher, The perception of face gender: the role of stimulus structure in recognition and classification, *Memory and Cognition* 26 (1997) 146–160.
- [37] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, J.C. Bartlett, Classifying faces by race and sex using an autoassociative memory trained for recognition, *Proceedings of 13th Annual Conference on Cognitive Science*, Hillsdale, NJ, 1991, pp. 847–851.
- [38] C. Padgett, G.W. Cottrell, A simple neural network models categorical perception of facial expressions, *Proceedings of the 20th Annual Cognitive Science Conference*, Madison, WI, 1998.
- [39] J.W.B. Peters, H.M. Koot, R.E. Grunau, J. de Boer, V. Druenen, M.J.D. Tibboel, H.J. Duivenvoorden, Neonatal facial coding system for assessing postoperative pain in infants: item reduction is valid and feasible, *Clinical Journal of Pain* 19 (6) (2003) 353–363.
- [40] P.J. Phillips, Support vector machines applied to face recognition, *Advances in Neural Information Processing Systems* 11 (1998) 803–809.
- [41] K.M. Prkachin, S. Berzins, S.R. Mercer, Encoding and decoding of pain expressions: a judgement study, *Pain* 58 (2) (1994) 253–259.
- [42] 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 Research and Management* 6 (2) (2001) 105–112.
- [43] J.A. Rushforth, M.I. Levene, *Archives of Disease in Childhood* 70 (1994) F174–F176.
- [44] R.S. Sexton, R.E. Dorsey, Reliable classification using neural networks: a genetic algorithm and backpropagation comparison, *Decision Support Systems* 30 (2000) 11–22.
- [45] R.S. Sexton, R.E. Dorsey, J.D. Johnson, Toward a global optimum for neural networks: a comparison of the genetic algorithm and backpropagation, *Decision Support Systems* 22 (1998) 171–185.
- [46] R.S. Sexton, R.S. Sriram, H. Etheridge, Improving decision effectiveness of artificial neural networks—a modified genetic algorithm approach, *Decision Sciences* 34 (3) (2003) 421–442.
- [47] R. Sexton, R. Dorsey, N. Sikander, Simultaneous optimization of neural network function and architecture algorithm, *Decision Support Systems* 36 (2004) 283–296.
- [48] J.G. Shade, B.A. Joyce, J. Gerkensmeyer, J.F. Keck, Comparison of three preverbal scales for postoperative pain assessment in a diverse pediatric sample, *Journal of Pain and Symptom Management* 12 (1996) 348–359.
- [49] L. Sirovich, M. Kirby, Low dimensional procedure for the characterization of human faces, *Journal of the Optical Society of America* 4 (3) (1987) 519–524.
- [50] J.B. Stevens, K.J.S. Anand, An overview of neonatal pain, 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. 1–7.
- [51] B.J. Stevens, C. Johnston, P. Petryshen, A. Taddio, Premature infant pain profile: development and initial validation, *Clinical Journal of Pain* 12 (1) (1996) 13–22.
- [52] B. Stevens, C. Johnston, S. Gibbins, Pain assessment in 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. 101–134.
- [53] L.I. Swafford, D. Allen, Pain relief in the pediatric patient, *Medical Clinics of North America* 52 (1968) 131–136.

- [54] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (8) (1996) 831–837.
- [55] A. Taddio, J. Katz, A.L. Illersich, G. Koren, Effect of neonatal circumcision on pain response during subsequent routine vaccination, *The Lancet* 349 (1997) 599–603.
- [56] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3 (1) (1991) 71–86.
- [57] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Silver Spring, MD, 1991, pp. 586–591.
- [58] D. Valentin, H. Abdi, A.J. O'Toole, G.W. Cottrell, Connectionist models of face processing: a survey, *Pattern Recognition* 27 (9) (1994) 1209–1230.
- [59] D. Valentin, H. Abdi, B.E. Edelman, A.J. O'Toole, Principal component and neural network analyses of face images: what can be generalized in gender classification? *Journal of Mathematical Psychology* 41 (4) (1997) 398–413.
- [60] L. Van Cleve, L. Johnson, P. Pothier, Pain responses of hospitalised infants and children to venipuncture and intravenous cannulation, *Journal of Pediatric Nursing* 11 (3) (1996) 161–168.
- [61] M. van Dijk, H.M. J.B.d. Boer, D. Koot, J. Tibboel, H.J. Passchier, The reliability and validity of the comfort scale as a postoperative pain instrument in 0 to 3-year-old infants, *Pain* 84 (2–3) (2000) 367–377.
- [62] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [63] F. Warnock, J. Lander, Foundations of knowledge about neonatal pain, *Journal of Pain and Symptom Management* 27 (2) (2004) 170–179.
- [64] F. Warnock, D. Sandrin, Comprehensive description of newborn distress behavior in response to acute pain (newborn male circumcision), *Pain* 107 (3) (2004) 242–255.
- [65] D. Wong, C. Baker, Pain in children: comparison of assessment scales, *Pediatric Nursing* 14 (1) (1988) 9017.
- [66] R. Xavier Balda, R. Guinsburg, M.F.B.d. Almeida, C.P.d. Araujo, M.H. Miyoshi, B.I. Kopelman, The recognition of facial expression of pain in full-term newborns by parents and health professionals, *Archives of Pediatrics and Adolescent Medicine* 154 (10) (2000) 1009–1016.
- [67] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, *Face Recognition: A Literature Survey*, Univ. of Maryland, 2000.

Sheryl Brahnam is an Assistant Professor of Computer Information Systems at Missouri State University. She received her PhD in Computer Science at the Graduate Center of the City University of New York. She is on the editorial review board of several journals and has had papers accepted and published in SIAM, in *Artificial Intelligence in Medicine* and in many well-known conference proceedings. Her research interests include face recognition, face synthesis, medical decision support systems, embodied conversational agents, computer abuse and artificial intelligence.

Chao-fa Chuang is a PhD candidate in the Department of Computer Science at New Jersey Institute of Technology. His research interests include image processing, pattern recognition and machine learning, especially support vector machines and radial basis function neural networks.

Randall S. Sexton is an Associate Professor of Computer Information Systems at Missouri State University. He received his PhD in Management Information Systems at the University of Mississippi. His research interests include computational methods, algorithm development, artificial intelligence and neural networks. His articles have been accepted or published in *Decision Support Systems*, *Decision Sciences*, *INFORMS Journal on Computing*, *European Journal of Operational Research*, *Journal of Computational Intelligence in Finance*, *Journal of End User Computing* and *OMEGA*.

Frank Y. Shih received his PhD degree from Purdue University, West Lafayette, Indiana, in Electrical Engineering. He is presently a professor jointly appointed in the Departments of Computer Science, Electrical and Computer Engineering, and Biomedical Engineering at New Jersey Institute of Technology, Newark, NJ. Dr. Shih is currently on the Editorial Board of the *International Journal of Pattern Recognition*, the *International Journal of Pattern Recognition and Artificial Intelligence*, the *International Journal of Internet Protocol Technology* and the *Journal of Internet Technology*. He won the Honorable Mention Award from the International Pattern Recognition Society for Outstanding Paper and also won the Best Paper Award in the International Symposium on Multimedia Information Processing. He has published over 180 technical papers in well-known prestigious journals and conferences. His current research interests include image processing, computer vision, sensor networks, computer graphics, artificial intelligence, bioinformatics, information security, robotics, fuzzy logic and neural networks.