Automated Pain Detection in Facial Videos of Children using Human-Assisted Transfer Learning

Xiaojing Xu[1], Kenneth D. Craig[2], Damaris Diaz[3], Matthew S. Goodwin[4], Murat Akcakaya[5], Büşra Tuğçe Susam[5], Jeannie S. Huang[3], Virginia R. de Sa[6]

[1] Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, USA, xix068@ucsd.edu

[2] Department of Psychology,University of British Columbia Vancouver, BC, Canada, kcraig@psych.ubc.ca

[3] Rady Childrens Hospital and Department of Pediatrics, University of California San Diego, CA, USA, dad003@ucsd.edu,jshuang@ucsd.edu

[4] Department of Health Sciences, Northeastern University, Boston, MA, USA, m.goodwin@northeastern.edu

[5] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA, akcakaya@pitt.edu,tugcebusraiu@gmail.com

[6] Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA, desa@ucsd.edu

Abstract. Accurately determining pain levels in children is difficult, even for trained professionals and parents. Facial activity provides sensitive and specific information about pain, and computer vision algorithms have been developed to automatically detect Facial Action Units (AUs) defined by the Facial Action Coding System (FACS). Our prior work utilized information from computer vision, i.e. automatically detected facial AUs to develop classifiers to distinguish between pain and no-pain conditions. However, application of pain/no-pain classifiers based on automated AU codings across different environmental domains resulted in diminished performance. In contrast, classifiers based on manually coded AUs demonstrated reduced environmentally-based variability in performance. To improve classification performance in the current work, we applied transfer learning by training another machine learning model to map automated AU codings to a subspace of manual AU codings to enable more robust pain recognition performance when only automatically coded AUs are available for the test data. With this transfer learning method, we improved the Area under the ROC Curve (AUC) on independent data (new participants) from our target data domain from 0.69 to 0.72.

Keywords: automated pain detection, transfer learning, facial action units

1 Introduction

In the classic model of machine learning, scientists train models on training data to accurately detect a desired outcome and apply learned models to new data measured under identical circumstances to validate their performance. Given the real world and its notable variation, it is tempting to apply learned models to data measured under similar but not identical circumstances. However, performance in such circumstances often deteriorates because of unmeasured factors not accounted for between original and new datasets. Nevertheless, lessons can be learned from similar scenarios. Transfer learning or inductive transfer in machine learning focuses on storing knowledge gained while solving one problem and applying it to a different but related problem [1]. We describe application of transfer learning to the important clinical problem of pain detection in children.

Accurate measurement of pain severity in children is difficult, even for trained professionals and parents. This is a critical problem as over-medication can result in adverse side-effects, including opioid addiction, and under-medication can lead to unnecessary suffering [2].

The current clinical gold standard and most widely employed method of assessing clinical pain is patient self-report [3]. However, this method is subjective and vulnerable to bias. Consequently, clinicians often distrust pain self-reports, and find them more useful for comparisons over time within individuals, rather than comparisons between individuals [4]. Further, infants and older children with communication/neurologic disabilities do not have the ability or capacity to self-report pain levels [3][5][6]. As a result, to evaluate pain in populations with communication limitations, observational tools based on behavioral nonverbal indicators associated with pain have been developed [7].

Of the various modalities of nonverbal expression (e.g., bodily movement, vocalizations), facial activity can provide the most sensitive, specific, and accessible information about the presence, nature, and severity of pain across the life span, from infancy [8] through to advanced age [9]. Observers largely consider facial activity during painful events to be a relatively spontaneous reaction [7].

Evaluation of pain based on facial indicators requires two steps: (1) Extraction of facial pain features and (2) pain recognition based on these features. For step (1), researchers have searched for reliable facial indicators of pain, such as the anatomically-based, objectively coded Facial Action Units (AUs) defined by the Facial Action Coding System (FACS) [10][11] (Visualizations of the facial activation units can be found at https://imotions.com/blog/facial-action-coding-system/). However, identifying these AUs traditionally requires time intensive offline coding by trained human coders, limiting their application in real-time clinical settings. Recently, algorithms to automatically detect AUs [11] have been developed and implemented in software such as iMotions (imotions.com) allowing automatical output of AU probabilities in real-time based on direct recording of face video. In step (2), Linear models [5], SVM [12] and Neural Networks [13] have been used to recognize pain based on facial features. In this paper, we first combine iMotions and Neural Networks to build an automated pain recognition model.

Although a simple machine learning model based on features extracted by a well-designed algorithm can perform well when training data and test data have similar statistical properties, problems arise when the data follow different distributions. We discovered this issue when training videos were recorded in one environment/setting and test videos in another. One way to deal with this problem is to use transfer learning, which discovers "common knowledge" across domains and uses this knowledge to complete tasks in a new domain with a model learned in the old domain [14]. In this paper, we show that features extracted from human-coded (manual) AU codings are less sensitive to domain changes than features extracted from iMotions (automated) AU codings, and thus develop a simple method that learns a projection from automated features onto a subspace of manual features. Once this mapping is learned, future automatically coded data can be automatically transformed to a representation that is more robust between domains.

To summarize, our contributions in this paper include:

- Demonstration that environmental factors modulate the ability of automatically coded AUs to recognize clinical pain in videos
- Demonstration that manually coded AUs (especially previously established "pain-related" ones) can be used to successfully recognize pain in video with machine learning across different environmental domains
- Development of a transfer learning method to transfer automated features to the manual feature space that improves automatic recognition of clinical pain across different environmental domains

2 Methods

2.1 Participants

143 pediatric research participants (94 males, 49 females) aged 12[10,15] (median [25%, 75%]) years old and primarily Hispanic (78%) who had undergone medically necessary laparoscopic appendectomy were videotaped for facial expressions during surgical recovery. Participating children had been hospitalized following surgery for post-surgical recovery and were recruited for participation within 24 hours of surgery at a pediatric tertiary care center. Exclusion criteria included regular opioid use within the past 6 months, documented mental or neurologic deficits preventing study protocol compliance, and any facial anomaly that might alter computer vision facial expression analysis. Parents provided written informed consent and youth gave written assent. The local institutional review board approved the research protocol.

2.2 Experimental Design and Data Collection

Data were collected over 3 visits (V): V1 within 24 hours after appendectomy; V2 within the calendar day after the first visit; and V3 at a follow-up visit 25 [19, 28] (median [25%, 75%]) days postoperatively when pain was expected to have fully subsided. Data were collected in two environmental conditions: V1 and V2 in hospital and V3 in the outpatient lab. At every visit, two 10-second videos (60 fps at 853x480 pixel resolution) of the face were recorded while manual pressure was exerted at the surgical site for 10 seconds (equivalent of a clinical examination). In the hospital visits (V1, V2), the participants were lying in the hospital bed with the head of the bed raised. In V3, they were seated in a reclined chair. Participants rated their pain level during the pressure using a 0-10 Numerical Rating Scale, where 0 = no-pain and 10 = worst pain ever. Following convention for clinically significant pain [15], videos with pain ratings of 0-3 were labeled as no-pain, and videos with pain ratings of 4-10 were labeled as pain, for classification purposes. 324 pain videos were collected from V1/2, 195 no-pain videos were collected from V1/2, and 235 no-pain videos were collected from V3. Figure 1 demonstrates the distribution of pain and no-pain videos across environmental domains.

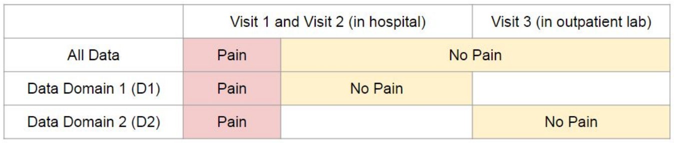
2.3 Feature Extraction

For each 10-second video sample, we extracted AU codings per frame to obtain a sequence of AUs. This was done both automatically by iMotions software (www.imotions.com) and manually by a trained human in a limited subset. We then extracted features from the sequence of AUs.

Automated Facial Action Unit Detection: The iMotions software integrates Emotient's FACET technology (www.imotions.com/emotient) which was formally known as CERT [16]. In the described work, the iMotions software was used to process the videos to automatically extract 20 AUs (AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, 43) and three head pose indicators (yaw, pitch and roll) from each frame. The values of these codings are the estimated log probabilities of AUs, ranging from -4 to 4.

Manual Facial Action Unit Detection: A trained human FACS AU coder manually coded 64 AUs (AU1-64) for each frame of a subset of videos by labeling the AU intensities (0-5, 0 = absence).

Feature Dimension Reduction: The number of frames in our videos were too large to use full sequences of frame-coded AUs. To reduce dimensionality, we applied 11 statistics (mean, max, min, standard deviation, 95th, 85th, 75th, 50th, 25th percentiles, half-rectified mean, and max-min) to each AU over all frames to obtain 11×23 features for automatically coded AUs, and 11×64 features for manually coded AUs. We call these automated features and manual features, respectively. The range of each feature was rescaled to $[0, 1]$ to normalize features over the training data.



	Visit 1 and Visit 2 (in hospital)		Visit 3 (in outpatient lab)
All Data	Pain	No Pain	
Data Domain 1 (D1)	Pain	No Pain	
Data Domain 2 (D2)	Pain		No Pain

Fig. 1. Data Domain Illustration

2.4 Machine Learning Models

Neural Network Model to Recognize Pain with Extracted Features: A neural network with 1 hidden layer was used to recognize pain with extracted automated or manual features. The number of neurons in the hidden layer was twice the number of neurons in the input layer, and the Sigmoid activation function $\sigma(x) = 1/(1 + \exp(-x))$ was used with batch normalization for the hidden layer. The output layer used cross-entropy error.

Neural Network Model to Predict Manual Features with Automated Features: A neural network with the same structure was used to predict manual features from automated features, except that the output layer was linear and mean squared error was used as the loss function.

Model Training and Testing: Experiments were conducted in a participant-based (each participant restricted to one fold) 10-fold cross-validation fashion. Participants were divided into 10 folds, and each time 1 fold was used as the test set, and the other 9 folds together were used as the training set. We balanced classes for each participant in each training set by duplicating samples from the under-represented class. 1/9 participants in the training set were picked randomly as a nested-validation set for early stopping in the neural network training. A batch size of 1/8 the size of training set was used. We examined the receiver operating characteristic curve (ROC curve) which plots True Positive Rate against False Positive Rate as the discrimination threshold is varied. We used the Area under the Curve (AUC) to evaluate the performance of classifiers. We considered data from 3 domains (D) as shown in Figure 1: (1) D1 with pain and no-pain both from V1/2 in hospital, (2) D2 with pain from V1/2 in hospital and no-pain from V3 from outpatient lab, and (3) All data, i.e., pain from V1/2 and no-pain from V1/2/3. The clinical goal was to be able to discriminate pain levels in the hospital; thus evaluation on D1 (where all samples were from the hospital bed) was the most clinically relevant evaluation.

3 Analysis and Discussion

Data of 73 participants labeled by both human and iMotions were used through section 3.1 to 3.5, and data of the remaining 70 participants having only automated (iMotions) AU codings were used for independent test set evaluation in the results section.

Table 1. AUC for Classification with SEM (Standard Error of The Mean)

Train on	Test on	Automated	Manual	Automated "Pain" Features	Manual "Pain" Features
All	D1	0.61 ± 0.007	0.66 ± 0.007	0.63 ± 0.007	$\mathbf{0.69 \pm 0.006}$
D1	D1	0.6 ± 0.009	0.61 ± 0.008	0.62 ± 0.007	$\mathbf{0.65 \pm 0.007}$
D2	D1	0.58 ± 0.006	0.66 ± 0.006	0.61 ± 0.005	$\mathbf{0.7 \pm 0.005}$
All	D2	0.9 ± 0.005	0.8 ± 0.006	0.89 ± 0.005	0.8 ± 0.004
D1	D2	0.72 ± 0.006	0.69 ± 0.008	0.75 ± 0.008	0.74 ± 0.007
D2	D2	0.93 ± 0.005	0.79 ± 0.008	0.91 ± 0.003	0.8 ± 0.006

3.1 Automated Classifier Performance Varies by Environment

Using automated features, we first combined all visit data and trained a classifier to distinguish pain from no-pain. This classifier performed well in general (AUC $= 0.77 \pm 0.011$ on All data), but when we looked at different domains, the performance of D1 (the most clinically relevant in-Hospital environment) was much inferior to that on D2, as shown in data rows 1 and 4 under the "Automated" column in Table 1.

There were two main differences between D1 and D2, i.e. between V1/2 and V3 no-pain samples. The first was that in V1/2, patients usually still had some pain and their self-ratings were greater than 0, while in V3, no-pain ratings were usually 0 reflecting a "purer" no-pain signal. The second difference was that V1/2 happened in hospital with patients in beds and V3 videos were recorded in an outpatient lab with the patient sitting in a reclined chair. Since automated recognition of AUs is known to be sensitive to facial pose and lighting differences, we hypothesized that the added discrepancy in classification performance between D1 and D2 was mainly due to the model classifying based on environmental differences between V1/2 and V3. In other words the classifier when trained and tested on D2, might be classifying "lying in hospital bed" vs "more upright in outpatient chair" as much as pain vs no-pain. (This is similar to doing well at recognizing cows by recognizing a green background).

In order to investigate this hypothesis and attempt to improve classification on the clinically relevant D1, we trained a classifier using only videos from D1. Within the "Automated" column, row 2 in Table 1 shows that performance on automated D1 classification doesn't drop much when D2 samples are removed from the training set. At the same time, training using only D2 data results in the worst classification on D1 (row 3), but the best classification on D2 (last row) as the network is able to exploit the environmental differences (no-pain+more upright from V3, pain+lying-down from V1/2).

Figure 2 (LEFT) shows ROC curves of within and across domain tests for models trained on automated features in D2. The red dotted curve corresponds to testing on D2 (within domain) and the blue solid curve corresponds to testing on D1 (across domain). The model did well on within domain classification, but failed on across domain tasks.

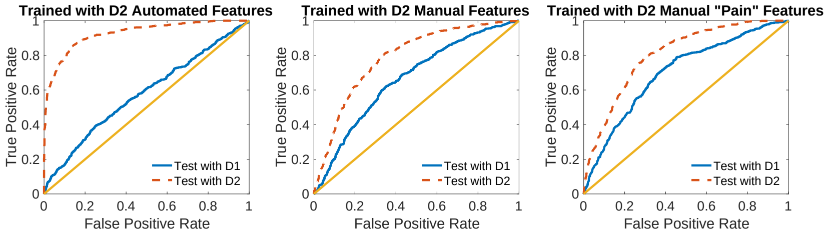


Fig. 2. ROC Curves for classification within and across domains using automated features (left), manual features (middle) and pain-related manual features (right). The red dotted lines are ROCs when the machine is able to use environment information to differentiate pain and no-pain conditions, and the blue solid lines show the machine's ability to discriminate between pain and no-pain based on AU information alone. The straight yellow line graphs the performance of random chance.

3.2 Classification Based on Manual AUs Are Less Sensitive to Environmental Changes

We also trained a classifier on manual AUs labeled by a human coder. Interestingly, results from the classifier trained on manual AUs showed less of a difference in AUCs between the domains, with a higher AUC for D1 and a lower AUC for D2 relative to those with the automated AUs (see Table 1 “Manual” and “Automated” columns). The manual AUs appeared to be less sensitive to changes in the environment reflecting the ability of the human labeler to consistently code AUs without being affected by lighting and pose variations.

When we restricted training data from All to only D1 or only D2 data, classification performance using manual AUs went down, likely due to the reduction in training data, and training with D2 always gave better performance than training with D1 on both D1 and D2 test data, which should be the case since D2 is higher in “pain” quality. These results appear consistent with our hypothesis that human coding of AUs is not as sensitive as machine coding of AUs to the environmental differences between V1/2 and V3.

Figure 2 (MIDDLE) displays the ROC curves for manual features. As discussed above, in contrast to the plot on the left for automated features, manual coding performance outperformed automated coding performance in the clinically relevant test in D1. The dotted red curve representing within domain performance is only slightly higher than the solid blue curve, likely due in part to the quality difference in no-pain samples in V1/2 and V3 and also possibly due to any small amount of environmental information that the human labeler was affected by. Note that ignoring the correlated environmental information in D2 (pain faces were more reclined and no-pain faces were more upright) resulted in a lower numerical performance on D2 but does not likely reflect worse classification of pain.

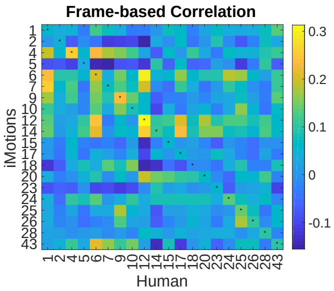


Fig. 3. Correlation matrix of AU pairs from automated and manual codings using All data

3.3 Restricting Manual AUs to Those Associated with Pain Improves Classification

In an attempt to reduce the influence of environmental conditions to further improve accuracy on D1, we restricted the classifier to the eight AUs that have been consistently associated with pain: 4 (Brow Lowerer), 6 (Cheek Raiser), 7 (Lid Tightener), 9 (Nose Wrinkler), 10 (Upper Lip Raiser), 12 (Lip Corner Puller), 20 (Lip Stretcher), and 43 (Eyes Closed) [17, 18] to obtain 11 (statistics) ×8 (AUs) features. Pain prediction results using these "pain" features are shown in the last two columns in Table 1. Results show that using only pain-related AUs improved the classification performances of manual features. However, it did not seem to help as much for automated features.

Similarly, Figure 2 (RIGHT) shows that limiting manual features to use only pain-related AUs further improved D1 performance when training with D2. We also performed PCA on these pain-related features and found that performance in the hospital environmental domain was similar if using 4 or more principal components.

3.4 iMotions AUs Are Different Than Manual FACS AUs

Computer Vision AU automatic detection algorithms have been programmed/ trained on manual FACS data. However, we demonstrated differential performance of AUs encoded automatically versus manually. To understand the relationship between automatically encoded v. manually coded AUs, we computed correlations between automatically coded AUs and manually coded AUs at the frame level as depicted in Figure 3. If two sets of AUs were identical, the diagonal of the matrix (marked with small centered dots) should yield the highest correlations, which was not the case. For example, manual AU 6 was highly correlated with automated AU 12 and 14, but had relatively low correlation with automated AU 6.

The correlation matrix shows that not only are human coders less affected by environmental changes, the AUs they code are not in agreement with the

Table 2. AUC (and SEM) with Transferred Automated Features

Train on	Test on	All Features	“Pain” Features	7 PCs	4 PCs	1 PC
All	D1	0.62 ± 0.007	0.64 ± 0.01	0.68 ± 0.009	**0.69 ± 0.006**	0.63 ± 0.008
D1	D1	0.64 ± 0.008	0.65 ± 0.009	**0.67 ± 0.011**	**0.67 ± 0.01**	0.63 ± 0.013
D2	D1	0.58 ± 0.006	0.59 ± 0.01	0.66 ± 0.007	**0.68 ± 0.006**	0.65 ± 0.006
All	D2	0.84 ± 0.011	0.83 ± 0.009	0.76 ± 0.01	0.75 ± 0.011	0.67 ± 0.012
D1	D2	0.69 ± 0.008	0.72 ± 0.013	0.71 ± 0.011	0.71 ± 0.013	0.66 ± 0.017
D2	D2	0.89 ± 0.004	0.85 ± 0.008	0.77 ± 0.007	0.74 ± 0.01	0.68 ± 0.009

automated AUs. (We separately had another trained human coder code a subset of the videos and observed closer correlation between the humans than between each human and iMotions). This likely explains the reduced improvement from restricting the automated features model to “pain-related AUs” as these have been determined based on human FACS coded AUs.

3.5 Transfer Learning via Mapping to Manual Features Improves Performance

We have shown that manual codings are not as sensitive to domain change. However, manual coding of AUs is very time-consuming and not amenable to an automated real-time system. In an attempt to leverage manual coding to achieve similar robustness with automatic AUs, we utilized transfer learning and mapped automated features to the space of manual features. Specifically, we trained a machine learning model to estimate manual features from automated features using data coded by both iMotions and a human. Separate models were trained to predict: manual features of 64 AUs, manual features of the eight pain-related AUs, principal components (PCs) of the manual features of the eight pain-related AUs. PCA dimensionality reduction was used due to insufficient data for learning a mapping from all automated AUs to all manual AUs.

Once the mapping network was trained, we used it to transform the automated features and train a new network on these transformed data for the pain/no-pain classification. The 10-fold cross-validation was done consistently so that the same training data was used to train the mapping network and the pain-classification network.

In Table 2, we show the classification AUCs when the classification model was trained and tested with outputs from the prediction network. We observed that when using All data to train (which had the best performance), with the transfer learning prediction network, automated features performed much better in classification on D1 ($0.68-0.69$ compared to $0.61-0.63$ in Table 1). Predicting 4 principal components of manual pain-related features gave the best performance on our data. Overall, the prediction network helped in domain adaptation of a pain recognition model using automatically extracted AUs.

Figure 5 (LEFT) plots the ROC curves using the transfer learning classifier within and across domains trained and tested using 4 predicted features. Com-

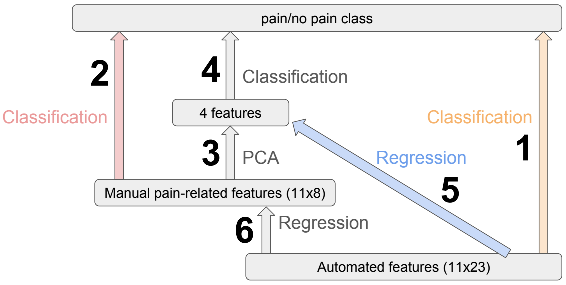


Fig. 4. Illustration of Machine Learning Models. 1/2 are classifications using automated/manual pain features, in which 2 does better than 1. 3-4 can be done to reduce feature dimensions while maintaining the performance. 6-2 and 5-4 are our transfer learning models, training a regression network to map automated features to a subspace of manual pain features before classification.

pared to Figure 2 (LEFT), the transferred automated features showed properties more like manual features, with smaller differences between performances on the two domains and higher AUC on the clinically relevant D1. Table 2 shows numerically how transfer learning helped automated features to ignore environmental information in D2 like humans, and learn pure pain information which can also be used in classification on D1.

Within domain classification performance for D1 was also improved with the prediction network. These results show that by mapping to the manual feature space, automated features can be promoted to perform better in pain classification.

4 Results

In the previous section, we showed that in Figure 4, classification with pain-related pain features (2) performed better than automated features (1) on D1, which was the clinically relevant classification. We have also found that applying PCA to manual features (3-4) didn't change the performance on D1 much. Thus we introduced a transfer learning model to map automated features first to manual pain-related features (or the top few principal components of them), and then used the transferred features for classification (6-2, or 5-4). We got similar results to manual features on D1 with transfer learning model (5-4) mapping to 4 principal components of manual features.

In this section we report on the results from testing our transfer learning method on a new separate dataset (new participants), which has only automated features. Table 1 shows that without our method, training on all data and restricting to pain-related AUs resulted in the best performance for D1. And cross-validation results in Table 2 shows that with our method, predicting 4 PCs yielded the best performance for D1. With these optimal choices of model structure and training domain, we trained two models

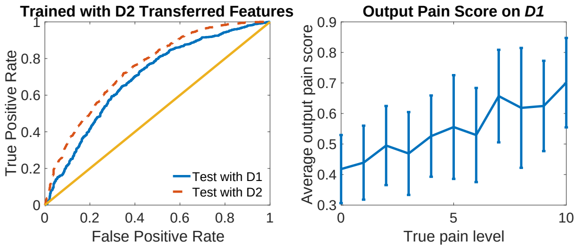


Fig. 5. ROC Curves for classification within and across domains using our transfer learning model (left) and plot of average model output pain score (with error bars indicating standard deviation) over true pain level (right)

using all the data in the previous sections labeled by both iMotions and human, and tested the model on a new separate data set (new participants) only labeled by iMotions (*D1, D2*). Our model with transfer learning (AUC=0.72 ± 0.009) performed better than the model without it (AUC=0.69 ± 0.033) on *D1* with p-value=$5.4585e - 04$.

Figure 5 (RIGHT) plots output pain scores of our model tested on *D1* versus 0-10 self-reported pain levels. The model output pain score increases with true pain level indicating that our model indeed reflects pain levels.

5 Conclusion

In the described work, we recognized differences in classifier model performance (on pain vs no-pain) across data domains that reflected environmental differences as well as differences reflecting how the data were encoded (automatically v. manually). We then introduced a transfer learning model to map automated features first to manual pain-related features (or principal components of them), and then used the transferred features for classification (6-2, or 5-4 in Figure 4). This allowed us to leverage data from another domain to improve classifier performance on the clinically relevant task of distinguishing pain levels in the hospital.

Acknowledgments

This work was supported by National Institutes of Health National Institute of Nursing Research grant R01 NR013500 and by IBM Research AI through the AI Horizons Network. Many thanks to Ryley Unrau for manual FACS coding and Karan Sikka for sharing his code and ideas used in [5].

References

1. Jeremy West, Dan Ventura, and Sean Warnick. Spring research presentation: A theoretical foundation for inductive transfer. *Brigham Young University, College of Physical and Mathematical Sciences*, 1, 2007.

2. Brenna L Quinn, Esther Seibold, and Laura Hayman. Pain assessment in children with special needs: A review of the literature. *Exceptional Children*, 82(1):44–57, 2015.
3. Ghada Zamzmi, Chih-Yun Pai, Dmitry Goldgof, Rangachar Kasturi, Yu Sun, and Terri Ashmeade. Machine-based multimodal pain assessment tool for infants: a review. *preprint arXiv:1607.00331*, 2016.
4. Carl L Von Baeyer. Childrens self-report of pain intensity: what we know, where we are headed. *Pain Research and Management*, 14(1):39–45, 2009.
5. Karan Sikka, Alex A Ahmed, Damaris Diaz, Matthew S Goodwin, Kenneth D Craig, Marian S Bartlett, and Jeannie S Huang. Automated assessment of childrens postoperative pain using computer vision. *Pediatrics*, 136(1):e124–e131, 2015.
6. Min SH Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, et al. The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal emopain dataset. *IEEE transactions on affective computing*, 7(4):435–451, 2016.
7. Kamal Kaur Sekhon, Samantha R Fashler, Judith Versloot, Spencer Lee, and Kenneth D Craig. Childrens behavioral pain cues: Implicit automaticity and control dimensions in observational measures. *Pain Res Manag.*, 2017.
8. Ruth VE Grunau and Kenneth D Craig. Pain expression in neonates: facial action and cry. *Pain*, 28(3):395–410, 1987.
9. Thomas Hadjistavropoulos, Keela Herr, Kenneth M Prkachin, Kenneth D Craig, Stephen J Gibson, Albert Lukas, and Jonathan H Smith. Pain assessment in elderly adults with dementia. *The Lancet Neurology*, 13(12):1216–1227, 2014.
10. Paul Ekman and Wallace V Friesen. Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1):56–75, 1976.
11. Brais Martinez, Michel F Valstar, Bihan Jiang, and Maja Pantic. Automatic analysis of facial actions: A survey. *IEEE Trans on Affective Computing*, 2017.
12. Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M Prkachin, and Patricia E Solomon. The painful face–pain expression recognition using active appearance models. *Image and vision computing*, 27(12):1788–1796, 2009.
13. Md Maruf Monwar and Siamak Rezaei. Pain recognition using artificial neural network. In *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, pages 28–33. IEEE, 2006.
14. Sinno J. Pan and Qiang Yang. A survey on transfer learning. *IEEE Trans on knowledge and data engineering*, 22(10):1345–1359, 2010.
15. DL Hoffman, A Sadosky, EM Dukes, and J. Alvir. How do changes in pain severity levels correspond to changes in health status and function in patients with painful diabetic peripheral neuropathy. *Pain*, 149(2):194–201, May 2010.
16. Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. The computer expression recognition toolbox (cert). In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 298–305. IEEE, 2011.
17. Kenneth M Prkachin. The consistency of facial expressions of pain: a comparison across modalities. *Pain*, 51(3):297–306, 1992.
18. Kenneth M Prkachin. Assessing pain by facial expression: facial expression as nexus. *Pain Res Manag.*, 14(1):53–58, 2009.