Real-time pain detection in facial expressions for health robotics

1st Laduona Dai
University of Twente
Human Media Interaction
Enschede, The Netherlands
laduona.dai@gmail.com

2nd Joost Broekens
LIACS
Leiden University
Leiden, The Netherlands
joost.broekens@gmail.com (corresponding)

2nd Khiet P. Truong
Human Media Interaction
University of Twente
Enschede, The Netherlands
k.p.truong@utwente.nl

***Abstract*—Automatic pain detection is an important challenge in health computing. In this paper we report on our efforts to develop a real-time, real-world pain detection system from human facial expressions. Although many studies addressed this challenge, most of them use the same dataset for training and testing. There is no cross-check with other datasets or implementation in real-time to check performance on new data. This is problematic, as evidenced in this paper, because the classifiers overtrain on dataset-specific features. This limits real-time, real-world usage. In this paper, we investigate different methods of real-time pain detection. The training data uses a combination of pain and emotion datasets, unlike other papers. The best model shows an accuracy of 88.4% on a dataset including pain and 7 non-pain emotional expressions. Results suggest that convolutional neural networks (CNN) are not the best methods in some cases as they easily overtrain if the dataset is biased. Finally we implemented our pain detection method on a humanoid robot for physiotherapy. Our work highlights the importance of cross-corpus evaluation & real-time testing, as well as the need for a well balanced and ecologically valid pain dataset.**

***Index Terms*—Pain detection, classification, generalization, cross validation, health**

I. Introduction

It has been shown that there is an increasing demand for rehabilitation services and therapeutic robotics due to the shortage of therapists and the increasing number of patients [1]. For example, the lifetime prevalence of frozen shoulder is about 2 to 5 percent of the general population [2]. With the use of therapeutic robotics, it is possible to increase the efficiency of resource use and give more patients therapy sessions at the right time. There is already much development in robotics for this purpose. For instance, Klevin et al. [3] developed robot-assisted rehabilitation of hand function, Burgar et al. [4] proposed three robot systems for post-stroke therapy, Lum et al. [5] compared robot-assisted rehabilitation of upper-limb motor function with conventional therapy. Most of these systems focus on the functionality of the robot and motor-related rehabilitation. The patients' feelings and demands are often neglected. Pain is an important feeling, also for recovery. In some cases, the patient can be demotivated about the therapy due to the fear of pain which limits the effect of rehabilitation [6]. Patients without pain catastrophizing (i.e., interpreting pain as threatening) usually lead to faster recovery [7]. Professional therapists would give motivating feedback to patients based on their behavior during therapy [8]. For diseases like chronic pain, it requires long term treatment, and even with successful treatment patients need to do self-management afterwards [9]. Therefore, a robotic system with pain detection and the ability to give proper feedback during therapy (like chronic pain) is needed.

By definition, pain is a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive and social components [10]. In practice, it is hard for health-care providers to directly measure a patient's pain intensity. Usually, this is done by the patient's self-report [11]. Since it has been shown that facial expressions are related to pain intensity [12], many have proposed machine learning methods to detect pain from facial expressions. For example, Sourav el at. [13] report 87.23% accuracy for the detection of pain at the frame level, Reza el at. [14] reported 87.2% accuracy for pain detection, and Pau et at. [15] use an LSTM network and report a 93.3 for area under curve (AUC) score. However, most papers on pain detection from facial expression use the same dataset (UNBC-McMaster Shoulder Pain expression archive database [16]) for training and testing, and there is no cross-corpus evaluation (testing with other pain datasets), and no implementation for real-time pain detection to validate the model.

In this work, we first show that this focus on one specific dataset limits generalisation of the trained models, and results in models that learned to detect dataset specific features. Then we show how mixing datasets can partly resolve this. Different training datasets are used to ensure optimal performance for real-time detection. A cross-corpus test is performed to test the model's pain prediction ability on other datasets. We also experiment with both facial action unit (AU) based (feature-based) methods and non-AU based (end-to-end) methods to predict pain in real-time. Finally we implement our method on a robot for frozen shoulder therapy demonstration.

II. Datasets

A. UNBC-McMaster Shoulder Pain Dataset

The UNBC-McMaster Shoulder Pain dataset [16] contains 200 video sequences of 25 subjects with spontaneous facial expressions who are suffering from shoulder pain. It has in

978-1-7281-3891-6/19/$31.00 ©2019 IEEE

total 48,398 FACS (Facial Action Coding System) [17] coded frames. The dataset also contains pain scores for each frame based on PSPI (Prkachin and Solomon pain intensity) scale [18]. The PSPI scale is defined in 17 levels with the help of FACS. The calculation is shown in below equation:

$$PSPI = AU4 + (AU6 \ or \ AU7) + (AU9 \ or AU10) + AU43$$

The PSPI score is the sum of AU4, AU6 or AU7 (whichever is higher in intensity), AU9 or AU10 (whichever is higher in intensity) and AU43. The AUs are: brow-lowering (AU4), cheek-raising (AU6), eyelid tightening (AU7), nose wrinkling (AU9), upper-lip raising (AU10) and eye-closure (AU43). Apart from AU43(0 = absent, 1 = present), each AU is coded in 6 level intensity(0 = absent, 5 = maximum). Figure 1 shows an example of the same face with different PSPI levels. According to the results from [13], the detection for strong pain(PSPI$\geq$3) would give a more stable result, therefore, the pain dataset is pre-processed into a NO-PAIN(PSPI=0) and PAIN(PSPI$\geq$3) subset. For the PAIN subset, there are about 3000 images from 25 subjects, for the NO-PAIN subset, there are about 40k images in total.

Fig. 1. Example of faces with different PSPI levels from UNBC-McMaster dataset [16]

B. BioVid Heat Pain Dataset

The BioVid Heat Pain [19] uses heat as the source for induced pain with 4 intensities. The temperatures for the heat are adjusted based on each subject's pain threshold and pain tolerance. In this report, part A of the dataset is used. For this part of the dataset, there are 87 subjects, 5 classes (baseline, pain level 1, pain level 2, pain level 3 and pain level 4). Each class of a subject has 20 samples, the average length of a sample is about 5 seconds. Figure 2 shows the pattern of the heat stimuli, all the heat stimuli last for 4s and connect with an 8-12s pause.

C. Facial Expression Recognition 2013 Dataset (FER-2013)

Because we want to investigate the mixing of face datasets in order to enhance pain detection reliability, we also use 2 facial affect datasets. The first is the FER-2013 dataset

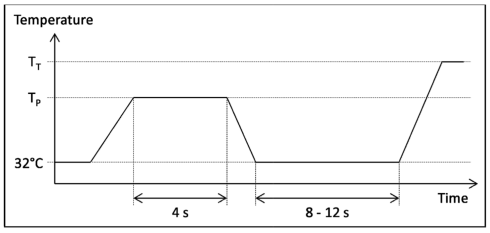

Fig. 2. Pattern of the heat stimuli in BioVid Heat Pain dataset [20]

[21], it is part of the ICML 2013 workshop "Challenges in Representation Learning". It consists of 35887 gray-scale images of faces(48x48 pixel), with 4953 'Anger' images, 547 'Disgust' images, 5121 'Fear' images, 8989 'Happiness' images, 6077 'Sadness' images, 4002 'Surprise' images, and 6198 'Neutral' images. Since the dataset was created using the Google image search API to search for images of faces, image labels contain errors, and human accuracy on the dataset is 65%. Figure 3 shows some example faces in FER-2013 dataset.

Fig. 3. Examples of some faces in FER-2013 dataset.

D. AffectNet Dataset

The second facial affect dataset used is AffectNet [22], it is a dataset containing images of facial expressions in the wild, and it contains about 1 million high-resolution facial images collected from the Internet by querying three major search engines(Google, Bing, and Yahoo). About half of the retrieved images ($\sim$440K) were manually annotated for the presence of seven discrete facial expressions (neutral, happy, sad, surprise, fear, disgust and anger).

III. Research Questions

We investigate if state-of-the-art machine learning in combination with frequently used pain datasets can result in real-time real-world performance. In more detail we focus on the following issues:

- How do models trained on just one pain dataset generalize?
- What is the effect of mixing affect and pain datasets on model generalization?

- What is the best method for real-time pain detection, given current dataset limitations?

IV. Experiments

The method, result, and discussion for each experiment will be presented in this section. Finally, a cross-check is performed on a different pain dataset to check generalization of the model. All experiments are conducted on a laptop configuration consisting of a 2.3 GHz Intel Core i5 and 16 GB memory. The real-time detection is tested on the laptop's camera with several subjects. In this paper, 5 new datasets are made for different experiments. Their composition is shown in Table I.

A. AU based methods

For the purpose of automatic facial action unit recognition, OpenFace 2.0 is utilized [23] [24]. This toolkit is the only recent free software that provides facial action unit recognition for both real-time video and a single image according to the author's knowledge. According to [23], the average detection accuracy for Action Units is about 60%.

The general steps for the AU based method are shown in Figure 4. The datasets that have been tried include the McMaster Pain dataset alone or in combination with FER-2013/AffectNet. The first step is to use OpenFace on all the frames in the dataset. The output would be a CSV file with information for each frame, like frame id, the success of face detection, and values for each AU. Before further processing, frames in which no face was detected by OpenFace were deleted. To train the dataset using a support vector machine (SVM), the labels for each frame are extracted. Therefore, the inputs for SVM are 6 AU intensity values with a label of 1 or 0(pain/no-pain). Table II shows an example of the inputs and label for the SVM.

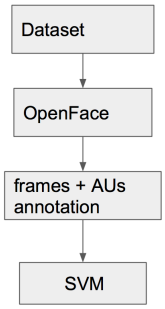

Fig. 4. Flowchart for general steps of AU based methods.

1) Dataset 1 using SVM: Since most papers about pain detection using facial expression only train and test the result on the McMaster Shoulder Pain dataset, it is reasonable to first check how the proposed method works on this dataset as the baseline. Considering this dataset is extremely uneven in its distribution between pain and non-pain frames (3k frames of pain vs 40k frames of non-pain), the same number of no-pain frames are randomly retrieved from PSPI=0 images. Then the no-pain frames and pain frames are combined as a new balanced Dataset 1. OpenFace is then used to detect AU4, AU6, AU7, AU9, AU10 and AU45 values for each frame (AU43 eye-closure is not supported so we use AU45, the eye-blink instead). These AUs values are used as input for a SVM with 8:2 train-test ratio, 5 fold cross-validation and default parameter setting from Scikit-learn library ('C': 1.0, 'class weight': None, 'gamma': 'auto', 'kernel': 'rbf'). We did not control for unseen subjects in the different folds, i.e., each folds might contain all subjects.

The final result on the test set is 85% accuracy, which is comparable with other papers' result of about 88% accuracy. However, when this model is used in real time to detect posed pain by the experimenters as a pilot, it classifies *all* frames that contain *facial movements* as pain. Since the training data only has pain faces and no-pain faces, and the no-pain faces in this dataset are neutral faces without any facial movement, the model will not be able to differentiate pain faces from other faces that have facial activity. This means that the high level of accuracy reported in many other papers is probably not pain detection accuracy but movement detection. In a real-life scenario, a patient would have more facial expressions than just pain and no-pain, so using only the McMaster shoulder pain dataset as training data will not work for real-time detection.

2) Dataset 2 using SVM: In order to make the detection model more robust to normal facial movements, the AffectNet dataset is used to increase the variation of training data. 3k images are randomly retrieved from each of the 7 emotions in AffectNet (neural, happy, sad, surprise, fear, disgust and angry) and then combined with 3k pain faces from McMaster to form Dataset 2 (in total 24k images). The AUs values are extracted by OpenFace and used as input features for SVM like baseline test. But this time SVM is using grid search for finding the best parameters.

The final result is shown in Figure 5, the parameters found from the grid search are: ('C': 245.82510572851524, 'class weight': None, 'gamma': 0.23273590717450615, 'kernel': 'rbf'). The average accuracy for this model is only 40%, but the purpose of this model is not to differentiate these emotions, the aim is to separate pain from other emotions. If only looking at the precision for pain, it's about 88.4% and reaches the same level for only pain/no-pain detection. In real-time detection, we used a binary output (pain or no-pain): if any of the 7 emotions are detected instead of pain then the output is mapped to no-pain. This model appeared more robust than the previous in pilot tests with posed pain faces.

B. Non AU based methods

Recently, many state-of-the-art methods in computer vision have utilized Convolutional Neural Networks (CNN) to achieve great results in face-detection and emotion classification. CNN's are able to extract features from the training set without manual feature construction, but need high number of examples.

TABLE I
COMPOSITIONS OF 5 NEW DATASETS

	Composition
Dataset 1	McMaster (3k pain + 3k no-pain)
Dataset 2	McMaster (3k pain) + AffectNet (3k neutral, 3k happy, 3k sad, 3k surprise, 3k fear, 3k disgust and 3k angry)
Dataset 3	McMaster (3k pain) + FER-2013 (3k neutral, 3k happy, 3k sad, 3k surprise, 3k fear, 3k disgust and 3k angry)
Dataset 4	AffectNet (3k neutral, 3k happy, 3k sad, 3k surprise, 3k fear, 3k disgust and 3k angry)
Dataset 5	McMaster (3k pain+ 3k no-pain) + AffectNet (3k happy, 3k sad, 3k surprise, 3k fear, 3k disgust and 3k angry)

TABLE II
EXAMPLE OF THE INPUTS AND LABEL FOR SVM

AU4	AU6	AU7	AU9	AU10	AU45	Label
2.10	0.81	1.16	0.49	0.65	0.66	0

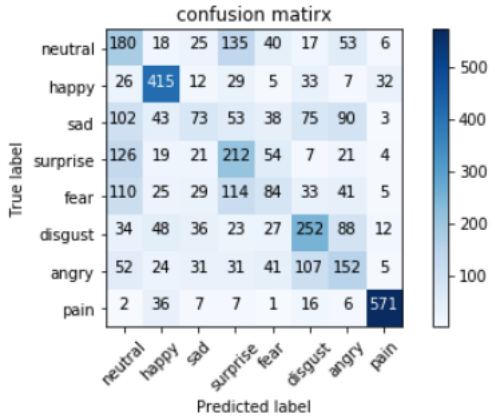

Fig. 5. Confusion matrix for pain + AffectNet 7 emotions(numbers in matrix are absolute numbers)

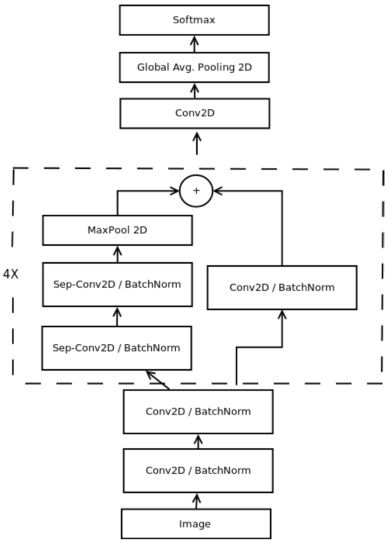

Fig. 6. The architecture of CNN [25]

Considering that a pain face is also a facial expression, it is worthwhile to try a CNN that performs well at emotion classification on pain detection. The chosen CNN model [25] contains 4 residual depth-wise separable convolutions (combination of residual modules [26] and depth-wise separable convolutions [27]). It achieves on average about 63% on classifying the 7 facial expressions on the FER-2013 dataset. Figure 6 shows the CNN architecture.

1) Dataset 3 using CNN: Since the CNN architecture mentioned above achieved a good result on the FER-2013 dataset, and since pain detection in real time needs a more varied dataset than just pain/no pain faces, it is reasonable to first try the combination of FER-2013 with pain frames from McMaster. For this purpose, in total, about 21k images are retrieved from FER-2013 dataset with an even distribution among the 7 emotions (neural, happy, sad, surprise, fear, disgust and angry) as no-pain faces and then combined with 3k pain faces from McMaster to form Dataset 3. By using 5 fold cross-validation, the final accuracy is about 97%. But in a real-time detection pilot (output is again binary, if any of the 7 emotions are detected then output no-pain), all posed emotional faces (including posed pain face) are classified as no-pain with 0.9 confidence. This result suggests that the pain features the CNN learned from the dataset are very specific for the individuals in the dataset, i.e. the CNN learned to recognize the difference between pain faces in McMaster and emotional faces in AffectNet.

2) Dataset 4 using CNN: In the AU based method, the combination of pain frames + AffectNet dataset gives good results(Section IV-A2). We now compare the result of the CNN architecture with the AU based method reported above and investigate whether the combination of pain frames + FER-2013 dataset caused the problem or not. This CNN structure is first validated on the AffectNet dataset alone, with 3000 images randomly retrieved from each of the 7 emotions in AffectNet to form a new Dataset 4. Then this dataset is trained with 5 fold cross-validation, and the final result is shown in Figure 7. The average accuracy is about 54%, for the 7 emotions classification (without pain), it performs better than the AU based method.

3) Dataset 2 using CNN: In order to check whether the CNN's performance on the FER/McMaster combined dataset is dataset specific, we trained the model on Dataset 2 (AffectNet/McMaster). The results are shown in Figure 8. The precision for pain detection is about 99%, but in real time all posed pain faces are again classified as no-pain with 0.9 confidence, just like FER-2013+ McMaster pain frames. This result confirms the CNN indeed learned to detect subject-specific (or dataset specific) features that identify the difference between the two datasets.

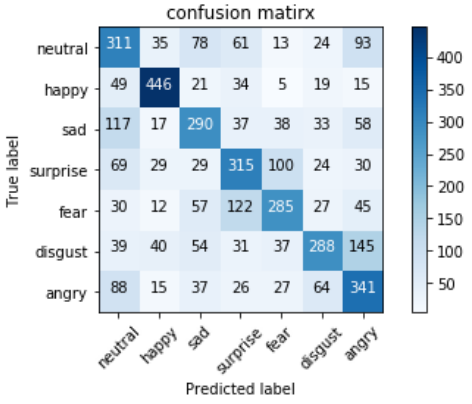

Fig. 7. Confusion matrix for AffectNet 7 emotions using CNN

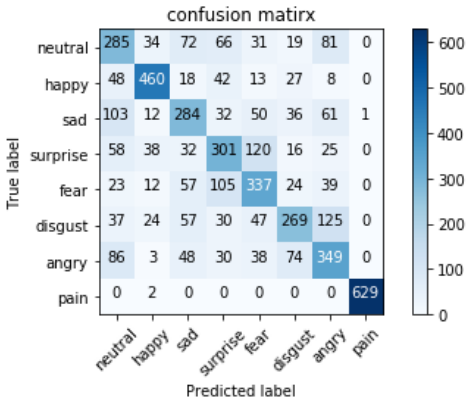

Fig. 8. Confusion matrix for pain + AffectNet 7 emotions using CNN

4) Dataset 5 using CNN: At this point, there are two possible explanations for the lack of real-time pilot performance. First, according to the behaviors of the model for Dataset 2 and Dataset 3 with the CNN, it is possible that the CNN learned to classify the 25 subjects in the McMaster dataset as pain, based on subject/dataset characteristic rather than pain characteristics. Second, it is possible that it learns to detect a hierarchically higher class due to specific differences in the images and then learned to classify *for that subclass* whether or not the participant expressed pain (e.g. based on movement again). To tear apart the two possible explanations, we construct a new dataset as follows. We combine 3000 pain frames + 3000 no-pain frames from McMaster and 3000 images randomly retrieved for each of the 6 emotions (happy, sad, surprise, fear, disgust and angry) in AffectNet. The 3000 no-pain frames are used as neutral faces. The result of this dataset is shown in Figure 9. If the CNN does learn 25 subjects face instead of pain faces, now the neutral face would be misclassified as pain. But there is no misclassified neutral to pain as seen in the confusion matrix.

In a real-time pilot we found that this model classifies all posed facial expressions including posed pain to the no-pain category. As such, the only possible explanation is that this CNN is too 'clever' and indeed learned to recognize pain expression for 25 *specific* subjects. The extreme result for pain classification is therefore produced by a combination of dataset features and subject specific pain expression. This is a classical example of over training due to dataset bias and size, there is not enough variation in individuals for the pain dataset, and as a result the network learns to predict pain for the specific individuals in that dataset. In this dataset images for the 6 emotions are from different people and pain images are from only 25 subjects. Therefore, by using this dataset, the CNN learned 25 specific pain faces. To fix this problem, a CNN with fewer layers could be used, or one could build a pain dataset with more subjects and different variations of facial expressions. Using a different split for the folds will not help that much as the number of subjects is simply not large enough for generalization to unseen subjects based on image data. It shows the importance of cross-checking with other datasets & real-time testing, as well as the need for a well balanced and ecologically valid pain dataset.

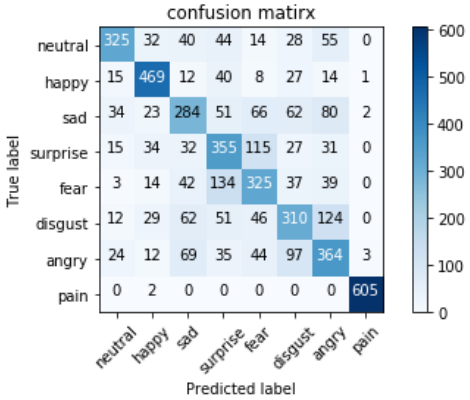

Fig. 9. Confusion matrix for pain + no-pain + AffectNet 6 emotions using CNN

C. Cross checking with the Biovid dataset

The only model that was able to classify real-time pain expressions in our pilots is the AU-based method(Section IV-A2) trained on Dataset 2 (AffectNet/McMaster). To test the performance of this pain detection model on novel data, the Biovid dataset is used for cross-corpus evaluation. 20 subjects are randomly chosen out of 87 in part A of the dataset. In total, there are 2000 video sequences of about 5 seconds. As the pain label is on video sequence level, all the frames in all videos are checked for the existence of pain using the method in Section IV-A2. To determine if one video sequence is pain or not, we can set a threshold M. If there are M consecutive frames of detected pain, then that video sequence is classified as containing pain. Figure 10 shows an example when M=5. Table III shows the result of AU based method on the 20 subjects from Biovid with different values for M.

With the increase of M, the accuracy for baseline (neutral) increases while the accuracy for all pain levels decreases. To further investigate the details on these 20 subject, we plotted

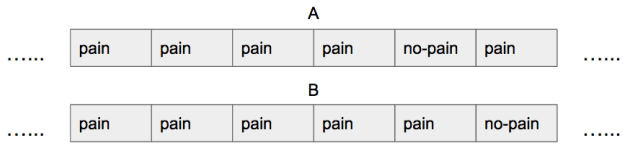

Fig. 10. An example for M=5, the video sequence A will not be classified as pain if it doesn't contain 5 consecutive pain frames, but video sequence B will be classified as pain

TABLE III
ACCURACY FOR M=1-5 AT DIFFERENT CLASSES. BL IS BASELINE, PA IS PAIN LEVEL. TAKING THE EXAMPLE OF PAIN LEVEL 4 WITH A 5 FRAME DETECTION WINDOW, THE TABLE READS AS FOLLOWS: 38,5 PERCENT OF THE CASES LABELED WITH PAIN=4 IN THE BIOVID DATASET WERE CLASSIFIED CORRECTLY AS PAIN, THE OTHER 61,5 WAS CLASSIFIED AS NO PAIN.

	BL	PA1	PA2	PA3	PA4
M=1	58.50%	41.25%	43.50%	47.00%	56.75%
M=2	64.50%	34.75%	38.25%	41.25%	53.25%
M=3	68.25%	31.50%	34.75%	38.50%	47.75%
M=4	70.75%	29.00%	30.75%	33.75%	43.50%
M=5	72.50%	26.25%	27.75%	31.50%	38.50%

a class accuracy (not shown) for all the subjects. From this we learned that performance is subject specific and for some subjects there is zero accuracy for 4 different pain levels and some have near zero accuracy for the baseline.

Further investigation into these 20 subjects video sequences showed an important finding in that most of these subjects have different behaviors than the patients in McMaster dataset. Among these 20 participants, many of them closed eyes for most of the time during the experiment with no facial movement, some even closed their eyes during the whole experiment (see the examples in Figure 11). However, in the McMaster dataset patients look at the camera during the whole experiment, only close their eyes when normal blink or pain. These different behaviors make it difficult to predict pain in Biovid with the model trained from McMaster, and again shows the importance of novel well-balanced datasets.

Fig. 11. Examples of participants in Biovid

V. A DEMONSTRATION ON A ROBOT FOR FROZEN SHOULDER THERAPY

To test how the best model works in real life, a demonstration with a humanoid robot is illustrated. The scenario is frozen shoulder therapy. The therapy is entirely conducted by the robot. The robot first shows the desired shoulder movements and asks the patient to follow it. Pain detection is activated while the patient follows the movement. If pain is detected from the patient's face during movement, the robot gives motivations by speech.

The robot used for this scenario is a Pepper(see Figure 12), a humanoid robot manufactured by SoftBank Robotics [28]. For this implementation, the forehead camera is used as image input.

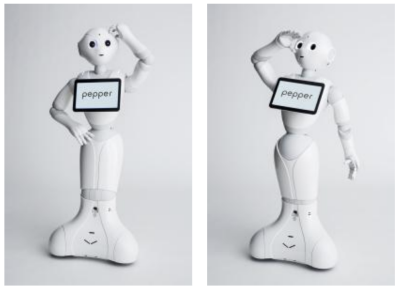

Fig. 12. The Pepper humanoid robot [28]

The overall workflow for the AU based pain detection with a robot is shown in Figure 13. During the therapy session, the camera on the robot will keep monitoring the patient's face. All the frames are sent back to a local computer, where OpenFace is utilized to detect AUs values for the face. Then, a pre-trained model uses those AUs values as input to predict if the corresponding face is in pain or not, and the robot decides whether to give motivation or not based on the prediction.

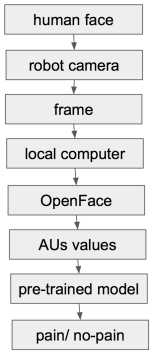

Fig. 13. The overall steps for pain detection in real time with Pepper robot

The streaming frames from the robot's camera are in the resolution of 640*480 with about 3-4 frames per second. This resolution is significantly lower than the images in the training datasets (AffectNet), and due to this specific scenario's setting there is a certain distance between the camera and patient's face. The prediction for pain could be highly sensitive to noise

data. Besides dropping the frames in which no front face is detected, a we used the majority vote over the last 6 frames to determine if pain is currently present. This method could help avoid misclassification from blurred images, sudden movement from the head, and noisy data cause the wrong prediction for 1 or 2 frames. Five different participants tried this pilot setup, participants reported a good posed-pain detection accuracy and a more enjoyable engagement in human-robot interaction compared to robot without pain detection. However, more testing and research is needed into this direction.

VI. Conclusions and future work

We investigated the effect of classification methods and datasets on the generalization of pain detection. We found important issues with current datasets that limit interpretation of earlier found model accuracy trained on these datasets. We also found that to achieve some level of real-time pain detection, variations of facial expressions had to be added to the pain/no-pain training data. We further found that the limited number of subjects present in McMaster and BioVid limits usability of models that are trained directly on raw data (in our case CNNs) as they quickly overtrain even when mixed with other facial expression data.

Overall our results suggest that earlier work that used only the McMaster dataset for training and testing with CNNs could be detecting subjects facial activity rather than pain.

The best approach we found is an AU based method trained on a combined dataset of AffectNet and McMaster. It achieves 88.4% accuracy. The real-time detection has been implemented on a humanoid robot as part of a frozen shoulder robot therapy pilot.

While we have used emotional faces as non-pain examples, it is quite possible that pain is associated with certain emotional expressions such as fear or distress [29]. This remains an interesting future topic to investigate.

Our work shows the importance of cross-checking with other datasets & real-time testing, as well as the need for a well balanced and ecologically valid pain dataset.

References

[1] S. Levy-Tzedek H. I. Krebs, L. Dipietro and et al. A paradigm shift for rehabilitation robotics. *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 4, pp. 61-70, 2008.

[2] R. A. Malik N. H. Zreik and C. P. Charalambous. Adhesive capsulitis of the shoulder and diabetes: a meta-analysis of prevalence. *Muscles Ligaments Tendons Journal*, 6, 26-34, 2016.

[3] Klein J. Balasubramanian S. and Burdet E. Robot-assisted rehabilitation of hand function. *Curr Opin Neurol*, 23(6):661-670, 2010.

[4] Shor P.C. Burgar C.G., Lum P.S. and Van der Loos M. Development of robots for rehabilitation therapy: the palo alto va/stanford experience. *J Rehabil Res Dev*, 37:663-673, 2000.

[5] Shor P.C. Majmundar M. Lum P.S., Burgar C.G. and Van der Loos M. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Arch. Phys Med Rehabil*, 83(7):952-959, 2002.

[6] Linton S.J. Crombez G. Boersma K. Leeuw M., Goossens M.E. and Vlaeyen J.W. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. *J Behav Med*, 30:77-94, 2007.

[7] Vlaeyen J. W. S. and Linton S. J. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3), 317-332, 2000.

[8] J. Jia A. Fidalgo A. Tajadura-Jimenez N. Kanakam N. Bianchi-Berthouze A. Singh, A. Klapper and A. Williams. Motivating people with chronic pain to do physical activity: Opportunities for technology design. *Proc. ACM Conf. Human Factors Comput. Syst.*, pp. 2803-2812, 2014.

[9] D. C. Turk and A. Okifuji. Psychological factors in chronic pain: Evolution and revolution. *J. Consult. Clin. Psychol.*, vol. 70, no. 3, pp. 678-690, 2002.

[10] A. C. Williams and K. D. Craig. Updating the definition of pain. *Pain*, vol.157 , Issue.11 , pp.2420, 2016.

[11] S. S. Jaywant and A. V. Pai. A comparison study of pain measurement scales in acute burn patients. *Indian J Occup. Ther.*, vol. 35, 2003, pp. 13-17, 2003.

[12] K. Schepelmann M. Kunz, V. Mylius and S. Lautenbacher. On the relationship between self-report and facial expression of pain. *Pain*, 5:368-76, 2004.

[13] P. Saha S. D. Roy, M. K. Bhowmik and A. K. Ghosh. An approach for automatic pain detection through facial expression. *Procedia Comput. Sci.*, vol. 84, pp. 99-106, 2016.

[14] A. Peiravi R. Kharghanian and F. Moradi. Pain detection from facial image using unsupervised feature learning approach. *in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 419-422, 2016.

[15] J. Gonzlez J.M. Gonfaus K. Nasrollahi P. Rodriguez, G. Cucurull and T. B. Moeslund. Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE TRANSACTIONS ON CYBERNETICS*, 2016.

[16] K.M. Prkachin P. Solomon P. Lucy, J.F. Cohn and I. Matthrews. Painful data: The unbc-mcmaster shoulder pain expression archive database. *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 2011.

[17] W. Friesen P. Ekman and J. Hager. Facial action coding system: Research nexus. *Salt Lake City, UT, USA: Network Research Information*, 2002.

[18] K. Prkachin and P. Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, vol. 139, pp. 267-274, 2008.

[19] The biovid heat pain database. *http://www.iikt.ovgu.de/BioVid.html*.

[20] S. Gruss H. Ehleiter J. Tan-H. C. Traue A. Al-Hamadi A. O. Andrade G. Moreira da Silva S. Walter, P. Werner and S. Crawcour. The biovid heat pain database: Data for the advancement andsystematic validation of an automated pain recognition system. *in IEEE Int'l Conf. on Cybernetics (CYBCONF)*, pages pp. 128–131, 2013.

[21] P. L. Carrier A. Courville-M. Mirza B. Hamner W. Cukierski Y. Tang D. Thaler I. J. Goodfellow, D. Erhan and D.H. Lee et al. Challenges in representation learning: A report on three machine learning contests. *International Conference on Neural Information Processing*, pp. 117-124, 2013.

[22] B. Hasani A. Mollahosseini and M.H. Mahoor. Affectnet: A new database for facial expression, valence, and arousal computation in the wild. *IEEE Transactions on Affective Computing*, 2008.

[23] Y. Lim T. Baltruaitis, A. Zadeh and L. Morency. Openface 2.0: Facial behavior analysis toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition*, 2018.

[24] M. Mahmoud T. Baltruaitis and P. Robinson. Cross-dataset learning and person-specific normalisation for automatic action unit detection. *Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition*, 2015.

[25] M. Valdenegro-Toro O. Arriaga and P. Plger. Real-time convolutional neural networks for emotion and gender classificationfacs - facial action coding system. *arXiv preprint arXiv:1710.07557*, 2017.

[26] S. Ren K. He, X. Zhang and J. Sun. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[27] Andrew G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR, abs/1704.04861*, 2017.

[28] Softbank robotics. *https://www.softbankrobotics.com/emea/en/pepper*.

[29] Temitayo A. Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min S. H. Aung, and Amanda C. De C. Williams. How can affect be detected and represented in technological support for physical rehabilitation. *ACM Trans. Comput.-Hum. Interact.*, 26(1):1–29, 2019.