2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

Convolution Neural Network for Pain Intensity Assessment from Facial Expression

Elahe Hosseini[1], Ruijie Fang[1], Ruoyu Zhang[1], Chen-Nee Chuah[1],
Mahdi Orooji[1], Soheil Rafatirad[2], Setareh Rafatirad[3], Houman Homayoun[1]

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871770

Abstract— **Pain is an unpleasant feeling that can reflect a patient's health situation. Since measuring pain is subjective, time-consuming, and needs continuous monitoring, automated pain intensity detection from facial expression holds great potential for smart healthcare applications. Convolutional Neural Networks (CNNs) are recently being used to identify features, map and model pain intensity from facial images, delivering great promise in helping practitioners detect disease. Limited research has been conducted to determine pain intensity levels across multiple classes. CNNs with simple learning schemes are limited in their ability to extract feature information from images. In order to develop a highly accurate pain intensity estimation system, this study proposes a Deep CNN (DCNN) model using the transfer learning technique, where a pre-trained DCNN model is adopted by replacing its dense upper layers, and the model is tuned using painful facial. We conducted experiments on the UNBC-McMaster shoulder pain archive database to estimate pain intensity in terms of seven-level thresholds using a given facial expression image. The experiments show our method achieves a promising improvement in terms of accuracy and performance to estimate pain intensity and outperform the-state-of-the-arts models.**

I. Introduction

Pain is a vital indicator of the health condition and the primary reason that prompts people to seek medical attention [1]. According to The International Association for the Study of Pain (IASP), pain is "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" [2]. A valid, and reliable pain assessment is vital for understanding the severity of the patient's situation, diagnosis, and treatments. Usually *self-reports* or *observational* scales are used to assess pain. However, there are some challenges in subjective pain assessments. They may cause high hospitals' staff workload because of continuous patients monitoring. Moreover, these subjective measurements may be unreliable since they are based on patients or caregivers' opinions. Furthermore, they are not always possible in some situations, for example infants, unconscious patients, or those with language impairments who can not communicate to express their pain intensity. Therefore, an accurate and reliable automatic pain measurement has an important potential diagnostic in population with limited linguistic ability and can provide an economical and unbiased option in hospital health monitoring.

Nonverbal expression such as facial expressions are valuable source of information. Craig *et al.* [3] indicate the facial expressions as a reflective and spontaneous reaction of painful experiences in humans. Therefore, many of automatic pain assessment research focus on analyzing facial expression. Studies investigating facial expressions of pain have most often used Facial Action Coding System (FACS) [4]. FACS is an objectively coding system that differentiates between 44 facial movements known as Action Units (AU). Each AU is determined by the onset, offset, and an intensity on a five-point scale. The facial expressions of pain are composed of a small subset of AUs using FACS. Prkachin *et al.* [5] developed the Prkachin and Solomon Pain Intensity (PSPI) metric based on this observation, which is a 16-level scale based on the contribution of individual intensity of pain related AUs.

Recently, deep learning algorithms have been demonstrated to be an important technique in the image classification field. Deep neural networks have been applied in pain recognition recently since they demonstrated greater effectiveness in both the feature selection and feature extraction such as the enhancement of pre-training methods and transfer learning algorithms [6], [7], [8]. In this paper, we continue this research by applying deep learning algorithms for pain detection using facial expression images.

In order to reduce development effort, this study applies inductive transfer learning and re-train a deep convolutional neural networok (DCNN) to estimate pain intensity. The transfer learning technique is a popular method for building models while saving time by using previously-learned patterns as the basis of the learning process [9], [10], [11]. By repurposing pre-trained models, we avoid training from sketch that requires a large amount of data and reduce computational efforts associated with training models. In other words, transfer learning makes advantage of previously learned knowledge by using pre-trained models that have been trained on a large benchmark dataset for a similar problem. The study's motivations are as follows: because the UNBC-McMaster database is a small dataset for training an accurate DCNN, and because a similar method can be applied to facial emotion recognition and pain detection systems, a previously trained model should be useful for classification by only fine-tuning the high-level features. In the proposed pain intensity estimation method, a pre-

[1]Elahe Hossein, Ruijie Fang, Ruoyu Zhang, Chen-Nee Chuah, Mahdi Orooji and Houman Homayoun are with the Department of Electrical and Computer Engineering, University of California, Davis, USA ehosseini@ucdavis.edu

[2]Soheil Rafatirad is with the Department of Computer Science, University of California, Los Angeles, USA

[3]Setareh Rafatirad is with the Department of Computer Science, University of California, Davis, USA

trained DCNN (e.g., ResNet-18), originally modeled for facial emotion recognition, is adopted by replacing its upper layers with the dense layer to make it compatible with pain detection. Next, with the painful face dataset, the model is fine-tuned using the pipeline strategy, where the dense layers are tuned first, followed by tuning of each DCNN block successively. Such fine-tuning gradually improves the accuracy of pain estimation to a high level without the need to develop and train a DCNN from scratch with random weights.

We designed an efficient and effective deep learning approach, based on transfer learning, to estimate the seven levels of pain from facial expression images. We accomplished this by fine-tuning a pretrained ResNet-18 and extracting the most important features. We introduced an image-pre-processing pipeline that includes cropping, centering, resizing, and data augmentation. Our pipeline achieves classification accuracy of 85%.

The rest of the paper is organized as follows. Section II provides a summary of the related works used in the automated pain detection from facial expressions. Section III describes an overview of the proposed methodology for pain assessment. Section II presents the experimental setup and the obtained results. Finally, Section V concludes the paper.

II. Related Works

In recent years, many methods have been developed for automatic pain assessment from facial expressions. The existing works can be divided into two categories based on the output of the method: binary pain classification (pain/no pain) and estimating pain intensity. Early studies tend to focus on distinguishing a painful face from a non-painful one. In [12], [13], Active Appearance Model (AAM) based features are feed into the Support Vector Machine (SVM) classifiers for frame-level pain recognition. They performed pain detection in frame-level and sequence-level. For sequence-level prediction, they aggregated frame level predictions by averaging. Gholami *et al.* [14] applied a correlation vector machine (RVM), a Bayesian extension of SVM to distinguish pain from no pain in neonates as well as assess their pain intensity levels. Hammal *et al.* [15] identified four level of pain intensity using Log-Normal filter based features and SVM classifier. Kaltwang *et al.* [16] compared three approaches by using facial landmarks, DCT, and Local Binary Patterns (LBP) features to train three Relevance Vector Regression (RVR) models separately to estimate PSPI. The best performance was achieved by training an RVR model to combine the predictions of all three trained RVR models.

More recently, deep learning methods have also been applied to pain estimation directly from images or a sequence of frames. Zhou *et al*. [17] exploited AAM to warp all facial images of different poses into the frontal pose and converted each facial images into a 1D vector using flattening feature vectors. Then the flatten features are fed into the Recurrent Convolutional Neural Network (RCNN) to estimate pain intensity. Bellantonio1 *at al*. [18] first extracted self-learned features of each frame via the fully connected layer of a CNN architecture, then fed the extracted features to a Long-Short Term Memory (LSTM) to obtain the temporal information. Xin *et al*. [19] proposed an end-to-end attention network with spatial transformation network (STN) followed by an attention mechanism and CNN to estimate 4 levels pain intensity.

Rezaei *et al*. [20] proposed a neural network architecture that compares two images from the same person and estimates the difference in the pain level in them based on facial expressions of older adults with dementia. Rudovic *et al*. [21] presented a Personalized Federated Deep Learning (PFDL) approach for pain estimation from face images by using a lightweight CNN architecture across different clients without sharing their face images. Instead of sharing all parameters of the model, PFDL retains the last layer locally.

Some studies exploited the transfer learning technique by applying the customized VGG-Face pre-trained with millions of faces, and then used a hybrid joint deep learning approach to estimate pain intensity [6], [7], [8]. Rodriguez *at al.*[6] used VGG-Faces to learn facial features and linked it to a Long-Short Term Memory (LSTM) to exploit the temporal relation between video frames. Haque *et al.* [7] employed depth and Thermal along with RGB images for pain analysis. They present 'Multimodal Intensity Pain (MIntPAIN)' database. They recognized 5 levels of pain by analyzing independent visual modalities and fused with VGG-Face and LSTM. [8] presented an enhanced joint hybrid CNN-BiLSTM (EJH-CNN-BiLSTM) algorithm by linking a VGG-face to a joint Bidirectional Long Short Term Memory (BiLSTM) to estimate 4 levels pain intensity.

III. Proposed Method

As pain detection is a form of facial expression recognition, similar methods can be applied to the more general task of emotion recognition. Therefore, in this paper we propose a transfer learning based DCNN framework to estimate pain intensity utilizing knowledge from facial emotion data. The block diagram of the proposed framework is shown in Fig. 1. Basically, it is divided into three primary steps that aim to improve the overall efficacy of the algorithm. In the first step, the DCNN is trained on a facial emotion recognition dataset (FER+) to obtain the pre-trained model for the transfer learning task. Subsequently, a preprocessing step applied procedures such as cropping, resizing, centralizing, and data augmentation techniques to the original images from the UNBC-McMaster pain database. The pre-processing provides clean images to feed into the model for the training step. Finally, the pre-trained DCNN is fine-tuned in order to obtain seven distinct pain intensity levels. The details of the proposed framework are explained in the following subsection.

A. Data Processing

In a real-world environment, there is a lot of inherent heterogeneity in an image dataset. For instance, a face may

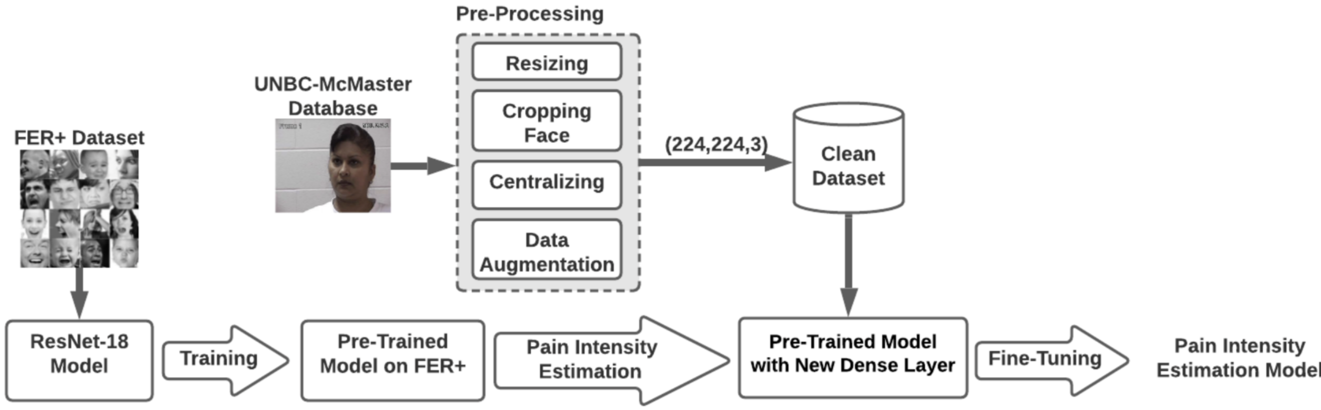

Fig. 1. Framework of the proposed pain intensity estimation system based on transfer learning and ResNet-18.

be captured in different orientations and locations, scales, and brightness. Therefore, by using face detection and alignment in the Dlib toolbox, we extend the face bounding box with a ratio of 25% to crop patients' faces from raw images. The pre-processing stage improves the identification of the images during any experimental phase. For this purpose, the raw images are applied in the pre-processing part of our model and each original image in the database is cropped, and then centralized. We have therefore resized the images to $224 \times 224 \times 3$ pixels because this representation is the most common input size for most of the deep neural network models after cropping.

B. CNN and Transfer Learning

The residual neural network (ResNet) [22] introduced the skip-connection concept, which is used in most of the later models. Skip connections work by taking input from one layer and adding it to the output from another layer after several layers. As a result, the layer contains more information and the vanishing gradient problem is resolved. ResNet models exist with different depths, including ResNet-18, 34, 50, and 152, the numbers indicate the number of convolutional layers, and the depth size is one less than the mentioned number.

Since the network has many parameters to be optimized, training any large DCNN model is a complex and time-consuming task. A deep network usually demands a large training dataset. Training with small or inadequate amounts of data can lead to overfitting. For some tasks, it is difficult to obtain a sufficient amount of data for proper training of the DCNN. However, research has shown that transfer learning [11], [10] can be very useful to solve this issue. Transfer learning is a concept to use the knowledge representation learned from different tasks but similar applications. It is reported that the transfer learning technique works better when both tasks are similar [11], [9] which is the motivation behind this study.

In this study, to extract the features prevalent on facial images, a feature extraction technique denoted as the ResNet method was utilized. It consists of a fine-tuned ResNet18 CNN as a pre-trainer. The use of a pre-trained CNN algorithm as a facial feature extractor is expected to be useful as a basis for training the CNN algorithm to accurately detect pain from the facial images.

C. Pain intensity estimation using transfer learning in DCNN

According to the CNN layers visualization [23], the first layers of the CNN capture basic features like the edge and corners of an image. As layers go deeper, more complex features are detected. Since the basic features from facial images are similar for both facial emotion recognition and pain estimation tasks, the lower layers in CNN can be identical. Training a CNN model from scratch with randomly initialized weights is a time-consuming task and a pre-trained model on another task can be fine-tuned employing the transfer learning approach for pain estimation. A ResNet-18 model trained with a large emotion recognition dataset (FER+ [24]) is suitable for pain intensity estimation. The following subsections describe transfer learning concepts for pain estimation and the proposed method in detail with required illustrations.

The convolutional layers are a part of pre-trained ResNet-18 excluding its own classifier, and the classifier is the newly added layers for pain intensity estimation. Generally, using a pretrained CNN model in transfer learning consists of two steps: 1) replacing the pre-trained original classifier with a new one, and 2) fine-tuning the model. The new classifier is a combined dense layer(s) of those that are fully connected. In particular, three strategies are used for fine-tuning the model for the target task: 1) training the entire model, 2) training some layers and leaving other layers frozen, and 3) training the classifier (added dense layers) and freezing the convolution layers. Since in this scenario, facial emotion recognition and pain estimation tasks are similar, training only the classifier layers is enough. Therefore, fine-tuning is performed on the added classifier (new dense layer).

Fig. 1 illustrates the proposed automatic pain intensity estimation with ResNet-18, the well-known pre-trained CNN model. The available ResNet-18 model is trained with the FER+ dataset for emotion recognition. The pre-trained model is modified for pain intensity estimation by redefining the dense layers, and then fine-tuning is performed with a painful face dataset. In defining the architecture, the last dense layers of the pre-trained model are replaced with the new dense layer(s) to recognize a facial image into one of seven levels of pain as defined in Table I. The dense layer is the regular fully connected neural network layer, which means each neuron in the dense layer receives input from all neurons of its previous layer. It takes some dimension as input and produces an output with the required dimension. Therefore, the output layer contains only seven neurons. A cleaned pain dataset prepared through preprocessing (described in Section III-A) is used to train in fine-tuning. In the case of testing, a facial image is placed to the input of the system, and the highest output probability is considered to be the decision.

Fig. 2 shows the detailed schematic architecture of the pre-trained ResNet-18 model and dense layer for pain intensity estimation. It s made up of small *residual blocks* and the purpose for the direct connection with the input to the output is to give a direct path for the error to backpropagate through the entire network. In DCNN, skip connection can avoid the vanishing gradient. ResNet-18 performed convolution operation on the input, followed by 4 residual blocks, and finally performed full connection operation. The full model in the same pipeline of pre-trained DCNN and the added dense layer give the opportunity for fine-tuning the dense layer.

Stochastic gradient descent algorithm, the popular optimization algorithm in computer vision and natural language processing applications, is used in the re-training process. Image cropping and data augmentation are considered for training our model. Cropped face portion from the image is considered as input in the model to enhance facial properties. On the other hand, data augmentation is an effective technique for creating new data from available data. In this case, we utilize flipping and rotating techniques to generate new face images from the original image. This is very helpful, especially when the dataset is not very large, like the case of UNBC-McMaster database.

IV. Experimental Results

A. Dataset and Implementation Details

The UNBC-McMaster Shoulder Pain Expression Archive Database [25] is the most commonly used database for visual analysis of pain assessment. It consists of 200 videos showing facial expressions of 25 participants suffering from shoulder pain, who underwent a series of range-of-motion tests with their affected and unaffected limbs. The dataset features rich annotation with self-report and observer measures of the pain intensity at video level and FACS coding at frame level. The PSPI score was computed to quantify pain intensity in 16 discrete levels (0-15) based on AUs. In this paper, we used the videos of active tests to perform pain intensity estimation experiments, with the 16-level PSPI as the ground truth. The database provides 200 sequences across 25 subjects, which totals 48,399 images. Fig. 3 shows some of the images indicated by PSPI.

The database used is unbalanced and hence, it has been very challenging to perform the modelling experiments. As can be seen from Fig. 4, examples of test subjects experiencing no pain significantly outnumber those experiencing pain. Therefore, based on the specific character of the database it is likely that any model gets biased towards the prediction of no-pain at the cost of missing pain frames. Using imbalance data is basically intentionally biasing data to get an interesting result. To deal with this issue, we upsample the positive examples by applying the data augmentation technique by flipping and rotating images. Moreover, full sequences included only no pain (PSPI = 0) frames were removed. Then, We downsample the negative examples by random selection 20 percent of them. For classifying pain into seven levels, the database was divided into seven parts as shown in Table I.

We use the SGD method for optimization with a momentum of 0.8 and a weight decay of 10^{-4}. We initialize the learning rate (lr) to $4e$, and modify it to 8e-7 at 60 epochs and 1.6e-7 at 120 epochs, and stop training after 180 epochs.

TABLE I
DIVIDED LEVELS OF PAIN IN THE DATABASE FOR SEVEN LEVELS BASED ON PSPI CODES OF IMAGES' FRAMES.

Pain level	**PSPI score**	**Augmentation**
0	0	Downsampling
1	1-2	-
2	3-4	Flipping and Rotating
3	5-6	Flipping and Rotating
4	7-9	Flipping and Rotating
5	10-12	Flipping and Rotating
6	13-15	Flipping and Rotating

B. Performance evaluation metrics

To evaluate the quality of our model predictions, several performance evaluations measures were used. The performance was quantified via the Pearson correlation coefficient (PCC) between the predictions and the human-annotated ground-truth. Moreover, classification accuracy, Area under Curve (AUC), and F1 score were reported for the performance of our model.

$$PCC = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}}_i)^2 (y_i - \bar{y}_i)^2}} \tag{1}$$

where N is the total number of frames of testing sequences. y_i and $\hat{y}_i$ are the ground-truth and the pain intensity estimation of the i frame, respectively. $\bar{\hat{y}}_i$ and $\bar{y}_i$ are the sample mean of $\{y_1, ..., y_N\}$ and $\{\hat{y_1}, ..., \hat{y_N}\}$.

True positive rate (TPR) or Recall corresponds to the probability of an actual positive that are correctly considered as positive. False positive rate (FPR) corresponds to the probability of an actual negative that are mistakenly considered as positive. FPR and TPR both have values in the range [0, 1]. Accuracy is the number of correct predictions out of the total number of predictions. The ROC curve is plotted with TPR vs. FPR where TPR is on the y-axis and FPR is on

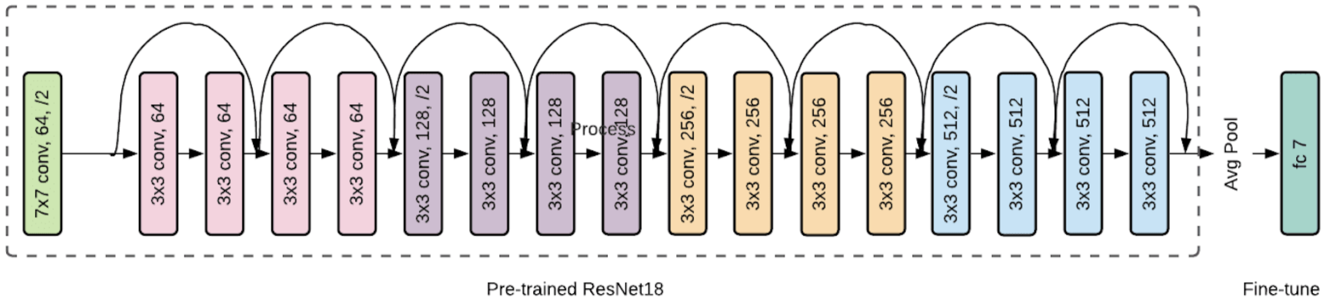

Fig. 2. Architecture of ResNet-18, the model is trained on FER+ dateset and the last layer are replaced and weights are tuned.

Fig. 3. The image frame samples of the UNBC-McMaster Shoulder Pain Achieve database [25] used in this paper.

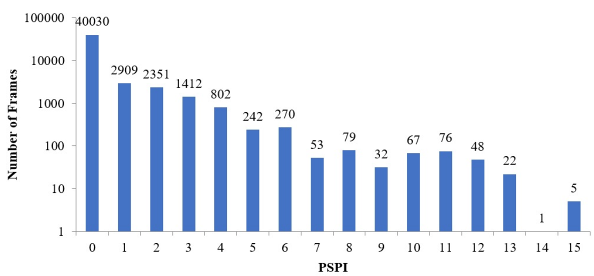

Fig. 4. Number of frames at each pain level (PSPI score) in the UNBC-McMaster dataset. Note that the y-axis is in log scale.

TABLE II

THE AVERAGE PERFORMANCE OF THE ALGORITHM PER EACH CLASS.

Pain level	Recall(%)	F1-score(%)	Precision(%)
0	88.34	78.75	71.03
1	75.72	78.01	80.46
2	77.57	70.26	64.21
3	87.69	82.49	77.88
4	78.41	77.04	75.71
5	73.13	81.27	91.45
6	81.30	89.69	100

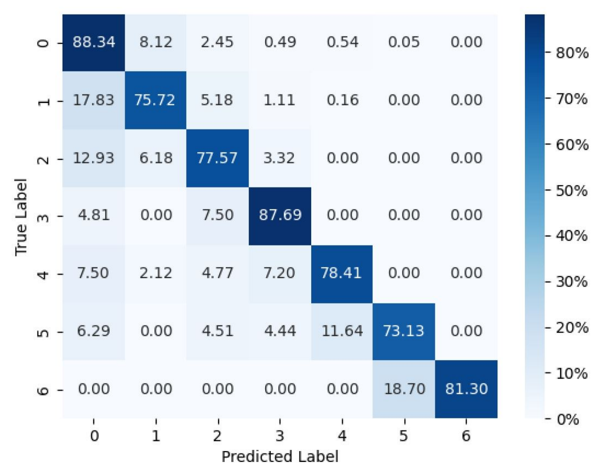

Fig. 5. The confusion matrix corresponding for the predicted pain intensity level using our model.

the x-axis. The Area Under the ROC Curve (AUC) is the measure of the ability of a classifier to distinguish between classes. The F1-score is a combination of the precision and recall of the model, and this score takes both false positives and false negatives into account.

C. Results

The proposed pain intensity estimation system was evaluated for each class based on recall, F1-score, and precision. Table II shows the average performance measuring of the proposed framework for each pain level. As for our proposed method, we used transfer learning based DCNN model to conduct pain intensity estimation. We got promising results of the average accuracy and PCC of 85.71% and 0.83, respectively. It indicates that our method is effective.

The proposed framework is compared with state-of-the-art methods in this section. The comparison results of the proposed framework with other methods' results on the UNBC-McMaster database are shown in Table III. The proposed framework shows promising performance and effectiveness compared to the state-of-the-art methods. As shown in Table III, the proposed framework outperforms the existing methods in pain assessment in terms of accuracy and PCC. Our model achieves superior Accuracy and PCC compared to CNN-LSTM [6] but a lower AUC, but in our work we estimate pain intensity in seven levels while [6] estimates pain/no-pain faces. Furthermore, the high PCC of this proposed framework indicates that it is capable of extracting detailed information from the face structure in order to estimate pain intensity.

TABLE III

COMPARISON OF THE PROPOSED FRAMEWORK RESULTS WITH THE STATE-OF-THE-ART RESULTS IN THE UNBC-MCMASTER SHOULDER PAIN DATABASE

	Classifier	#Level	AUC	Acc	PCC
Lucey *et al.* [13]	SVM	2	0.84	-	-
Kaltwang *et al.* [16]	RVR	2	-	-	0.60
Rodriguez *et al.* [6]	CNN-LSTM	2	0.93	0.83	0.78
Rezaei *et al.* [20]	CNN	4	0.86	-	0.71
Xin *et al.* [19]	STN	4	-	0.51	-
Bargshady *et al.* [8]	CNN-BiLSTM	4	0.88	0.85	-
Zhou*et al.* [17]	RCNN	16	-	-	0.65
Our model	**DCNN**	**7**	**0.88**	**0.85**	**0.83**

V. CONCLUSION

A growing area of research in health informatics focuses on appropriate pain management strategies and the automatic detection of pain intensity using facial images. This paper introduces a framework for automatic pain intensity estimation from facial images using deep convolutional neural networks based on the transfer learning technique. Since the facial emotion recognition and pain intensity estimation tasks are similar, we trained ResNet-18 on a FER+ dataset to achieve satisfactory results in the pain intensity estimation domain. The fully connected layer of a pre-trained ResNet-18 was optimized for the pain estimation task by fine-tuning the weights of the last fully connected layer of the model. In this work, we accelerate the training time of the DCNN for pain intensity estimation by using transfer learning. The our model is evaluated on the UNBC-McMaster Shoulder Pain Expression Archive Database. Experimental results demonstrated that our DCNN significantly improves the performance achieved by using the conventional approaches like SVM, RVR, and CNN. The enhanced algorithm obtained an AUC of 88%, an accuracy of 85%, and PCC of 0.83. The DCNN framework developed in this study can have useful implications for the medical diagnostic fields, particularly, for clinicians and other medical researchers who wish to implement automatic pain management practices.

REFERENCES

[1] P. Mäntyselkä, E. Kumpusalo, R. Ahonen, A. Kumpusalo, J. Kauhanen, H. Viinamäki, P. Halonen, and J. Takala, "Pain as a reason to visit the doctor: a study in finnish primary health care," *Pain*, vol. 89, no. 2-3, pp. 175–180, 2001.

[2] H. Merskey, "The definition of pain," *European psychiatry*, vol. 6, no. 4, pp. 153–159, 1991.

[3] K. D. Craig, K. M. Prkachin, and R. E. Grunau, "The facial expression of pain." 2011.

[4] P. Ekman and W. V. Friesen, *Facial action coding system: Investigator's guide*. Consulting Psychologists Press, 1978.

[5] K. M. Prkachin and P. E. Solomon, "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain*, vol. 139, no. 2, pp. 267–274, 2008.

[6] P. Rodriguez, G. Cucurull, J. Gonzàlez, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. Roca, "Deep pain: Exploiting long short-term memory networks for facial expression classification," *IEEE transactions on cybernetics*, 2017.

[7] M. A. Haque, R. B. Bautista, F. Noroozi, K. Kulkarni, C. B. Laursen, R. Irani, M. Bellantonio, S. Escalera, G. Anbarjafari, K. Nasrollahi *et al.*, "Deep multimodal pain recognition: a database and comparison of spatio-temporal visual modalities," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 250–257.

[8] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Systems with Applications*, vol. 149, p. 113305, 2020.

[9] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[10] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *arXiv preprint arXiv:1411.1792*, 2014.

[11] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717–1724.

[12] A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, "The painful face–pain expression recognition using active appearance models," *Image and vision computing*, vol. 27, no. 12, pp. 1788–1796, 2009.

[13] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, "Automatically detecting pain in video through facial action units," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 3, pp. 664–674, 2010.

[14] B. Gholami, W. M. Haddad, and A. R. Tannenbaum, "Relevance vector machine learning for neonate pain intensity assessment using digital imaging," *IEEE Transactions on biomedical engineering*, vol. 57, no. 6, pp. 1457–1466, 2010.

[15] Z. Hammal and J. F. Cohn, "Automatic detection of pain intensity," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, 2012, pp. 47–52.

[16] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous pain intensity estimation from facial expressions," in *International Symposium on Visual Computing*. Springer, 2012, pp. 368–377.

[17] J. Zhou, X. Hong, F. Su, and G. Zhao, "Recurrent convolutional neural network regression for continuous pain intensity estimation in video," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 84–92.

[18] M. Bellantonio, M. A. Haque, P. Rodriguez, K. Nasrollahi, T. Telve, S. Escalera, J. Gonzalez, T. B. Moeslund, P. Rasti, and G. Anbarjafari, "Spatio-temporal pain recognition in cnn-based super-resolved facial images," in *Video Analytics. Face and Facial Expression Recognition and Audience Measurement*. Springer, 2016, pp. 151–162.

[19] X. Xin, X. Lin, S. Yang, and X. Zheng, "Pain intensity estimation based on a spatial transformation and attention cnn," *Plos one*, vol. 15, no. 8, p. e0232412, 2020.

[20] S. Rezaei, A. Moturu, S. Zhao, K. M. Prkachin, T. Hadjistavropoulos, and B. Taati, "Unobtrusive pain monitoring in older adults with dementia using pairwise and contrastive training," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1450–1462, 2020.

[21] O. Rudovic, N. Tobis, S. Kaltwang, B. Schuller, D. Rueckert, J. F. Cohn, and R. W. Picard, "Personalized federated deep learning for pain estimation from face images," *arXiv preprint arXiv:2101.04800*, 2021.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 233–255, 2016.

[24] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 279–283.

[25] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 57–64.