

Received July 2, 2021, accepted July 24, 2021, date of publication July 30, 2021, date of current version August 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101396

# 2D Facial Expression and Movement of Motion for Pain Identification With Deep Learning Methods

KORNPROM PIKULKAEW<sup>1</sup>, WARAPORN BOONCHIENG<sup>2</sup>,  
EKKARAT BOONCHIENG<sup>3,4</sup>, (Member, IEEE), AND VARIN CHOUVATUT<sup>4</sup>

<sup>1</sup>Graduate Program in Computer Science, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>2</sup>Faculty of Public Health, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>3</sup>Data Science Research Center, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>4</sup>Department of Computer Science, Faculty of Science, Graduate School, Chiang Mai University, Chiang Mai 50200, Thailand

Corresponding authors: Varin Chouvatut (varinchouv@gmail.com) and Ekkarat Boonchieng (ekkarat.boonchieng@gmail.com)

This work was supported in part by the Thailand Research Fund, and in part by Chiang Mai University, Thailand, through the Royal Golden Jubilee Ph.D. Program under Grant PHD/0226/2561.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Local Ethics Committee at the Faculty of Public Health, Chiang Mai University.

**ABSTRACT** As pain is an inevitable part of life, this study examines the use of facial expression technology in assisting individuals with pain. The self-reporting system commonly used to detect discomfort is ineffective and cannot be utilized by patients of all ages; a standardized formula for measuring pain would resolve this issue. Facial monitoring technology is an important tool for measuring pain because it is both easy to use and incredibly precise. Accordingly, this article uses deep learning techniques to examine the use of 2D facial expressions and motion to sense pain. Sequential pictures from the University of Northern British Columbia (UNBC) dataset were used to train a deep learning model, as deep learning can detect motion and assist patients in self-reporting. Our mechanism is capable of classifying pain into three categories: not painful, becoming painful, and painful. The system's performance was evaluated by comparing its findings to those of a specialist physician. The precision rates of the not painful, becoming painful, and painful classifications were 99.75 percent, 92.93 percent, and 95.15 percent, respectively. In sum, our study has developed an alternative way to test for pain prior to hospitalization that is straightforward, cost effective, and easily understood by both the general population and healthcare professionals. Additionally, this analysis technique could be applied to other screening methods, such as pain detection for infectious diseases.

**INDEX TERMS** Assessment system, deep learning, facial expression, face recognition, pain detection.

## I. INTRODUCTION

Pain is something that everyone experiences, sometimes at excruciating levels. As a result, a diagnosis of pain requires payment to a doctor. Traditionally, many patients and healthcare workers measure levels of discomfort using the self-rating approach [1] to describe the degree of pain. However, this method has drawbacks, primarily inaccuracy. Patients' and medical staff's perceptions of pain can vary widely, especially in children [2]. After the self-report approach was introduced, the observer rating scale (ORS) was developed to assist individuals in improving the accuracy

of their self-reporting [3]. In certain instances, however, pain control according to the ORS requires the skills of a specialist. Such cases frequently occur in intensive care units (ICU) and when doctors and nurses are unable to monitor patients for an extended period of time. Both self-reporting and ORS are applicable in many cases, but people can experience pain from a wide range of causes, from stomach cramps and shoulder injuries to sore throats. Simply describing a patient's condition is therefore inadequate for accurate treatment, despite the additional assessments of pain that may be available. A wide variety of techniques can now be used to assess pain, including neurologic diagnostic and imaging technologies, but such procedures are expensive for patients. Accordingly, Vijayanandh and Balakrishnan [4] proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai.

facial recognition as a potential solution, specifically using a facial action coding scheme (FACS), which identifies facial expression.

The FACS visualization system uses 44 action units (AUs) to show muscle activity in the face. A significant amount of research has been conducted on this approach to help people discern human emotional states, such as sadness, anger, impartiality, and calm [6]–[8]. The Prkachin and Solomon Pain Intensity (PSPI) metric was proposed by Lucey *et al.* [9], Prkachin and Solomon [10], who were the first to bring FACS into the medical profession. PSPI can quantify pain using FACS, which typically classifies pain using facial muscles and codes them using the corresponding AUs. Each AU is scored on a scale of 0 to 5, except for AU43, which has two values (0, 1). The facial expression data in this study has been extracted from the UNBC, a widespread pain recognition dataset. However, pain levels are inconsistent across the UNBC.

One of the most critical aspects of deep learning, first offered by Goodfellow *et al.*, is generative adversarial networks (GANs) [11]. Many analysts have used GANs to complete a variety of tasks, such as creating iris models or a modified national institute of standards and technology (MNIST). Since GANs are composed of two deep neural networks, they perform the same task as a neural network; in essence, they are also a discriminator that attempts to match a fictitious picture against a genuine image created by the generator. This process is repeated until the separator is unable to differentiate between the initial and generated images. As a result, the GAN can address issues of class balance and accuracy improvement. Pikulkaew *et al.* [12] applied the Wasserstein generative adversarial network (WGAN) to human faces and discovered that it may improve the efficiency of the pain discovery technique. They also used the WGAN to improve pain detection techniques, which could increase efficiency due to the limited data in the relevant datasets [13].

This research aims to offer a system for 2D facial pain detection and classification that utilizes both deep learning techniques and motion, specifically a motion technique that allows individuals to maintain a high degree of orientation when shifting from side to side. This comes from the system's ability to predict the axis of the decoded picture based on the coordinates from a reference image. Our primary contributions are as follows:

- A strategy for detecting pain that relies on deep learning for classification and recognition. This system is more capable of reaching state-of-the-art performance than other classification methods, such as support vector machines (SVMs)
- Classification of pain into three categories: not painful, becoming painful, and painful.

We implemented a three-level scale because we wanted to create a method that can be used as a standard tool for pain measurement in daily life. In the past, certain researchers were only capable of performing either painless

or painful procedures, but the proposed method's evaluation system explores pain detection with both experimental and ground truth (doctor analysis). The results of the investigation matched the medical personnel's ground truth data.

We believe that the techniques presented here offer significant contributions to the fields of computer vision and image processing, among others, especially in terms pain-sensing methods. These techniques can also be used, with appropriate modifications, to address other detection issues. For example, a system that can help people with everyday life could include primary pain screenings for issues such as abdominal pain, shoulder pain, or COVID-19.

The rest of this article is structured as follows: the relevant works are proposed in Part II; our materials and methods are explained in Part III; the studies and findings are given in Part IV and discussed in Part V; and, finally, our accomplishments and subsequent works are summarized in Part VI.

## II. RELATED WORK

The FACS, proposed by Ekman *et al.* [15], is a standard facial movement estimation tool used by researchers in many fields for facial expression identification. FACS defines facial expressions by encoding facial muscle movements as fourth-fourth AUs. While it seems that it can both distinguish pain and identify its potency, some researchers [8] have observed a correlation between individual emotional states, such as sadness and surprise, making it harder to use the AUs to accurately detect pain. In-depth studies of the mood-detection process can be found in the work of Zeng *et al.* [16] and, most recently, Cohn and Torre [17]. These studies are generally tied to three database characteristics: (1) the classification of a pain expression (COPE) dataset [6] in terms of an infant experiencing pain; (2) the Biovid heat pain dataset [18], which archives pain instances; and (3) the UNBC pain dataset [9] of adult patients with shoulder pain.

Conversely, Gholami *et al.* [7] used a relevant vector machine (RVM) for binary pain detection on the face of a manually selected child and Guo and Zhang [19] suggested a local binary pattern (LBP) and extension to enhance facial imaging and accuracy. During experiments on the Biovid dataset, Werner *et al.* [18] obtained information from various outlets and then used a head posture estimator to assist in pain management. They showed that specific individuals had different pain thresholds. Littlewort *et al.* [20] discussed the perceptions of facial expressions in those who are experiencing pain by using a previously established AU detector implemented with Gabor, AdaBoost, and SVM filters. Their analysis focused on AUs and was developed using the UNBC pain database. Lucey *et al.* [5] used active appearance models (AAMs) to manually track and adjust faces on a labeled keyframe and enter them as a device classification SVM. If any pain-related AUs previously identified by Prkachin *et al.* were reported, the frame was described as indicating pain [10]. Methods explored in Kaltwang *et al.* [21] and Rudovic *et al.* [22] estimated

numerous types of pain. In comparison, Kaltwang *et al.* used LBP, discrete cosine transformations (DCT), and AAM together to determine pain intensity either by AU or by directly processing all device frames while Rudovic *et al.* suggested a more time-specific conditional random field (CRF).

Prkachin and Solomon [10] demonstrated that four activities—lowered eyebrows, eyelid tightening or cheek elevation, nostril or upper lip lift, and eye closure—are consistent enough to be considered significant expressions of pain. Various studies have reported that other actions can also be related to pain. LeResche *et al.* [23], for instance, stated that horizontal stretching of the lip caused by the risible muscle correlated with suffering. Similarly, Craig *et al.* [24] indicated that lifting the lip, which includes the zygomaticus major muscle (which is also activated when smiling), increased as the pain level increased. Given the well-reviewed recommendations for self-measuring pain based on the FACS, Prkachin and Solomon [10] suggested the PSPI. In their study, pain expression was thoroughly defined by the activation of a limited set of facial muscles and coded by a community of related AUs: AU4, AU6-AU7, AU9-AU10, and AU43. The PSPI metric is presented in equations 1 and 2:

$$\text{PSPI} = \text{AU4} + \max(\text{AU6} - \text{AU7}) + \max(\text{AU9} - \text{AU10}) + \text{AU43}, \quad (1)$$

$$\text{Pain} = \text{intensity}(\text{AU4}) + (\max \text{intensity AU6, AU7}) + (\max \text{intensity AU9, AU10}) + \text{intensity}(\text{AU43}) \quad (2)$$

Except for AU43, which has two values (0, 1), each operation is rated on a six-point scale (0 = imperceptible, 5 = maximum) using FACS. PSPI measurements yield 16 levels.

Significant advancements were made in 2001, when Viola and Jones [25] designed a hair-based cascade classifier for entity detection, and in 2002, when Lienhart and Maydt enhanced it [25]. The target detection method is both quick and accurate, though performs better from the front and is limited when placed to the side of the face. In 2020, Bargshady *et al.* [30] proposed a new method for facial expression deep learning to detect pain in four phases, and found that the enhanced joint hybrid - convolutional neural network - bidirectional long short-term memory (EJH-CNN-BiLSTM) system that they proposed substantially outperformed the traditional method in terms of efficiency. Additionally, Bargshady *et al.* [35] developed an ensemble deep learning model - CNN - recurrent neural network (EDLM-CNN-RNN) that is capable of correctly classifying pain and generating multi-class pain levels, offering several possible applications in the field of medical informatics.

Hussein *et al.* [36] suggested using a CNN model to comprehend three facial emotions: neutral, negative, and positive emotions. This model, inspired by Xception, uses residual blocks and depth-separable convolutions; it achieved

an accuracy rating of 81% for unseen outcomes while neutral emotion recognition has an accuracy rating of only 51%. In the same year, Egede *et al.* [37] developed the EmoPain dataset, a novel platform for assessing chronic pain using multimodal facial and body expressions. Dufourq [38] conducted a study on facial expression recognition using CNN and found that CNN enables researchers to reach a higher level of classification accuracy than the conventional techniques. With sufficient effort, future CNNs will be capable of achieving state-of-the-art performance while also requiring less computing power. Li and Xu [39] hypothesized that the emotion-based categorization of facial expression recognition is heavily reliant on the quality of the available data. As a result, they proposed a new framework for pre-selecting relevant pictures based on reinforcement learning.

Ravi and Yadhukrishna [40] presented a balanced comparison of two of the most commonly utilized facial expression recognition (FER) methods, LBP and CNN, noting that the lighting conditions in the images affected accuracy. Their findings showed that CNN outperformed LBP. In the area of animation, Paier *et al.* [41] pioneered the use of a deep neural network (DNN). The method's central idea is to enable animators to easily manipulate the facial expressions of a virtual human character through the use of deep neural networks. This data-driven method allows for the generation of physically accurate facial animations and representations. Ameer *et al.* [42] presented a method for increasing the detection rate of faces using monogenic binary patterns, (MBPs) and CNN, as well as DCNN, which is one of the most effective methods for enhancing large-scale picture recognition.

However, as shown in the summary of these techniques presented in Table 1, none of the existing methods consider 2D facial pain detection and motion. An important advantage of this approach is that it decreases the amount of time that hospital staff requires to monitor patients, including impaired people, children, and ICU cases, or to, for example, conduct a primary screening for COVID-19. Moreover, this method can reduce costs by decreasing the need for expensive tissue analysis.

### III. MATERIALS AND METHODS

#### A. DATA COLLECTION

This study started with data collection to determine the extent of the difficulties involved in pain detection. Since we focused on human subjects, we used the shoulder pain dataset (UNBC), which included validation experiments, for our work. This photograph collection includes over 49,000 images of adults of all genders. Under painful conditions, sequential images of people and objects are taken with an effective image resolution of  $320 \times 240$  pixels. The subject turns their shoulder independently to perform dynamic movement, though the operator may assist the patient with passive movement.

**TABLE 1.** Compendium of the current strategies for facial expression recognition.

Author	Methods	A	B	C	D
Ashraf et al. [26]	AAM	PSPI	SVM	UNBC	82%
Lucey et al. [5]	AAM, 3D Head Pose Changes	PSPI	SVM	UNBC	84%
Brahnam et al. [6]	DCT	PSPI	SVM	COPE	85.51%
Gholami et al. [7]	Pixels	PSPI	SVM, RVM	COPE	82%
Hammal and Cohn [27]	Canonical Appearance	PSPI	SVM	UNBC	82%
Kaltwang et al. [21]	DCT, LBP	PSPI	RVR	UNBC	MSE = 1.39, PCC = 0.59
Hong et al. [28]	LBP+ Variants	PSPI	NN	UNBC	88.2%
Rathee et al. [29]	Distance Metric Learning, Thin Plate Spline	PSPI	SVM	UNBC	DML = 96% NO DML = 79%
Florea et al. [34]	Histograms of Topographical features	PSPI	SVM	UNBC	MSE = 1.18 PCC = 0.55

A = Measures of pain intensity, B = Classifier, C = Dataset, D = Precision, DML: Distance Metric Learning, MSE: Mean Square Error, NN: Nearest Neighbor, PCC: Pearson Correlation Coefficient, RVR: Relevance Vector Regression.

Additionally, analysts concentrate on the front-facing images during the dynamic state, while video recording takes place approximately 70 degrees from front during the passive state. Examples are shown in Figure 1. Due to the benefits of this technique in real-life situations, including basic discomfort like a stomachache or menstruation, we only



**FIGURE 1.** Pain level classifications using the PSPI scale.

examine our proposed technique in the context of human facial images.

**B. DATA AUGMENTATION**

It is important to ensure data remains almost equal, thereby enhancing the overall dataset and handling any potential errors. This approach is known as the imbalance technique [33]. Instead of modifying all of the pictures in the collection, the data augmentation procedure will choose one image, apply simple geometric transformations like



FIGURE 2. An example of AAM landmarks to locate AUs.

TABLE 2. Data augmentation parameter.

Augmentation Type	Parameter
Rotation	30°
Scaling	15
Shearing	15°
Width Shifting	2
Height Shifting	2

translation and rotation, and then proceed to the next step. By using this method, images in the training set that have been changed are treated as belonging to the same class as the original image. When we enhance data, our goal is to increase the generalizability of our models. We must utilize one of the imbalanced algorithms because the UNBC dataset is deficient in some classifications, meaning that the model may show bias when some classes are less common than others. Thus, before proceeding with the pain detection procedure, we must utilize data augmentation to expand the dataset; the parameter we used is shown in Table 2.

**C. PAIN ASSESSMENT SCALES**

Both careful evaluations and the formal diagnosis of a patient’s condition are required for adequate pain control. As disease may have individual side effects, pain assessment approaches typically rely on a patient’s understanding of

TABLE 3. The overview of pain classification.

Level/Type	ORS	VAS	AMS	SS
Painless	0-1	0-3	0-4	0-4
Becoming Painful	2-3	4-7	5-9	5-9
Painful	4-5	8-10	10-16	10-16

TABLE 4. An example of a DCNN-based algorithm for facial expression recognition.

```

procedure facial_recognition

    for all image (i), face expressions representation ∈ UNBC dataset

        do
            data that has been cropped, scaled, and normalized
            reformat the picture (i) as grayscale
            identify (i) and crop just the face (x)
            using DCNN, extract ResNet features from a cropped face (x)

        for (epoch = 0, epoch = 5, epoch++)
            batch ← 8
            learning rate ← 1e-1
            train, test data
        end for

        convert the image (i) to RGB

    end procedure
    
```

TABLE 5. A DCNN-based algorithm for pain detection is shown below.

```

procedure pain_detection

    preprocess (input)
    cropping, scaling, and normalization of data

    compute the measure for pain intensity
    calculated motion vector of images
    DCNN modeling to identify pain
    enhancing the DCNN dataset via data augmentation
    for (epoch = 0, epoch = 50, epoch++)
        batch ← 128
        learning rate ← 1e-1
        train, test DCNN
    end for

    identify three distinct categories of pain severity
    evaluation process

end procedure
    
```

pain and degree of severity thereof. Typical self-reporting approaches therefore suffer from pitfalls such as reactivity to feedback, duplicity, and differentiation between a patient’s sense of discomfort and the clinician’s assessment, leading to the addition of an ORS to address these concerns. In situations where the examination takes a long time, such as when supervising individuals in clinics, the ORS’s main disadvantage is its poor reliability and validity. We can categorize initial pain in three ways: (1) by computing pain using the PSPI equation; (2) by using ORS, the Visual Analog Scale (VAS), the Affective Motivational Scale (AMS), and the Sensory Scale (SS) to determine the source of the initial pain; and (3) by instructing a deep convolutional neural network (DCNN) to locate a point that changes based on pain.

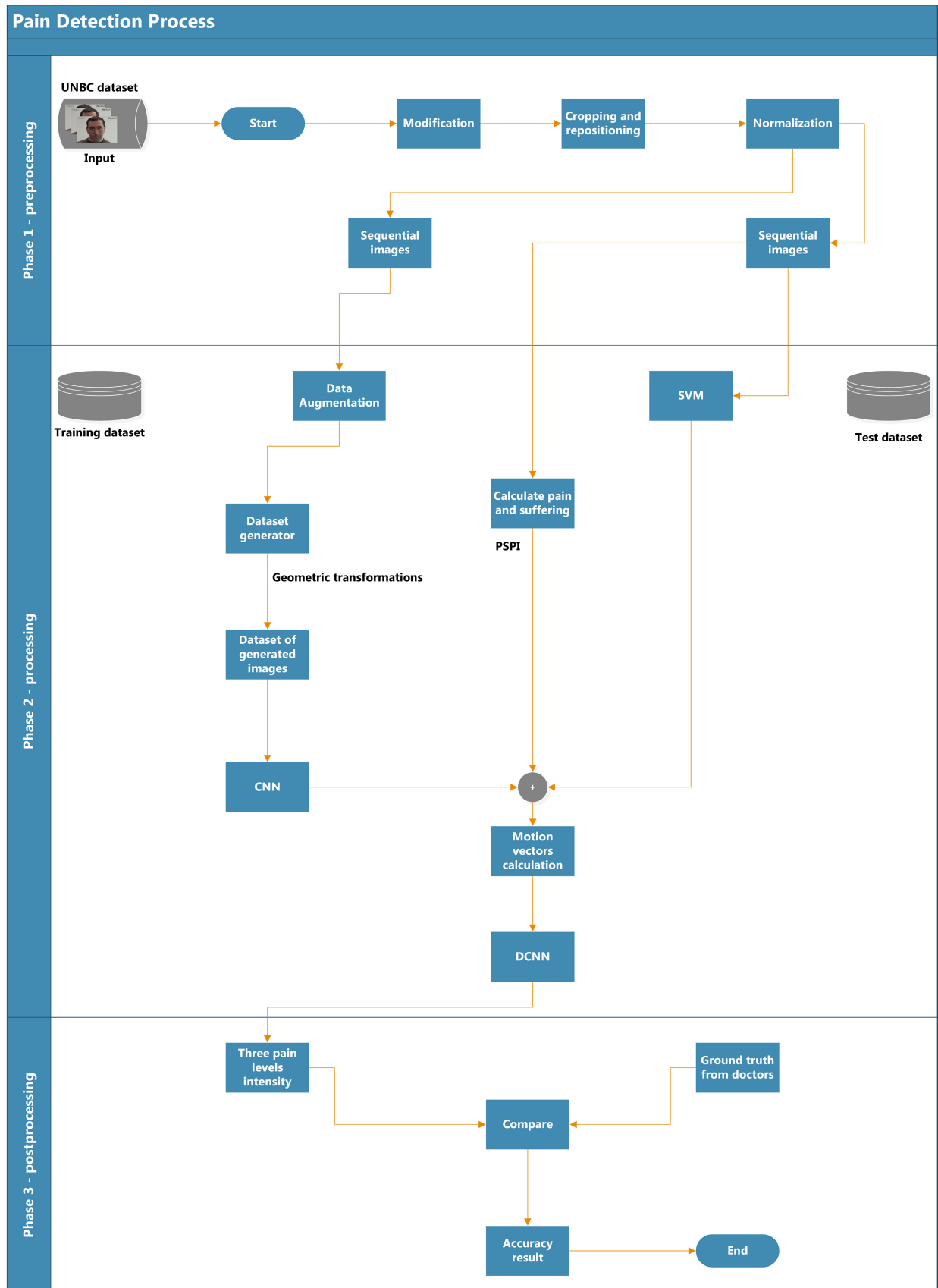


FIGURE 3. The suggested method's flowchart.

**TABLE 6.** Our findings regarding the perception of pain in human faces.

Samples	Level	A	B	C	D	E
	No pain	0	4	6	7	8
	Initial pain	7	4	6	7	8
	No pain	0	3	3	3	4
	Pain	10	5	10	14	14
	Initial pain	5	4	9	14	13

Table 3 summarizes how to grade pain. The pain severity measures we used are described below:

- The VAS is an adult version of a one-dimensional pain scale. On this measure, patients indicate their current level of suffering on a scale ranging from 0 to 10, with 0 representing no pain and 10 representing the most severe pain.

- The ORS is a multidimensional instrument that is related to past or recent performance and practice. Medical personnel will measure the patient’s distress on a simple 0 to 5 scale, with 0 representing no discomfort and 5 representing the highest amount of pain [3].
- The AMS is a defined distinction measure that expresses emotional-control inclinations. The scale runs from

TABLE 6. (Continued.) Our findings regarding the perception of pain in human faces.

	Initial pain	8	5	4	4	8
	Pain	10	5	8	8	10
	No pain	0	3	3	4	6
	Initial pain	5	3	4	3	4
	No pain	0	3	6	7	10

A = Prkachin and Solomon Pain Intensity Metric (PSPI), B = Observer Rating Scale (ORS), C = Visual Analog Scale (VAS), D = Affective Motivational Scale (AMS), E = Sensory Scale (SS).

0 to 16, showing increasing discomfort from the initial pain-free value of 0 to the final value of 16, the highest level of pain.

- The SS is capable of monitoring and scoring sensory environment reactions and tactile input responses

in ways that can be used by patients in everyday circumstances. Patients indicate their actual degree of discomfort using a scale ranging from 0 to 16, with 0 representing no pain and 16 representing the most pain.



**TABLE 7.** The variables used to calculate the evaluation scores.

Variable	Meaning
Sufferer	Positive for the illness.
Healthy	Negative for the illness.
True negatives (TN)	The proportion of patients properly classified as healthy.
False negative (FN)	The number of patients misclassified as healthy.
True positives (TP)	The proportion of patients properly classified as ill.
False positives (FP)	The number of patients misclassified as ill.

#### D. FACIAL DETECTION PROCESS

We first used OpenCV to improve facial recognition, but it is important to note that OpenCV is not designed to recognize faces. We attempted to plot an AAM to find AUs in every patient image before using the DCNN, demonstrating the primary role used to analyze pain, as shown in Figure 2. Our face recognition architecture is built on ResNet-34, a deep residual learning method developed by He *et al.* [14] for image recognition. Over three million pictures of tagged faces from the wild dataset were used to train the network. The training process then collected sequential photos from the UNBC dataset to create the marker. The reason for drawing labels is to boost and increase the accuracy of the system's speed. We then used the movement to predict the axis in the decoded picture based on the coordinates in a reference image. A deep learning model can learn by itself while maintaining precision, though it has limited power. For example, it needs to process a significant amount of data to obtain good results. Time is an additional aspect to consider, as this approach takes much longer than others, including the SVM.

#### E. DEEP LEARNING PROCESS

Convolutional neural networks (CNN) are neural networks that are primarily used to categorize pictures, cluster images based on their similarity, and recognize objects within scenes; CNN are the most frequently used computational method for image identification. LeNet [32] was one of the first convolutional neural networks to be developed and was crucial in the development of deep learning. We can refer to it as the world's first deep neural network, as it was built in 1998 to address the problem of digit recognition. Following that, ResNet is regarded as a particularly unique architecture because it is a true deeper design, with 152 levels. It gained popularity because it bypasses the degradation problem, which occurs when the depth of the network is raised beyond a certain threshold in both the training and testing sets. The previous layer's output, known as the residual, is added to the current layer's output in ResNet. Our technology for pain detection is based on CNN ResNet-34, with the number 34 indicating that the CNN has 34 layers.

A deep convolutional neural network (DCNN) is composed of many layers of neural networks. Usually, two distinct layer types, convolutional and pooling, are used alternately. Each filter in the network has a depth that rises from left to right. Typically, the final level is composed of one or more fully connected layers. The layer count is the distinction between CNN and DCNN; a deep learning model with a larger number of layers is said to be deeper. Thus, DCNN is just a CNN with more layers.

#### F. PARTICULARS OF APPLICATION

Python 3.5.1 was used to establish pain-sensing via deep learning and motion algorithms on Windows 7 (64 bits) with OpenCV, and the UNBC dataset was used to train the pain-sensing system. To prepare for research, we used 70% of the sequential images for training and the rest for testing. Another key factor is the processing time required for the Intel i7 processor with DDR4 16 GB RAM; identifying both pain classes takes about three months. Our system flowcharts are shown in Figure 3 and our algorithm for facial expression is displayed in Table 4.

### IV. EXPERIMENT AND RESULTS

#### A. PAIN INTENSITY METRIC

Including an alarm device with audible effects for emergency situations involving ICU patients or disabled individuals was essential for this research, as we wanted to make it possible for doctors and nurses to apply our software in real life. The results are summarized in Table 6. Six AUs—AU4, AU6, AU7, AU9, AU10, and AU43—controlled the intensity of PSPI pain. The PSPI scale can be given using equations (1) and (2). Other assessments, such as the VAS, ORS, AMS, and SS, were often used to rate people as well. The data was collected from medical professionals and others by a physician/specialist chosen by the National Center for Community Health Informatics. Table 5 illustrates our method for pain identification.

#### B. EVALUATION PROCEDURES

Determining whether or not a model is performing adequately in terms of belief and validity is, of course, essential, and often the only way to do so is via empirical data. Four possible outcomes may occur while conducting categorization predictions. Table 7 contains a list of the factors that we utilized in this study.

Precision, accuracy, and recall are the three essential characteristics used to evaluate a model. Table 8 summarizes the variables used in the evaluation measurement.

#### C. PROCESS OF VALIDATION

Our model was validated against classification criteria as well as a variety of batch sizes and epochs. Additionally, we assessed the proposed method's performance both with and without data augmentation. The findings from the tests are summarized in Table 9 while Table 10 compares the proposed model to the unbalanced methodology in terms of

**TABLE 8.** The variables, definitions, and equations used in the evaluation process.

Variable	Definition	Equation
Accuracy	The percentage of correctly predicted data from tests. It is simply calculated by dividing the total number of accurate predictions by the total number of predictions.	$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$ (3)
Precision	The percentage of excellent instances among all expected examples belonging to a particular class.	$Precision = \frac{TP}{(TP+FP)}$ (4)
Recall	The percentage of instances that were intended to belong to a class divided by the total number of occurrences.	$Recall = \frac{TP}{(TP+FN)}$ (5)
F1-Score	The term is used to refer to the accuracy of a test. The highest possible F1-score is 1, indicating excellent precision and recall, while the minimum possible value is zero.	$F1 - Score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$ (6)

**TABLE 9.** The results of our pain detection algorithm.

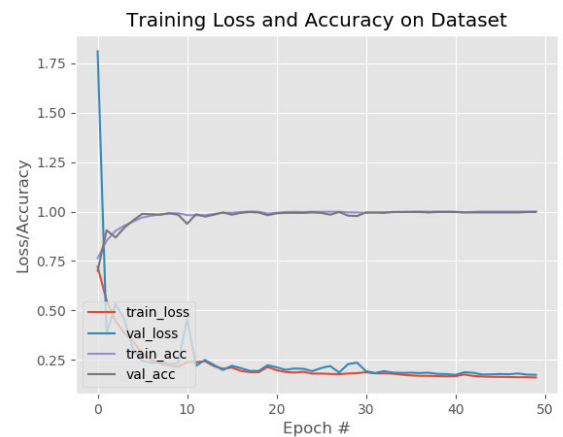
Epoch	Batch Size	F1-score	Recall	Precision
5	8	0.97	0.97	97%
	32	0.98	0.98	98%
	128	0.89	0.89	91%
20	8	0.99	0.99	99%
	32	0.98	0.98	98%
	128	0.99	0.99	99%
50	8	0.98	0.99	98%
	32	0.98	0.98	98%
	128	0.99	0.99	99%

**TABLE 10.** The outcome of our pain detecting algorithm without the use of augmented data.

Epoch	Batch Size	F1-score	Recall	Precision
5	8	0.94	0.97	92%
	32	0.99	1.00	99%
	128	0.70	0.67	89%
20	8	0.98	0.99	97%
	32	0.99	1.00	99%
	128	0.99	1.00	99%
50	8	0.97	0.99	95%
	32	0.99	0.99	98%
	128	0.99	1.00	99%

precision. The graph in Figure 4 illustrates the relationship between training failure and the accuracy of our model.

Table 9 shows how our 2D facial expression and motion algorithm augments data. Ninety-nine percent of the results



**FIGURE 4.** Accuracy of pain detection via data augmentation.

**TABLE 11.** Correspondence with diagnoses made by a doctor.

Tests	Doctor analysis	Correctness
1. Painless	Painless	98.63%
2. Beginning to be painful	Beginning to be painful	89.54%
3. Painless	Painless	99.42%
4. Painful	Painful	95.15%
5. Painless	Painless	99.75%
6. Beginning to be painful	Beginning to be painful	92.93%
7. Beginning to be painful	Beginning to be painful	88.43%
8. Painful	Painful	82.17%
9. Painless	Painless	98.75%
10. Beginning to be painful	Beginning to be painful	89.54%

were correct; accuracy was limited when working with small batch sizes but was excellent when working with large batch sizes. Additionally, our objectives could be accomplished within a short time period.

**V. DISCUSSION**

Table 11 shows a comparison of experimental and ground truth (doctor analysis) data in terms of patients' pain in three modules: no pain, initial pain, and pain. The initial pain and pain classes influenced the imbalanced data from the UNBC dataset, showing that these two classes

**TABLE 12.** Comparison of suggested approaches to state-of-the-art pain detection strategies.

Citation	Classifier	Dataset	A	B	C
Lucey et al., 2012	SVM	UNBC	48,398	2	-
Rodriguez et al., 2017	CNN-LSTM	UNBC	48,000	2	83.1%
Zhou et al., 2016	RCNN	UNBC	16,657	2	-
Bargshady et al., 2020	CNN-BiLSTM	UNBC	10,783	4	85%
Bargshady et al., 2020	EDLM-CNN-RNN	UNBC	10,783	5	86%
Bargshady et al., 2020	EDLM-CNN-RNN	MIntPAIN	34,800	5	92%
Haque et al., 2018 [31]	CNN-LSTM	MIntPAIN	34,800	5	32.4%
Bellantonio et al., 2016	CNN-RNN	UNBC	-	3	61.9%
Hammal & Cohn, 2012 [27]	SVM	UNBC	16,157	4	80%
Our proposed	DCNN-SVM	UNBC	10,595	3	93.4%

A = Data size, B = Pain level, C = Accuracy, LSTM: Long Short-Term Memory, RCNN: Recurrent Convolutional Neural Network, RNN: Recurrent Neural Network.

sometimes demonstrated similar patterns for the two values of accuracy. Nevertheless, it was important to explain both the benefits and drawbacks of deep learning, which require more information to process, and so we decided to demonstrate that some experiments could drop below 90% accuracy, though every experiment actually achieved above 90% accuracy in practice.

For example, Experiment 5 in the no-pain category had the highest accuracy rating, 99.75%, with a low of 98.63%. The initial pain category accounted for 92.93% of this group's maximum value, while the accuracy rating in terms of classification of the type of pain reached in 95.15% in Experiment 4.

Overall, the most significant decrease was found in the pain category. In Experiment 7, the accuracy of the initial pain hit a low of 88.43%, while pain accuracy plummeted to 82.17% in Experiment 8. However, all experimental results matched the results of the doctor's analysis, demonstrating that this experiment was reliable. Reliability could be further improved in the future by incorporating sequential images or using GAN or data augmentation to increase the deep learning model's accuracy and limits. Table 12 summarizes the obtained findings and compares them to the state-of-the-art output.

## VI. CONCLUSION

This article proposed a novel strategy for detecting pain using deep learning that can be used in both patients' everyday lives and in hospitals. Individuals with shoulder discomfort were used as a model to validate the technique. The proposed approach works effectively in conditions

that cannot be handled using standard methods. We also proposed a method that can increase the efficiency and accuracy of pain perception by applying deep convolutional neural networks to monitor conditions and other complex environments.

We plan to add accompanying photos to one of our potential studies, as the levels of pain severity between the initial pain and pain classes are currently unbalanced. Accordingly, the GAN, imbalance technique, and data augmentation will be applied to increase the accuracy of the process [31]. Additionally, the deep learning methods presented here, mainly 2D facial expression and movement, were superior to conventional methods (doctor analysis) in terms of both accuracy and cost. This approach could be applied to either other fields or to further medical research in the future; potential examples include facial recognition for criminal identification and pain detection in terms of contagious diseases.

## REFERENCES

- [1] C. L. Hicks, C. L. von Baeyer, P. A. Spafford, I. van Korlaar, and B. Goodenough, "The faces pain scale—Revised: Toward a common metric in pediatric pain measurement," *Pain*, vol. 93, no. 2, pp. 173–183, Aug. 2001.
- [2] O. Doody and M. E. Bailey, "Understanding pain physiology and its application to person with intellectual disability," *J. Intellectual Disabilities*, vol. 23, no. 1, pp. 5–18, May 2017.
- [3] H. J. Moller, "Observer rating scales," in *Guide to Assessment Scales in Major Depressive Disorder*. Cham, Switzerland: Springer, 2014.
- [4] R. Vijayanandh and G. Balakrishnan, "Human face detection using color spaces and region property measures," in *Proc. 11th Int. Conf. Control Autom. Robot. Vis.*, Singapore, Dec. 2010, pp. 1605–1610.
- [5] 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 *Proc. Face Gesture*, Santa Barbara, CA, USA, Mar. 2011, pp. 197–205.
- [6] S. Brahmam, L. Nanni, and R. Sexton, "Introduction to neonatal facial pain detection using advanced face classification techniques," in *Advanced Computational Intelligence Paradigms in Healthcare*. Berlin, Germany: Springer, 2007.
- [7] B. Gholami, W. M. Haddad, and A. R. Tannenbaum, "Relevance vector machine learning for neonate pain intensity assessment using digital imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1457–1466, Jun. 2010.
- [8] J. Huang, X. Xu, and T. Zhang, "Emotion classification using deep neural networks and emotional patches," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Kansas City, MO, USA, Nov. 2017, pp. 958–962.
- [9] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, S. Chew, and I. Matthews, "Painful monitoring: Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database," *Image Vis. Comput.*, vol. 30, no. 3, pp. 197–205, Mar. 2012.
- [10] 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, Oct. 2008.
- [11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2014, pp. 2672–2680.
- [12] K. Pikulkaew, E. Boonchieng, W. Boonchieng, and V. Chouvatut, "Network-simulated generation of human faces with expressions and orientations for deep learning classification," *J. Innov. Technol. Exploring Eng.*, vol. 9, no. 2, pp. 2178–2185, Dec. 2019.
- [13] K. Pikulkaew, E. Boonchieng, W. Boonchieng, and V. Chouvatut, "Pain detection using deep learning with evaluation system," in *Proc. 5th Int. Congr. Inf. Commun. Technol.*, London, U.K., 2020, pp. 426–435.

- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
- [15] P. Ekman, W. V. Friesen, and S. Ancoli, "Facial signs of emotional experience," *J. Pers. Social Psychol.*, vol. 39, no. 6, pp. 1125–1134, 1980.
- [16] Z. Zeng, Y. Hu, and T. S. Huang, "Audio-visual spontaneous emotion recognition, in artificial intelligence for human computing," in *Artificial Intelligence for Human Computing*, Berlin, Germany: Springer, 2007.
- [17] J. F. Cohn and F. De la Torre, "Automated face analysis for affective computing," in *The Oxford Handbook of Affective Computing*. New York, NY, USA: Oxford, 2015.
- [18] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue, "Automatic pain recognition from video and biomedical signals," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 4582–4587.
- [19] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
- [20] J. M. Susskind, G. Littlewort, M. S. Bartlett, J. Movellan, and A. K. Anderson, "Human and computer recognition of facial expressions of emotion," *Neuropsychologia*, vol. 45, no. 1, pp. 152–162, 2007.
- [21] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous pain intensity estimation from facial expressions," in *Proc. Int. Symp. Vis. Comput.*, Crete, Greece, 2012, pp. 368–377.
- [22] O. Rudovic, J. Lee, and R. W. Picard, "Personalized machine learning for robot perception of affect and engagement in autism therapy," *Science*, vol. 3, no. 19, pp. 1–11, Jun. 2018.
- [23] L. LeResche, S. F. Dworkin, L. Wilson, and K. J. Ehrlich, "Effect of temporomandibular disorder pain duration on facial expressions and verbal report of pain," *Pain*, vol. 51, no. 3, pp. 289–295, Dec. 1992.
- [24] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: An exploratory look into the role of affect in learning with autotutor," *J. Educ. Media*, vol. 29, no. 3, pp. 241–250, Oct. 2004.
- [25] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, Rochester, NY, USA, 2002, pp. 900–903.
- [26] A. Ashraf, P. 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 Vis. Comput.*, vol. 27, no. 12, pp. 1788–1796, Nov. 2009.
- [27] Z. Hammal and J. F. Cohn, "Automatic detection of pain intensity," in *Proc. Int. Conf. Multimodal Interact.*, New York, NY, USA, 2012, pp. 47–52.
- [28] X. Hong, G. Zhao, S. Zafeiriou, M. Pantic, and M. Pietikäinen, "Capturing correlations of local features for image representation," *Neurocomputing*, vol. 184, pp. 99–106, Apr. 2016.
- [29] N. Rathee and D. Ganotra, "A novel approach for pain intensity detection based on facial feature deformations," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 247–254, Nov. 2015.
- [30] 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 Syst. Appl.*, vol. 149, no. 1, pp. 1–10, Jul. 2020.
- [31] M. A. Haque, R. B. Bautista, F. Noroozi, K. Kulkarni, C. B. Laursen, R. Irani, M. Bellantonio, S. Escalera, G. Anbarjafari, K. Nasrollahi, O. K. Andersen, E. G. Spaich, and T. B. Moeslund, "Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Xi'an, China, May 2018, pp. 250–257.
- [32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [33] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop*, Swinoujscie, Poland, May 2018, pp. 117–122.
- [34] C. Florea, L. Florea, and C. Vertan, "Learning pain from emotion: Transferred HoT data representation for pain intensity estimation," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland, 2015, pp. 778–790.
- [35] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Ensemble neural network approach detecting pain intensity from facial expressions," *Artif. Intell. Med.*, vol. 109, no. 1, pp. 1–12, Sep. 2020.
- [36] E. S. Hussein, U. Qidwai, and M. Al-Meer, "Emotional stability detection using convolutional neural networks," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol.*, Doha, Qatar, Feb. 2020, pp. 136–140.
- [37] J. O. Egede, S. Song, T. A. Olugbade, C. Wang, A. C. D. C. Williams, H. Meng, M. Aung, N. D. Lane, M. Valstar, and N. Bianchi-Berthouze, "EMOPAIN challenge 2020: Multimodal pain evaluation from facial and bodily expressions," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Buenos Aires, Argentina, Nov. 2020, pp. 849–856.
- [38] E. Dufourq, "A survey on factors affecting facial expression recognition based on convolutional neural networks," in *Proc. Conf. South Afr. Inst. Comput. Sci. Inf. Technol.*, Cape Town, South Africa, Sep. 2020, pp. 168–179.
- [39] H. Li and H. Xu, "Deep reinforcement learning for robust emotional classification in facial expression recognition," *Knowl.-Based Syst.*, vol. 204, no. 1, pp. 1–9, Sep. 2020.
- [40] R. Ravi, S. V. Yadhukrishna, and R. Prithviraj, "A face expression recognition using CNN & LBP," in *Proc. Int. Conf. Comput. Methodol. Commun.*, Erode, India, Mar. 2020, pp. 684–689.
- [41] W. Paier, A. Hilsmann, and P. Eisert, "Interactive facial animation with deep neural networks," *IET Comput. Vis.*, vol. 14, no. 6, pp. 359–369, Sep. 2020.
- [42] B. Ameur, M. Belahcene, S. Masmoudi, and A. B. Hamida, "Unconstrained face verification based on monogenic binary pattern and convolutional neural network," in *Proc. Int. Conf. Adv. Technol. Signal Image Process.*, Sousse, Tunisia, Sep. 2020, pp. 1–5.



**KORNPROM PIKULKAEW** received the B.E. degree in computer engineering (Hons.) and the M.E. degree in computer engineering (international program) from King Mongkut's University of Technology Thonburi. He is currently pursuing the Ph.D. degree in computer science with Chiang Mai University. His research interests include computer vision, computer graphics, human-computer interfaces, 3-D games, animation, and data science.



**WARAPORN BOONCHIENG** received the B.S. and M.S. degrees in health promotion from Chiang Mai University, and the Ph.D. degree in health education and behavioral science from Mahidol University. She is currently an Associate Professor and the Head of the Public Health School, Chiang Mai University.



**EKKARAT BOONCHIENG** (Member, IEEE) received the B.S. degree from Khon Kaen University, the M.S. degree from the University of New Haven, and the Ph.D. degree from Illinois Institute of Technology, all in computer science. He is currently an Associate Professor of computer science with Chiang Mai University and the Director of the Center of Excellence in Community Health Informatics, CMU. His current research interests include networks, image processing, and data science.



**VARIN CHOUVATUT** received the B.E., M.E., and Ph.D. degrees in computer engineering from King Mongkut's University of Technology Thonburi (KMUTT), Thailand. She is currently an Assistant Professor of computer science with Chiang Mai University. Her current research interests include image processing, computer vision, computer graphics, and machine vision.

• • •