STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, May 2025, pp 0–14.
Published online in International Academic Press (www.IAPress.org)

Pain Intensity Recognition from Facial Expression Using Deep Learning

Nusrat Tania, Md. Mijanur Rahman*, A.H.M Saifullah Sadi*, Wahidur Rahman*

Department of Computer Science and Engineering, Uttara University, Bangladesh

Abstract Pain weaves its way into daily life, turning ordinary tasks into challenges and testing our strength in unexpected ways. Pain can be observed in a human's face and can be understood as pain intensity from patients' verbal acquisition. However, for non-verbal patients—such as those in the ICU, individuals with mental challenges, or even AI—detecting and interpreting pain remains a complex challenge. To extend this, many researchers have done remarkable research and are still trying to find a solution with acceptable accuracy. This research paper presents a hybrid parallel model to detect the intensity of pain from facial expressions. Following the Prkachin and Solomon Pain Intensity (PSPI) metric, we considered 16 pain levels, which were divided into four subranges. We call these sub ranges as " No Pain", "Mild Pain", "Moderate Pain", and "Severe Pain". Our parallel feature fusion model consists of a fully connected network with inputs from two deep CNN models, one being VGG19, and the other model can be ResNet50, DenseNet121, or InceptionV3. Thus, we have 3 parallel feature fusion models (PFFM), respectively, PFFM-1, PFFM-2 and PFFM-3. Besides, we trained and evaluated our models using the McMaster Shoulder Pain dataset, where PFFM-2 emerged as the top performer, achieving an 82.54% accuracy in assessing pain intensity from facial expressions. By outperforming existing pain detection systems, this breakthrough bridges the gap between human perception and AI, enabling more precise and reliable pain interpretation.

Keywords Facial Expression Recognition, Pain Intensity, Transfer Learning, Deep Learning, Healthcare.

DOI: 10.19139/soic-2310-5070-2664

1. Introduction

The pain is widely recognized as one of the most distressing sensations experienced by humans, often necessitating medical attention. Broadly, it can be categorized into two major types encountered in daily life: acute pain and chronic pain. Acute pain typically arises from injuries or illnesses and tends to subside with appropriate treatment. In contrast, chronic pain results from severe or untreated injuries, infections, or ongoing conditions such as cancer. Unlike acute pain, chronic pain is persistent and may linger for months or even years, despite ongoing medical interventions aimed at minimizing discomfort [1]. Moreover, pain exerts a profound impact not only on the affected individual but also on their family members. It can cause restlessness, alter personality traits, impair memory, disrupt attention, and diminish intellectual performance. The influence of pain extends to sleep quality, often leading to insomnia and, over time, contributing to depression [1]. Consequently, both the intensity and duration of pain are strongly linked to a decline in an individual's Quality of Life (QOL). The ability to adapt to pain plays a vital role in maintaining a satisfactory QOL, underscoring the importance of accurately measuring pain intensity. Such assessments are critical for identifying the root cause of pain and prescribing the most effective treatment strategies [2]. In this context, facial expressions emerge as a primary medium for conveying emotions, especially in social settings. According to American psychologist Paul Ekman (1993), there are seven basic human emotions—happiness, sadness, fear, disgust, anger, contempt, and surprise—which can be interpreted based on subtle facial muscle movements, eye gestures, and head positions. These emotions were systematically analyzed

*Correspondence to: Wahidur Rahman (Email: mwrahman@uttarauniversity.edu.bd); Md. Mijanur Rahman (Email: mijanur@manarat.ac.bd); A.H.M Saifullah Sadi (Email: sadi.swe@diu.edu.bd). Department of Software Engineering, Daffodil International University, Dhaka-1216, Bangladesh.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

through his Facial Action Coding System (FACS) [3]". Notably, facial expressions can serve as valuable indicators of a person's pain experience in real-life situations, providing non-verbal yet informative cues [4]. While humans can often intuitively interpret facial expressions, the task poses significant challenges for even the most advanced Artificial Intelligence (AI) systems [5]. Despite these challenges, the role of AI in healthcare is expanding rapidly. Various organizations are partnering with researchers and medical professionals to develop technologies aimed at enhancing pain recognition and management. These efforts are grounded in the shared belief that relieving pain is not only a medical necessity but also a fundamental human right [6]. In line with this vision, our research proposes a novel, integrated approach for identifying pain intensity through facial expression analysis. This automated pain detection system presents an intelligent solution that supports patients and caregivers in effectively monitoring and managing chronic pain. By combining emotional recognition with AI-driven assessment, this system contributes to the broader goal of improving patient care and overall quality of life [7]. Building upon the significance of accurately identifying pain through facial expressions, it is essential to consider the existing pain assessment methods typically employed in medical settings. Among these, self-reporting and clinical observation are the most common. Notably, self-reporting—utilizing either verbal responses or visual analog scales—is widely recognized as the most reliable method. However, this approach becomes impractical or unreliable for specific populations, such as patients in intensive care units, young children, infants, individuals with paralysis, or those with cognitive or neurological impairments, including dementia and autism [8],[6],[9]. Furthermore, individuals living with chronic pain often experience fluctuating pain levels due to coexisting medical conditions. As a result, they may overlook new symptoms, perceiving them as part of their usual discomfort. In such cases, any unexpected change in pain intensity could serve as an early warning sign of an underlying disease, which, if detected promptly, could lead to timely intervention and improved treatment outcomes [10]. In addition, research has demonstrated that patient outcomes in intensive care units improve significantly when pain is monitored on a regular basis [11]. Nevertheless, current observation-based systems are inherently subjective, as they rely on the perceptions and judgments of healthcare professionals, who may be influenced by personal biases. Moreover, these systems require specialized training, and in overcrowded or understaffed hospitals, it is often not feasible for caregivers to monitor patients continuously for signs of distress or emergency situations [8],[6]. Given these limitations, there is a clear and urgent need for an automated pain assessment system that operates reliably and consistently. Such a solution could not only support healthcare workers in clinical environments but also play a crucial role in elderly care, especially in home-based settings where continuous professional supervision is not always possible[8],[12],[13].

On the other hand, many great contributors have placed significant contributions in pain intensity recognition from the facial expression leveraging the deep learning approaches. Rathee N. & Ganotra D.[14], used McMaster Shoulder Pain Database in 2016 to identify pain intensity. They used four pain levels: PSPI 0, 1, 2-3, and ≥ 4. First, they had detected pain then pain intensity. For this purpose, they applied SVM, achieving an accuracy of 89.59% for detecting pain and 75% for estimating pain intensity. Similarly, Werner et al.[15], utilized SVM for both pain detection and intensity evaluation. They conducted their study on both datasets UNBC McMaster and BioVid. Their model obtained an accuracy of 51.6%. In 2017, Lopez-Martinez et al. [16], utilized the BioVid heat-stimulated Pain dataset to measure pain intensity using a multi-task learning framework with neural networks. Since the original dataset contained raw signals, feature extraction was performed prior to feeding the data into the model. They employed binary classification for estimating pain intensity and evaluated their approach using 10-fold cross-validation, achieving an accuracy of 66.68%. Sourav Dey Roy et al. [4], utilized Principal Component Analysis (PCA) to reduce the dimensionality of feature vectors and implemented an SVM model to distinguish between pain and no-pain categories. They further classified pain into four levels: "No Pain", "Tolerable Pain", "Weak Pain", and "Strong Pain". Their study, conducted on the UNBC McMaster Shoulder Pain Archive Database, achieved an accuracy of 87.23% for pain detection and 82.43% for pain level classification. UNBC McMaster Shoulder Pain Archive Database is also used by Ghazal Bargshady et al [17], to classify four pain levels: "No Pain", "Weak Pain", "Mild Pain", and "Strong Pain". They used 224 x 224 pixels of image. Instead of using PCA, they have used ZCA (Zero-phase Component Analysis). They used a combined VGG Face model and RNN, from which they received an average accuracy of 75.2%.

Zakia Hammal et al.[18], researched the UNBC McMaster Shoulder Pain Archive Database, where the image pixel was 96 x 96. They used the SVM model for classification and PSPI pain labels. Four levels were identified:

"PSPI = 0 for No Pain", "PSPI = 1 for Pain Trace", "PSPI = 2 for Weak Pain", and "PSPI $\geq$ 3 for Strong Pain". For each pain level, the respective accuracies were 97%, 96%, 96%, and 98%. In 2020, Tavakolian et al.[19], conducted a study using two datasets: the UNBC McMaster Shoulder Pain Database and the BioVid Database. They applied a Convolutional Neural Network (CNN) to train their model. For pain intensity estimation, they categorized the UNBC McMaster dataset into 16 pain levels and the BioVid dataset into 5 levels. After training, they tested their model on a different dataset. In 2021, Othman et al.[20] classified four levels of pain: "No pain", "Low pain", "Medium Pain", and "Severe Pain" from the X-ITE Pain database. They used two CNNs with sample weighting and received an accuracy of 51.7%.

In 2022, Barua et al.[21], categorized 16 levels of PSPI into 4 levels like PSPI 0, 1, 2-3, and ≥ 4. They have conducted their research on the UNBC McMaster database. They trained their model with K-Nearest Neighbor and received an accuracy of 95%. In 2023, Wu et al.[22], introduced a convolutional neural network framework called GLA-CNN, incorporating both global and local attention mechanisms through GANet and LANet. This model was created to accurately assess pain intensity at four distinct levels by analyzing facial expression images from the UNBC McMaster Shoulder Pain Dataset. The GLACNN model reached an accuracy of 56.45%. Sharafi et al. [23] conducted their study using videos from the UNBC-McMaster database. They employed Source-Free Domain Adaptation (SFDA), utilizing neutral images for adaptation, while both neutral and pain images were used during testing. The authors introduced six pain levels: 0 (neutral), 1 (PSPI = 1), 2 (PSPI = 2), 3 (PSPI = 3), 4 (PSPI = 4–5), and 5 (PSPI = 6–15) and achieved accuracy 79.56%.

From the above discussion, it is evident that no existing research has successfully achieved accurate detection of pain intensity or effectively categorized intensity levels according to the PSPI scale. Nearly all studies classify severe pain within a broad range—from values above 3 up to 15, which creates an imbalanced distribution, especially when compared to lower pain levels that have more image data. Moreover, methods such as SVM, K-means nearest neighbor, or RNN may not be the most suitable approaches for enabling machines to recognize facial expressions, as they often result in low accuracy.

To address these challenges, the primary objective of this research is to develop an intelligent system capable of detecting pain intensity through the analysis of human facial expressions using image processing techniques. The proposed system focuses on several key goals, outlined as follows:

- To design a robust classification method that utilizes the Facial Action Coding System (FACS) and the Prkachin and Solomon Pain Intensity (PSPI) scale, enabling a four-level categorization of pain intensity, including a 'no pain' category.
- To construct predictive models by applying both traditional statistical techniques and deep learning-based feature extraction to classify facial expressions into defined pain levels: no pain, low pain, mild pain, and severe pain.
- Finally validate the performance of these models using test datasets, analyze their training and testing accuracy and loss, and examine additional metrics to assess their effectiveness.

This manuscript is partitioned into four sections. Section two illustrates the materials and methods of this manuscript. Section three provides the results of the proposed methodology and model. Finally, Section four represents the conclusion of this manuscript with the existing drawbacks and future scopes.

2. Materials and Method

This section outlines the structure and operational flow of the proposed model, presented in two main segments: the dataset description and the core processing pipeline. Initially, this section begins with a comprehensive overview of the system's operational framework, highlighting the fundamental technologies that underpin its development—namely, image processing, transfer learning, and artificial neural networks. The overall architecture of the proposed system is illustrated in Fig. 1. Specifically, the dataset comprises images organized into two separate folders, each annotated with a distinct labeling scheme: Facial Action Coding System (FACS) and Prkachin and Solomon Pain Intensity (PSPI). For this study, only the PSPI labels were utilized. During the image processing phase, each image was labeled according to its corresponding PSPI value, and images classified as 'no pain'

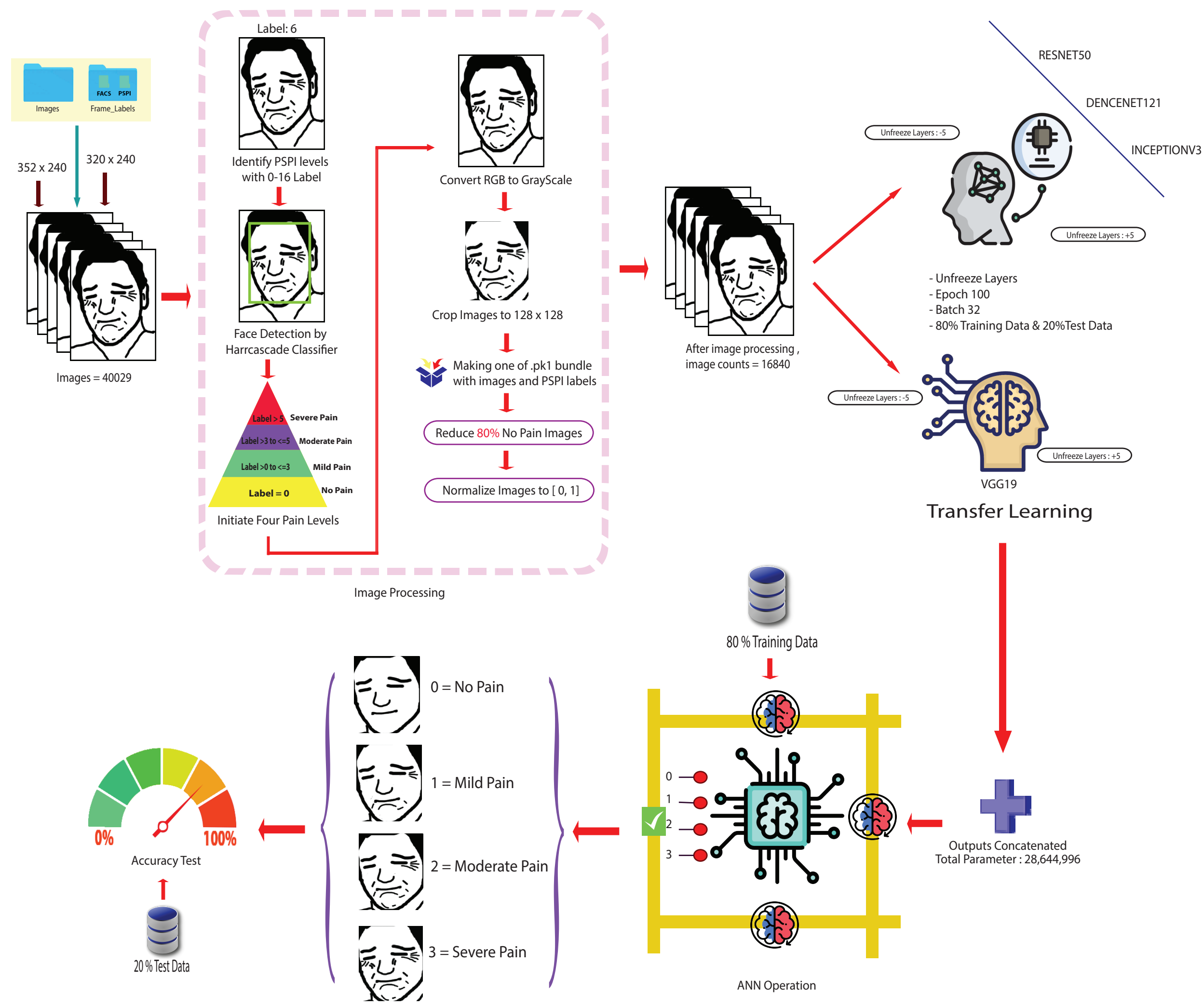

Figure 1. Pain intensity detection model

were selectively reduced to balance the dataset. Following this, the pre-processed images were introduced into the transfer learning pipeline. In this stage, various modified convolutional neural network (CNN) architectures were employed. CNNs, a type of feed-forward neural network, are widely recognized for their effectiveness in image recognition tasks. They encode input data into multidimensional arrays and perform particularly well when trained on large volumes of labeled image data. Moreover, two distinct CNN models were used in the transfer learning process. Their outputs were combined and subsequently passed into an artificial neural network for further classification. To train and evaluate the model, 80% of the data was allocated for training, while the remaining 20% was used for testing. Finally, the performance of the proposed system was assessed using the test set, with accuracy measured on a scale from 0 to 100 percent.

2.1. *Dataset*

Table-1 presents the datasets utilized by researchers for pain detection. In this study, we used the UNBC-McMaster Shoulder Pain Expression Archive Dataset, a unique collection of data from 25 real patients experiencing shoulder

Table 1. Datasets for Detecting Pain Intensity [24]

Datasets	Year of Creation	Patient Condition	Size of Sample
UNBC McMaster shoulder pain dataset	2011	Shoulder pain	Total Frames: 48,398 and pain frames: 8,369
BioVid Heat Pain Dataset	2013	Healthy	Total: 8700 videos and pain videos: 6960
DISFA Dataset	2013	Healthy	Video Frames: 4845
EmoPain Dataset	2015	Chronic lower back pain / Healthy	Total Frames: 585,487 and pain frames: 50,071
X-ITE Pain Dataset	2019	Healthy	-

pain. A total of 48,398 frames were extracted from the video sequences and annotated with Action Units (AUs) by certified FACS coders [25],[26].

Action Units (AUs) are the fundamental actions of certain feelings of expression, such as lips parting, nose wrinkling, and blinking. ”There are 44 AUs by which changes of expression can be described, which was first developed by two scientists in the 1970s (Ekman and Friesen, 1978) [3]”. AUs are encoded using the FACS. Facial Action Coding System (FACS) assigns every AU a unique code like AU number 7 corresponds to lid tightening, AU number 23 corresponds to lip tightening, AU number 27 corresponds to mouth stretching [27].

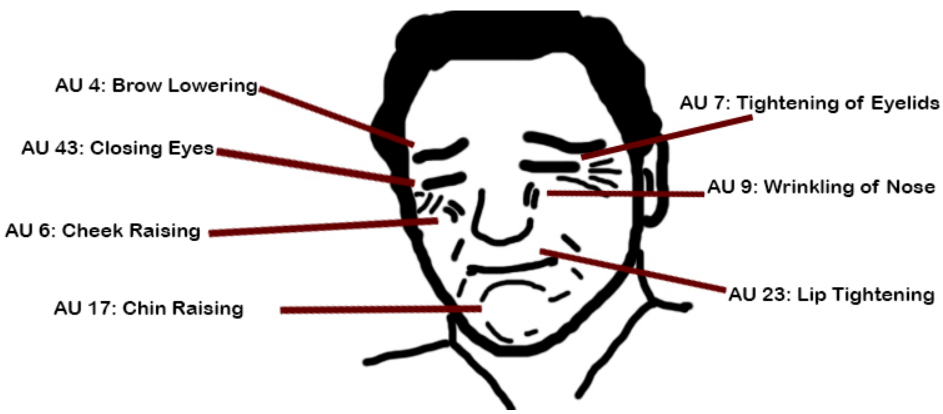

Figure 2. Pain expressive face with action units

In 1992, the Prkachin and Solomon Pain Intensity Scale (PSPI) was developed by Prkachin and Solomon [28]. They discovered that out of 44 AU, only 4 actions are responsible for pain expression and defined one formula [29],[30]:

$$\text{Pain} = \text{AU4} + (\text{AU6} \vee \text{AU7}) + (\text{AU9} \vee \text{AU10}) + \text{AU43}$$

Here, AU4 = brow lowering, AU6 = cheek raising, AU7 = eyelid tightening, AU9 = Nose wrinkling, AU10 = Upper lip raising, AU43 = eye closure [29]. AU4 and AU43 always present in pain expression, one of AU6 and AU7 and one of AU9 and AU10 also be present [29],[17],[31]. In the case of AU6, AU7, and AU9, AU10, if both are present then the highest intensity would be selected [32], [33],[30].

Each AU intensity is scored in the range of 0-5, where 0 is AU absent and 5 represents the highest intensity in "a-f" scale. Only AU 43 belongs to binary intensity: 0 or 1. If a frame is coded as:

$$\text{FACS} = \text{AU4a} + \text{AU6d} + \text{AU7d} + \text{AU12d} + \text{AU43} \tag{1}$$

Then the PSPI score would be:

PSPI = 1 + max (4,4) + 1 = 6 [15]

2.2. Core pipeline of the proposed model

In recent years, transfer learning has emerged as a powerful technique for image classification, owing to its ability to utilize prior knowledge derived from related tasks [34], [35], [36]. Among the most widely adopted pre-trained convolutional neural network (CNN) models in transfer learning are VGG16, VGG19, ResNet50, DenseNet121, and InceptionV3 [37]. Each of these models has demonstrated exceptional performance across various benchmark datasets and continues to be instrumental in deep learning-based image classification research.

To begin with, the VGG16 model—developed by the Visual Geometry Group (VGG) at the University of Oxford—comprises 13 convolutional layers followed by 3 fully connected layers [38], [39]. Its architecture is characterized by a sequence of stacked convolutional layers interspersed with max-pooling layers, with an increasing depth across layers to facilitate the extraction of hierarchical features. Notably, VGG16 achieved significant success in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), ranking highly in both object detection (across 200 categories) and image classification (across 1000 categories) [18].

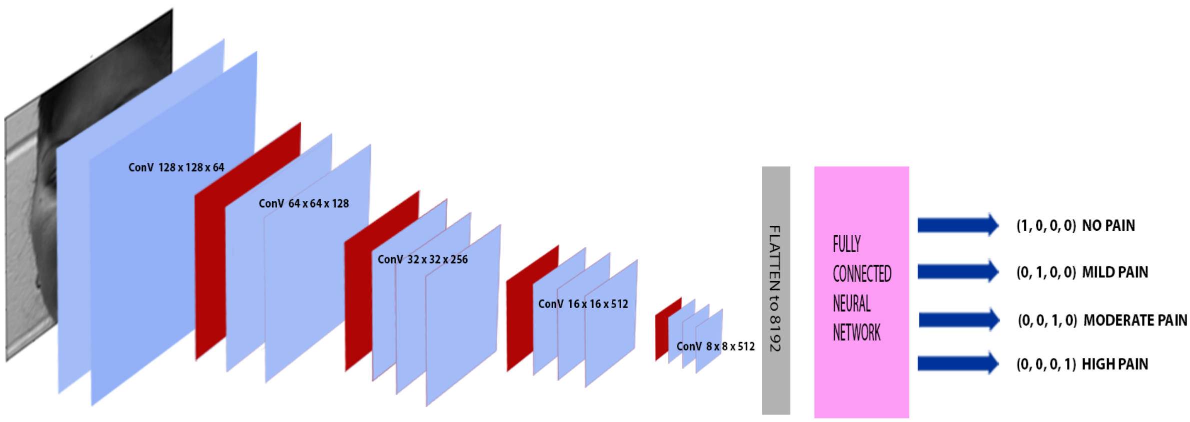

Figure 3. VGG16 in transfer learning process

Rather than having 19 layers consisting of 16 convolutional layers and 3 fully connected layers, VGG-19 follows the same basic structure as the VGG-16 model and was developed by the Visual Geometry Group. It includes 5 max-pooling layers and uses 3 x 3 convolutional filters (similar to the VGG-16 model) throughout its convolutional layers, employing max-pooling layers for downsampling spatial dimensions [40]. The additional layers in VGG-19 are designed to capture more complex patterns in images, allowing it to learn richer representations.

In contrast, ResNet-50—proposed by Microsoft Research Asia—introduced the concept of residual learning, which enables the training of very deep networks by addressing the vanishing gradient problem. The ResNet-50 architecture consists of 50 layers organized into four main components: a convolutional layer, identity blocks, convolutional blocks, and fully connected layers. One of its key innovations is the shortcut connection, wherein the input to a layer is added to the output of a convolutional block, allowing gradients to flow more effectively during backpropagation [37]. ResNet-50 secured first place in the 2015 ILSVRC for classification, detection, and localization tasks.

DenseNet121, short for "Densely Connected Convolutional Networks," presents another innovative architecture. It connects each layer to every other layer in a feedforward fashion, promoting efficient feature reuse and mitigating the vanishing gradient issue. Introduced in 2016, DenseNet121 consists of 121 layers and has achieved state-of-the-art performance in various image classification tasks.

Similarly, InceptionV3, developed by Google researchers, belongs to the Inception family that originated with GoogLeNet. It builds upon its predecessors by employing factorized convolutions, auxiliary classifiers, and label smoothing techniques. Notably, InceptionV3 achieved top-tier performance in the ILSVRC and is recognized for its efficiency and accuracy in large-scale image recognition.

Leveraging the strengths of these individual models, this study explores hybrid configurations that combine the complementary features of multiple architectures. For instance, integrating VGG19's consistent and deep feature extraction with DenseNet121's dense connectivity enhances the model's representational richness and gradient flow, ultimately leading to improved classification performance [41]. VGG19's uniform convolutional structure ensures fine-grained feature learning, while DenseNet121's layered interconnections allow for deeper network training without degradation [34].

To further evaluate the effectiveness of feature fusion models, three parallel feature fusion models (PFFM-1, PFFM-2, and PFFM-3) were developed. Specifically, PFFM-1 integrates VGG19, ResNet50, and an artificial neural network (ANN), whereas PFFM-2 and PFFM-3 comprise VGG19 with DenseNet121 and ANN, and VGG19 with InceptionV3 and ANN, respectively. Both models independently extract features from the same input image, and their outputs are then combined. Here's a detailed explanation of why this setup is considered a feature fusion parallel configuration:

- Input Shape: The input shape is defined as (128, 128, 3), which is a common input size for both VGG19 and DenseNet121 models.

- Loading Models: Both the VGG19 and DenseNet121 models are initialized with pre-trained weights from ImageNet, with the exception of their top classification layers. As a result, they are utilized primarily as feature extractors.

- Parallel Feature Extraction: The models will process the input images in parallel, each extracting its own set of features independently.

In this setup:

- Parallel Processing: Both VGG19 and DenseNet121 process the same input image in parallel.
- Feature Combination: The features extracted by each model are flattened and concatenated.
- Fully Connected Layers (ANN): The concatenated features are routed through fully connected layers for the final classification process.

This architecture allows us to combine the strengths of both models, leveraging VGG19's fine-grained feature extraction and DenseNet121's deep-learning capabilities to improve overall performance.

3. Result and Discussion

This section is structured into three cohesive segments. The first segment elaborates on the sequential steps undertaken during the image processing phase. The second provides a comprehensive overview of the implementation methodology, supported by detailed discussion. The final segment presents a comparative evaluation table to validate the research contribution and demonstrate the effectiveness of the proposed approach. All experiments for pain intensity recognition from facial expressions were conducted using Google Colab Pro Plus, a cloud-based platform offering high-performance GPUs like NVIDIA Tesla P100 and A100. With up to 52GB of RAM and extended runtime limits, it enabled efficient processing of large image datasets and training of feature fusion-based CNN models without local hardware limitations. Moreover, its seamless integration

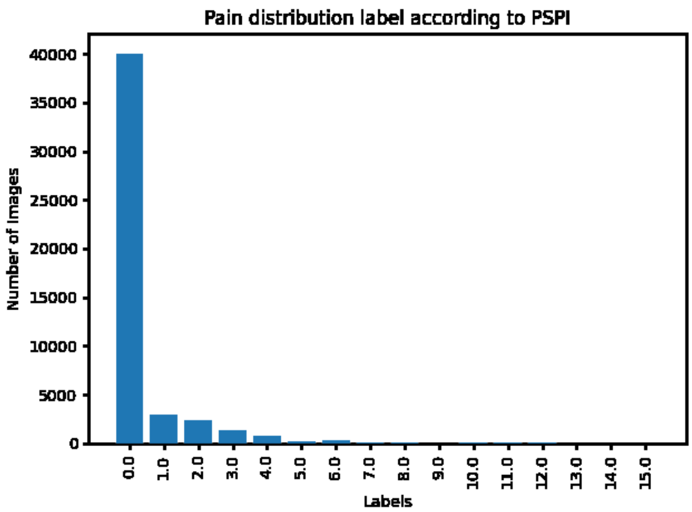

Figure 4. Number of images under 0-15 labels of PSPI scores

with Python libraries such as TensorFlow, Keras, OpenCV, and scikit-learn streamlined both development and evaluation. This setup was particularly beneficial for managing the computational demands of the entire model development process.

Table 2. Pain Intensity Levels and Image Count for Various Categories

Pain Intensity level according to PSPI scale	Number of images for 16 levels	Number of images for 4 levels with 20% no pain images	Number of images for 4 levels with 10% no pain images
0	40029	8471	4017
1	2909		
2	2351	6672	6672
3	1412		
4	802	1044	1044
5	242		
6	270		
7	53		
8	79		
9	32		
10	67		
11	76	653	653
12	48		
13	22		
14	1		
15	5		

3.1. Image Processing

Most of the 48,398 color frames in the UNBC-McMaster shoulder pain dataset have an average resolution of 320 x 240 pixels. The dataset is significantly imbalanced, with 82.71% (40,029 frames) categorized as ”no pain,” while the remaining 17.29% (8,369 frames) correspond to various pain levels [8].

Using the entire dataset can lead to biased classification and increased computational costs. As a result, many researchers opt to use a subset of the UNBC-McMaster dataset [8]. In this study, we randomly reduced the number

Table 3. Pain Intensity Levels and Corresponding Pain Classes

Pain Intensity Levels	Pain Classes
0 = 0	No Pain
1 = (<0 to >=3)	Mild Pain
2 = (<3 to >=5)	Moderate Pain
3 = (<5 to 15)	Severe Pain

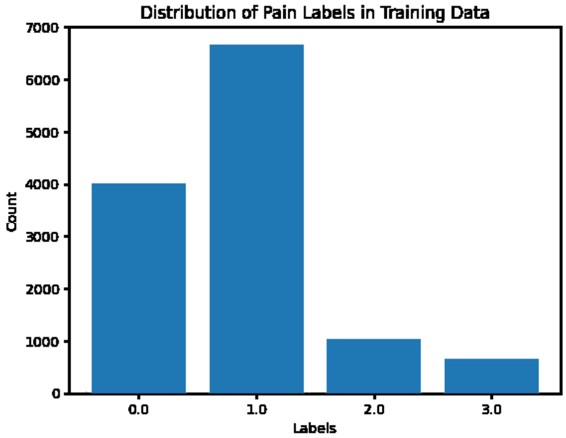

Figure 5. Image counts with 10 percent ”No pain” Images

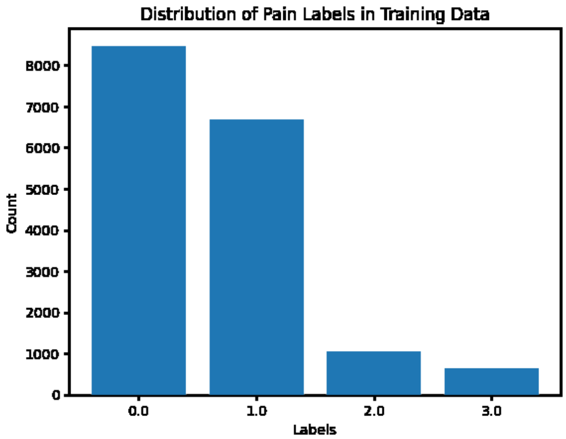

Figure 6. Image counts with 20 percent ”No pain” Images

of ”no pain” images from our dataset to evaluate the accuracy of our model. We have redefined the 16 PSPI pain levels into four clear categories: 0, 1, 2, and 3, which correspond to “No Pain,” “Mild Pain,” “Moderate Pain,” and “Severe Pain,” respectively. The ”Severe Pain” category includes PSPI values above 5 up to 15. However, due to the limited number of images available starting from level 10, it becomes difficult to train and test the model effectively. To address this, we have expanded the ”Severe Pain” category to include all levels above 5, ensuring a sufficient number of images for robust model training and evaluation [42].

The tables show the total number of images used to train our model, including 20% and 10% labeled as “No Pain,” along with the counts for the other categories.

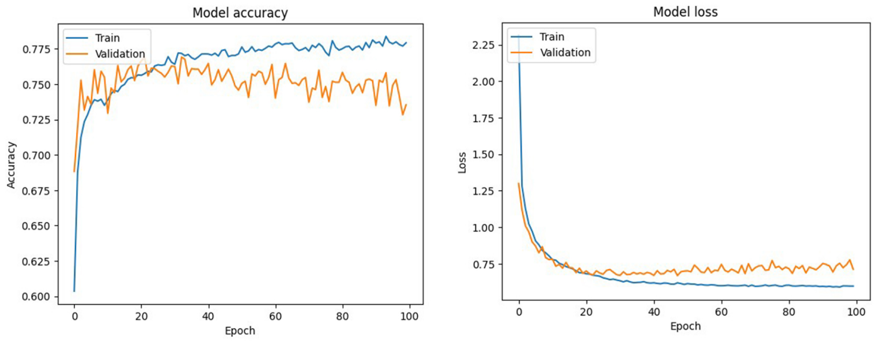

Fig-a: Model Accuracy and Loss of VGG16

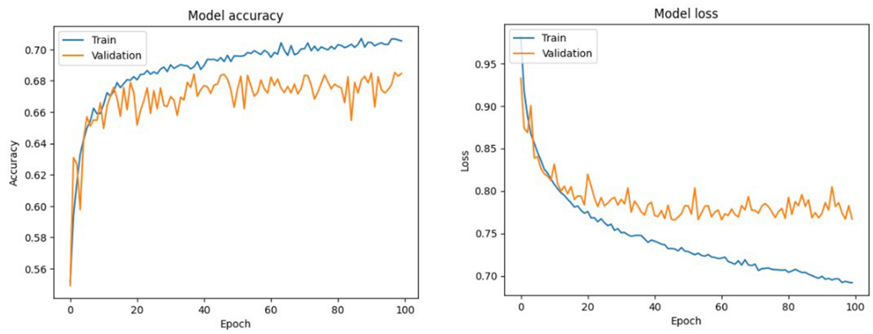

Fig-b: Model Accuracy and Loss of PHM-1

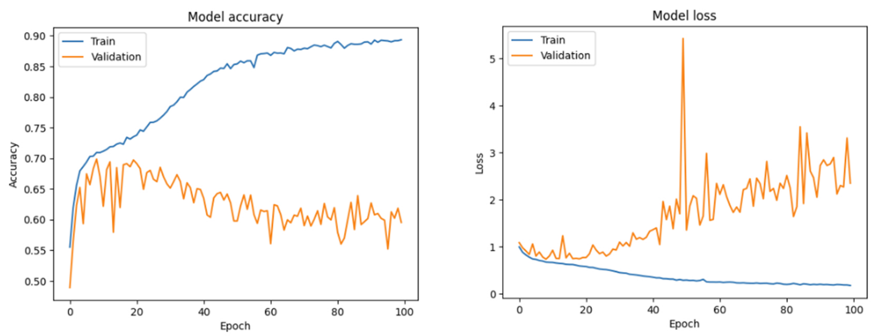

Fig-c: Model Accuracy and Loss of PHM-2

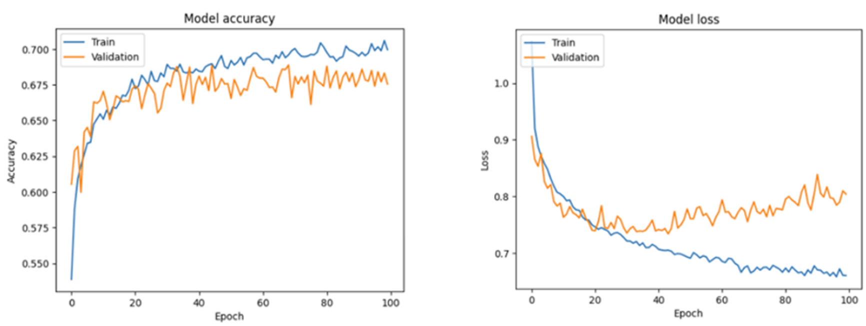

Fig-d: Model Accuracy and Loss of PHM-3

Figure 7. Model Accuracy and Losses for different Parallel Feature Fusion Models

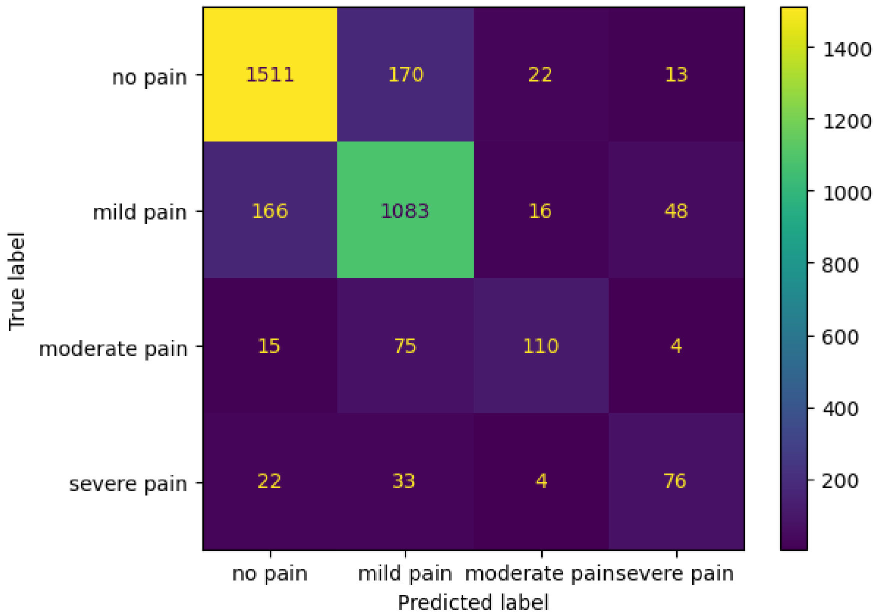

Figure 8. Confusion matrix of PFFM-2

To manage the large dataset and the similarity between "no pain" and pain expressions, we reduced the number of "no pain" images randomly. The figure shows samples from the UNBC McMaster Shoulder Pain Dataset [17].

Pain intensity was divided into four levels, and faces were detected using the Haarcascade Frontal Face Detector [42]. Images were cropped to 128x128 for the CNN architecture, converted to grayscale for simplicity, and normalized to a pixel range of [0, 1] [43]. Image normalization was used to improve training efficiency.

3.2. *Implementation Details*

Assessing a model's performance is essential to determine its effectiveness in making predictions. Commonly used evaluation metrics include Accuracy, Precision, Recall, and F1-Score. These metrics offer valuable insights, particularly in cases of imbalanced data or when the costs of false positives and false negatives differ [44].

In deep learning, having a larger dataset helps models generalize better, improving their performance on unseen data. Conversely, a smaller dataset may lead to overfitting, resulting in poor performance. To mitigate overfitting, dropout techniques were applied. The testing phase of this research focused on evaluating the performance and effectiveness of the developed pain intensity system. This section outlines the metrics (accuracy) obtained during the testing process. In this study, we count accuracy, precision, recall, and F1-score, which are the percentage of correctly classified instances among all instances [45].

Table 4. Performance metrics per pain category based on the confusion matrix of PFFM-2.

Pain Category	Precision	Recall	F1-Score	Accuracy
No Pain	0.8816	0.8805	0.8810	0.8789
Mild Pain	0.7957	0.8248	0.8100	0.8492
Moderate Pain	0.7237	0.5392	0.6180	0.9596
Severe Pain	0.5390	0.5630	0.5507	0.9632

Additionally, the model's robustness was validated using random images. For training and testing, 80% of the dataset was allocated for training, while 20% was reserved for testing [46],[45],[8].Furthermore, a 5-fold cross-validation strategy was employed to ensure a robust and generalizable evaluation of the proposed network.

Table 5. 5-Fold Cross-Validation Results for PFFM-2

Model	Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
PFFM-2	Fold 1	82.54	73.50	70.18	71.48
	Fold 2	82.54	73.50	70.18	71.48
	Fold 3	82.54	73.50	70.18	71.48
	Fold 4	82.54	73.50	70.18	71.48
	Fold 5	82.54	73.50	70.18	71.48
Average		**82.54**	**73.50**	**70.18**	**71.48**

Table 6. Performance of different models with varying percentages of no pain images

Models	Accuracy	Precision	Recall	F1 Score
Models for 20% no pain images				
VGG16	76.63%	0.58	0.48	0.5
VGG19	62.17%	0.42	0.42	0.42
ResNet50	45.25%	0.24	0.25	0.24
DenseNet121	44.00%	0.23	0.24	0.23
InceptionV3	44.39%	0.24	0.25	0.24
PHM-1	70.99%	0.5	0.41	0.41
PHM-2	82.54%	0.73	0.7	0.71
PHM-3	71.94%	0.53	0.45	0.47
Models for 10% no pain images				
VGG16	60.49%	0.41	0.37	0.37
VGG19	58.31%	0.43	0.35	0.36

We have taken 100 epochs based on initial experimental runs, which indicated stable convergence without signs of overfitting and set the batch size to 32 and ran the models. VGG16 had a training accuracy of 77.93% and a validation accuracy of 73.54%. After obtaining these results, we fine-tuned our VGG16 model by unfreezing the last five layers and applying regularization. Additionally, we increased the Dropout rate from 0.3 to 0.5, which led to an improvement in accuracy of 76.63%. As a result, the precision improved by 8%. 62.1%, 45.25%, 44.0% and 44.39% are accuracy of VGG19, ResNet50, DenseNet121 and InceptionV3 respectively. Furthermore, the model was trained using the Adam optimizer, applying its default learning rate of 0.001 as commonly used in deep learning tasks.

Our feature fusion parallel models have shown comparatively better accuracy than individual models.70.99%, 82.54%,71.94% are some accuracy of PFFM-1, PFFM-2, and PFFM-3 respectively. The features of two deep CNN models were concatenated here. PFFM-2 has the best accuracy result among the three parallel models from our end.

PFFM-3 has comparatively lower accuracy than PFFM-2, but better accuracy than PFFM-1. We have also checked our pain level with 10% no pain images, where VGG16 and VGG19 has an accuracy of 60.49% and 58.31%.

3.3. Comparison Table

As shown in Table 7, the proposed Parallel Feature Fusion Models (PFFMs) are evaluated against five established hybrid models to assess their effectiveness in classification tasks. The results indicate considerable variation in the performance of existing hybrid methods. GLA-CNN, introduced by Jiang Wu et al. (2023), reported the

Table 7. Comparison of Different hybrid Methods and Their Accuracy

Author / Year	**Method Used**	**Accuracy**
Bellantonio et al. 2017 [26]	RNN and CNN	63.47%
Bargshady et al. 2019 [17]	Vgg16 and RNN	75.20%
Tavakolian et al. 2020 [19]	Statistical Spatiotemporal Distillation	76.00%
Jiang Wu et al. 2023 [22]	GLA-CNN	56.45%
Sharaf et al. 2025 [23]	SFDA	79.56%
	PFFM-1	70.99%
Our Parallel Feature Fusion models	**PFFM-2**	**82.54%**
	PFFM-3	71.94%

lowest accuracy at 56.45%, suggesting limited capacity in capturing discriminative features. Bellantonio et al. (2017) achieved a modest improvement with their RNN-CNN hybrid, reaching 63.47%. A more substantial gain was observed in the Vgg16-RNN approach by Barshayoy et al. (2019), which attained an accuracy of 75.20%. Likewise, Tavakolian et al. (2020) reported 76.00% using a statistical spatiotemporal distillation method. The highest-performing among these was the SFDA model by Sharaf et al. (2025), achieving 79.56%. In contrast, the PFFM architecture—while not a hybrid in the conventional architectural sense—can be understood as a hybrid feature fusion strategy, as it integrates parallel streams of features extracted from different parts of the network. This design proves highly effective: PFFM-2 outperforms all models in the comparison with an accuracy of 82.54%, while PFFM-3 and PFFM-1 also yield strong results at 71.94% and 70.99%, respectively. These findings underscore the advantage of parallel feature fusion in enhancing the richness of learned representations and improving overall classification performance, positioning PFFM as a competitive alternative to more traditional hybrid modeling techniques.

4. Conclusion

This research presents a novel approach to detecting pain intensity from facial expressions through a feature fusion parallel model architecture. The models, combining features from deep CNNs and a fully connected network, demonstrated improved performance compared to existing pre-trained models, with PFFM-2 achieving the highest accuracy of 82.54%. This study highlights the potential of leveraging multiple deep learning frameworks in parallel to enhance the precision of pain detection systems. The significance of our findings lies in the advancement of AI-driven pain assessment tools, which could greatly benefit clinical environments by providing more accurate, non-verbal assessments of patient discomfort. However, there is still room for further improvement in model generalization and the exploration of additional datasets and architectures. Future research should refine these models and validate their effectiveness across diverse populations and varied clinical settings to ensure reliability and practical application.

Ethical Considerations

In this study, we employed the UNBC-McMaster Shoulder Pain Expression Archive Database, which is a publicly available resource collected and distributed by the original researchers in accordance with established ethical protocols. All data included in the dataset are fully anonymized, and informed consent was obtained by the dataset creators prior to its release. As no new data were collected and there was no direct involvement of human participants by the authors, additional ethical approval was not required for this work. Nevertheless, we recognize the critical importance of ethical transparency in research involving human-related data. Therefore, we have addressed potential sources of bias inherent in the dataset and discussed relevant limitations to promote responsible and fair use of the model in any future applications.

REFERENCES

1. Ahmad Al-Qerem, "An efficient machine-learning model based on data augmentation for pain intensity recognition," *Egyptian Informatics Journal*, vol. 21, no. 4, pp. 241–257, 2020, Elsevier.
2. Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F. Cohn, Tsuhan Chen, Zara Ambadar, Ken Prkachin, Patty Solomon, and Barry J. Theobald, "The painful face: Pain expression recognition using active appearance models," in *Proceedings of the 9th International Conference on Multimodal Interfaces*, pp. 9–14, 2007.
3. P. Ekman, "Facial expression and emotion," *American Psychologist*, 1993.
4. Sourav Dey Roy, Mrinal Kanti Bhowmik, Priya Saha, and Anjan Kumar Ghosh, "An approach for automatic pain detection through facial expression," *Procedia Computer Science*, **84**, pp. 99–106 (2016).
5. Kornprom Pikulkaew, Ekkarat Boonchieng, Waraporn Boonchieng, and Varin Chouvatut, "Pain detection using deep learning with evaluation system," in *Proceedings of the Fifth International Congress on Information and Communication Technology (ICICT 2020), London, Volume 2*, pp. 426–435 (2021), Springer.
6. Ioannis Karamitsos, IIham Seladji, and Sanjay Modak, "A modified CNN network for automatic pain identification using facial expressions," *Journal of Software Engineering and Applications*, **14**(8), pp. 400–417 (2021), Scientific Research Publishing.
7. Gioacchino D. De Sario, Clifton R. Haider, Karla C. Maita, Ricardo A. Torres-Guzman, Omar S. Emam, Francisco R. Avila, John P. Garcia, Sahar Borna, Christopher J. McLeod, Charles J. Bruce, et al., "Using AI to detect pain through facial expressions: a review," *Bioengineering*, **10**(5), p. 548 (2023), MDPI.
8. Thoria Alghamdi and Gita Alaghband, "Facial expressions based automatic pain assessment system," *Applied Sciences*, **12**(13), p. 6423 (2022), MDPI.
9. Gangireddy Prabhakar Reddy and M. Kalaiselvi Geetha, "Deep learning based model for pain recognition from facial expression," *Journal of Theoretical and Applied Information Technology*, **100**(3) (2022).
10. Wijdan Aljebreen, Dina M. Ibrahim, et al., "PainMeter: Automatic assessment of pain intensity levels from multiple physiological signals using machine learning," *IEEE Access*, (2024), IEEE.
11. Dong Huang, Zhaoqiang Xia, Lei Li, Kunwei Wang, and Xiaoyi Feng, "Pain-awareness multistream convolutional neural network for pain estimation," *Journal of Electronic Imaging*, **28**(4), pp. 043008–043008 (2019), Society of Photo-Optical Instrumentation Engineers.
12. Ghazal Bargshady, Xujuan Zhou, Ravinesh C. Deo, Jeffrey Soar, Frank Whittaker, and Hua Wang, "Ensemble neural network approach detecting pain intensity from facial expressions," *Artificial Intelligence in Medicine*, **109**, p. 101954 (2020), Elsevier.
13. D. Wong, "Pain assessment in children," *Anaesthesia & Intensive Care Medicine*, vol. 26, no. 3, pp. 139–142, 2025.
14. Neeru Rathee and Dinesh Ganotra, "Multiview distance metric learning on facial feature descriptors for automatic pain intensity detection," *Computer Vision and Image Understanding*, **147**, pp. 77–86 (2016), Elsevier.
15. Philipp Werner, Ayoub Al-Hamadi, Kerstin Limbrecht-Ecklundt, Steffen Walter, Sascha Gruss, and Harald C. Traue, "Automatic pain assessment with facial activity descriptors," *IEEE Transactions on Affective Computing*, **8**(3), pp. 286–299 (2016), IEEE.
16. Daniel Lopez-Martinez and Rosalind Picard, "Multi-task neural networks for personalized pain recognition from physiological signals," in *Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 181–184 (2017), IEEE.
17. Ghazal Bargshady, Jeffrey Soar, Xujuan Zhou, Ravinesh C. Deo, Frank Whittaker, and Hua Wang, "A joint deep neural network model for pain recognition from face," in *Proceedings of the 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pp. 52–56 (2019), IEEE.
18. Zakia Hammal and Jeffrey F. Cohn, "Automatic detection of pain intensity," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, pp. 47–52 (2012).
19. Mohammad Tavakolian, Miguel Bordallo Lopez, and Li Liu, "Self-supervised pain intensity estimation from facial videos via statistical spatiotemporal distillation," *Pattern Recognition Letters*, **140**, pp. 26–33 (2020), Elsevier.
20. Ehsan Othman, Philipp Werner, Frerk Saxen, Ayoub Al-Hamadi, Sascha Gruss, and Steffen Walter, "Automatic vs. human recognition of pain intensity from facial expression on the X-ITE pain database," *Sensors*, **21**(9), p. 3273 (2021), MDPI.
21. Prabal Datta Barua, Nursena Baygin, Sengul Dogan, Mehmet Baygin, N. Arunkumar, Hamido Fujita, Turker Tuncer, Ru-San Tan, Elizabeth Palmer, Muhammad Mokhzaini Bin Azizan, et al., "Automated detection of pain levels using deep feature extraction from shutter blinds-based dynamic-sized horizontal patches with facial images," *Scientific Reports*, **12**(1), p. 17297 (2022), Nature Publishing Group UK London.
22. Jiang Wu, Yi Shi, Shun Yan, and Hong-mei Yan, "Global-local combined features to detect pain intensity from facial expression images with attention mechanism," *Journal of Electronic Science and Technology*, p. 100260 (2024), Elsevier.
23. Masoumeh Sharafi, Emma Ollivier, Muhammad Osama Zeeshan, Soufiane Belharbi, Marco Pedersoli, Alessandro Lameiras Koerich, Simon Bacon, et al., "Disentangled source-free personalization for facial expression recognition with neutral target data," *arXiv preprint*, arXiv:2503.20771 (2025).
24. Uma Nath, "Facial expressions as behavioral indicators for assessing pain using machine learning models," *Journal of Student Research*, **12**(4) (2023).
25. Patrick Lucey, Jeffrey F. Cohn, Kenneth M. Prkachin, Patricia E. Solomon, and Iain Matthews, "Painful data: The UNBC-McMaster shoulder pain expression archive database," in *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 57–64 (2011), IEEE.
26. Marco Bellantonio, Mohammad A. Haque, Pau Rodriguez, Kamal Nasrollahi, Taisi Telve, Sergio Escalera, Jordi Gonzalez, Thomas B. Moeslund, Pejman Rasti, and Gholamreza Anbarjafari, "Spatio-temporal pain recognition in CNN-based super-resolved facial images," in *Proc. Int. Workshop on Video Analytics. Face and Facial Expression Recognition and Audience Measurement (VAAM & FFER)*, Cancun, Mexico, Dec. 2016, Revised Selected Papers 2, pp. 151–162, Springer, 2017.
27. Gioacchino D. De Sario, Clifton R. Haider, Karla C. Maita, Ricardo A. Torres-Guzman, Omar S. Emam, Francisco R. Avila, John P. Garcia, Sahar Borna, Christopher J. McLeod, Charles J. Bruce, et al., "Using AI to detect pain through facial expressions: a review," *Bioengineering*, **10**(5), p. 548 (2023), MDPI.

28. Kenneth M. Prkachin and Patricia E. Solomon, "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain*, **139**(2), pp. 267–274 (2008), Elsevier.
29. Kornprom Pikulkaew, Ekkarat Boonchieng, Waraporn Boonchieng, and Varin Chouvatut, "Pain detection using deep learning with evaluation system," in *Proceedings of Fifth International Congress on Information and Communication Technology: ICICT 2020, London, Volume 2*, pp. 426–435 (2021), Springer.
30. D. Green, Y. Shang, J. Cheong, Y. Liu, and H. Gunes, "Gender fairness of machine learning algorithms for pain detection," *arXiv preprint arXiv:2506.11132*, 2025.
31. Ilham Seladji, "Automatic Pain Assessment Through Facial Expressions," 2020.
32. Thoria Alghamdi and Gita Alaghband, "SAFEPA: an expandable multi-pose facial expressions pain assessment method," *Applied Sciences*, **13**(12), p. 7206 (2023), MDPI.
33. Daniel Lopez-Martinez et al., *Machine learning for pain assessment and management*, Ph.D. thesis, Massachusetts Institute of Technology, 2020.
34. Xinting Ye, Xiaokun Liang, Jiani Hu, and Yaoqin Xie, "Image-based pain intensity estimation using parallel CNNs with regional attention," *Bioengineering*, **9**(12), p. 804 (2022), MDPI.
35. Keng Wah Choo and Tiehua Du, "Pain detection from facial landmarks using spatial-temporal deep neural network," in *Proceedings of the Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, vol. 11878, pp. 593–597 (2021), SPIE.
36. Rachid Aliradi, Nidhaleddine Chenni, and Mohammed Touami, "Estimation for pain from facial expression based on XQEDA and deep learning," *International Journal of Information Technology*, **17**(1), pp. 655–663 (2025), Springer.
37. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016).
38. Ashish Semwal and Narendra D. Londhe, "Automated facial expression-based pain assessment using deep convolutional neural network," in *Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 366–370 (2020), IEEE.
39. Safaa El Morabit, Atika Rivenq, Mohammed-En-nadhir Zighem, Abdenour Hadid, Abdeldjalil Ouahabi, and Abdelmalik Taleb-Ahmed, "Automatic pain estimation from facial expressions: A comparative analysis using off-the-shelf CNN architectures," *Electronics*, **10**(16), p. 1926 (2021), MDPI.
40. U. Hwang, J. Han, O. Kwon, Y. S. Chu, and S. Yang, "Classifying office workers with and without cervicogenic headache or neck and shoulder pain using posture-based deep learning models: a multicenter retrospective study," *Frontiers in Pain Research*, vol. 6, p. 1614143, 2025.
41. E. Noori, S. Mutashar, and D. Ali, "Extract the pain features from facial expressions by utilizing two DCNN models," in *Proceedings of the Conference on Communication and Signal Processing (CoCoS)*, pp. 58–63 (2023), Research Gate. doi: 10.24086/cocos2022/paper.686
42. Muhammed Fatih Kaya, *Automated Pattern Recognition of Communication Behaviour in Electronic Business Negotiations*, Springer, 2023.
43. Mahmut Emin Çelik, "A novel deep learning model for pain intensity evaluation," *International Journal of Computational and Experimental Science and Engineering*, **9**(4), pp. 325–330 (2023), İskender AKKURT.
44. Dipesh Dilip Patil, *Facial Emotion Recognition Using Deep Learning*, Ph.D. thesis, California State University, Northridge, 2023.
45. Farkhund Iqbal, Ahmed Abbasi, Abdul Rehman Javed, Ahmad Almadhor, Zunera Jalil, Sajid Anwar, and Imad Rida, "Data augmentation-based novel deep learning method for deepfaked images detection," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 11, pp. 1–15, 2024.
46. Pu Yin, Xiaoxue Zhang, and Lvyuan Hao, "Deep learning assessment method for postoperative pain based on facial video data," *Journal of Physics: Conference Series*, vol. 2356, no. 1, pp. 012052, 2022.