

# GraphAU-Pain: Graph-based Action Unit Representation for Pain Intensity Estimation

Zhiyu Wang<sup>1</sup>, Yang Liu<sup>2,\*</sup> and Hatice Gunes<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Cambridge, Cambridge, UK

<sup>2</sup>Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland

## Abstract

Understanding pain-related facial behaviors is essential for digital healthcare in terms of effective monitoring, assisted diagnostics, and treatment planning, particularly for patients unable to communicate verbally. Existing data-driven methods of detecting pain from facial expressions are limited due to interpretability and severity quantification. To this end, we propose GraphAU-Pain, leveraging a graph-based framework to model facial Action Units (AUs) and their interrelationships for pain intensity estimation. AUs are represented as graph nodes, with co-occurrence relationships as edges, enabling a more expressive depiction of pain-related facial behaviors. By utilizing a relational graph neural network, our framework offers improved interpretability and significant performance gains. Experiments conducted on the publicly available UNBC dataset demonstrate the effectiveness of the GraphAU-Pain, achieving an F1-score of 66.21% and accuracy of 87.61% in pain intensity estimation. The code is available for re-implementation at [github.com/ZW471/GraphAU-Pain](https://github.com/ZW471/GraphAU-Pain).

## Keywords

Pain Intensity Estimation, Facial Expression Analysis, Graph Neural Networks, Deep Learning

## 1. Introduction

Pain detection is critical in clinical and caregiving settings for timely assessment and improved patient management. Current methods like self-reports and observational evaluations have notable limitations [1]. Self-reports are subjective and depend on patient communication ability, often impaired in nonverbal individuals, children, or those with cognitive impairments. Observational methods, while more objective, require extensive training to ensure accuracy and consistency.

Deep learning has driven interest in automated pain estimation via facial expression analysis, bypassing advanced medical equipment while providing objective measures. Methods like CNNs [2, 3] and hybrid frameworks [4, 5, 6] have been applied for pain estimation using facial features. Transformers [7, 8] perform remarkably in pain prediction from facial videos. However, most rely solely on image features, neglecting physiological insights, limiting clinical interpretability. Additionally, undersampling to address dataset imbalance reduces generalizability in diverse populations [9, 10].

Pain estimation can leverage features tied to facial expressions. Landmark-based methods like nose tip or eye corner coordinates outperform pixel-based techniques but lack clear physiological links to pain [11, 12]. The Facial Action Coding System (FACS) [13] maps facial movements into Action Units (AUs) with intensity levels, aggregating into the Prkachin and Solomon Pain Intensity (PSPI) score [14]. The UNBC dataset [15] provides AU and PSPI-labeled videos, supporting AU-informed methods like K-Nearest Neighbor [16] and Bayesian Networks [17], which report high accuracy but suffer from overoptimism due to class imbalance and overlook AU relationships. Recent approaches like GLA-CNN [18] and Multiple Instance Learning [19] have improved AU relation modeling. However, these methods rely on pixel-wise AU relationships, struggling with subtle facial changes and generalization across diverse racial groups, limiting real-world applicability.

MiGA@IJCAI25: International IJCAI Workshop on 3rd Human Behavior Analysis for Emotion Understanding, August 29, 2025, Guangzhou, China.

\*Corresponding author. Y. Liu contributed to this work while he was a visiting postdoctoral researcher at the AFAR Lab, Department of Computer Science and Technology, University of Cambridge, UK.

✉ zw471@cam.ac.uk (Z. Wang); yang.liu@oulu.fi (Y. Liu); hg410@cam.ac.uk (H. Gunes)

ORCID 0009-0001-8938-3663 (Z. Wang); 0000-0003-2157-0080 (Y. Liu); 0000-0003-2407-3012 (H. Gunes)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Recent advances in graph neural networks (GNNs), such as Multi-dimensional Edge Feature-based AU Relation Graph for AU (ME-GraphAU) [20] and Graph Relation Network (GRU) [21], have demonstrated promising AU prediction performance on the DISFA dataset [22] and BP4D dataset [23]. These methods use CNN backbones like ResNet [24] or VGG [25] to learn image features for each AU, then construct a relational graph based on the AU features. Each node in such graphs represents an AU, and each edge represents the relationship between a pair of AUs [26]. Through message propagation in GNN layers, the output AU features can capture individual AU information from neighbors as well as structural information [27]. Such graph-based AU features can then be aggregated to build a full-face representation informed by AUs and their relationships for downstream tasks [28], inspiring us to incorporate graph-based AU detection into pain intensity estimation.

Motivated by the above, we introduce **GraphAU-Pain** for accurate and interpretable pain intensity estimation and summarize our three contributions as follows:

- **GraphAU-Pain Model.** We propose a novel graph-based framework that transforms AU detection into pain estimation by modeling AU relationships as a dynamic graph structure, enabling more expressive and interpretable pain assessment compared to traditional image-driven approaches.
- **Cross-Dataset Transfer with Relabeling.** We introduce a novel transfer learning strategy that leverages DISFA-pretrained weights to address UNBC’s limited training data. By creating a hybrid UNBC+ dataset that combines original annotations with predicted labels for missing AUs, we enable effective knowledge transfer while preserving dataset integrity, significantly improving model performance in AU occurrence prediction.
- **High Performance & Interpretability.** Comprehensive experiments validate GraphAU-Pain for pain estimation, outperforming GLA-CNN [18] (current state-of-the-art work in this area) and demonstrating improved interpretability via explicit AU modeling.

## 2. Related Work

### 2.1. Facial Action Unit Detection

FACS [13] categorizes facial expressions into 44 fundamental components known as AUs, each corresponding to specific muscle movements with intensity ratings from 0 to 5. For instance, AU1 represents the “Inner Brow Raiser” and AU2 the “Outer Brow Raiser.” These AUs provide a systematic method for analyzing facial expressions, including pain and other affective states. Early approaches to AU detection relied on traditional machine learning methods, with OpenFace [29] being a widely used open-source tool based on SVMs [30]. Deep learning methods like Deep Region and Multi-label Learning [31] established early benchmarks but failed to model AU interdependencies. Later attention-based methods [32, 33] improved performance significantly, though their pixel-wise approach faced limitations in capturing subtle facial changes and generalizing across diverse racial groups.

Graph Neural Networks (GNNs) have emerged as a powerful architecture for modeling complex dependencies between facial landmarks or AUs. A GNN layer passes information between nodes through edges, allowing nodes to learn their neighborhood information. This makes GNNs essential for learning relational representations, as demonstrated in fields like knowledge graph construction [34] and recommender systems [35]. In facial AU analysis, some studies introduced prior AU co-occurrence knowledge via Graph Convolutional Networks [36, 37], while others like Graph Relation Network (GRN) [21] explored knowledge-free approaches. GRN constructs a fully connected directed graph with image features as nodes and uses attention-based edge functions, achieving MAE of 0.7 on BP4D and 0.2 on DISFA. Unlike GRN, which fully connects all nodes, ME-GraphAU [20] introduces a novel approach by using a CNN backbone to extract features for each node that represents an AU and establishing connections between nodes based on feature similarity. This architecture effectively captures facial feature relationships by modeling AUs as nodes and their interactions as edges, achieving strong performance with F1 scores of 65.5% on BP4D and 63.1% on DISFA. Despite these advances,

direct adaptation of AU detection models to pain estimation tasks remains challenging due to two key limitations. First, pain estimation datasets are typically much smaller than those used for AU detection. Second, these datasets exhibit a significant class imbalance in AU occurrences. Our GraphAU-Pain framework addresses both challenges through a novel transfer learning strategy, as detailed in Sec. 3.4.

## 2.2. Pain Prediction Based on Facial Expressions

Early approaches relied on feature extraction and classification techniques like KNN [16] and Random Forest [38]. These handcrafted feature approaches were limited by variations in head pose, lighting, and spontaneous expressions.

Recent deep learning approaches have evolved from traditional CNNs [39, 40] to more sophisticated architectures that incorporate AU features. While CNNs provide a foundation through pixel-level analysis, their lack of physiological knowledge limits both interpretability and generalizability. This limitation has driven the development of more advanced approaches, such as LSTM-based continuous pain monitoring [41]. However, this promising work faces practical constraints due to its reliance on a private dataset [42]. Similarly, while AU-based pain prediction has shown potential on BP4D [43], its binary classification approach may not capture the nuanced spectrum of natural pain experiences.

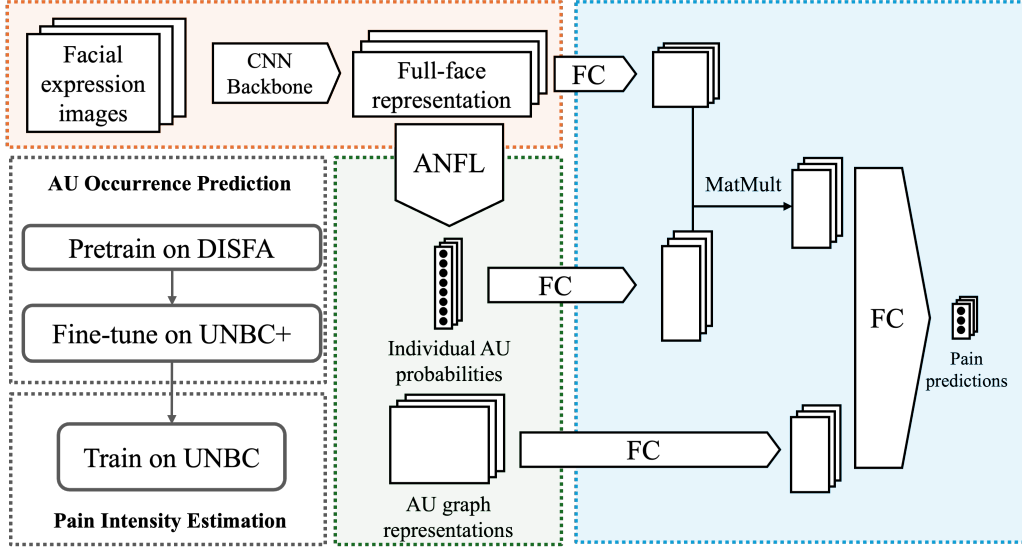
The state-of-the-art GLA-CNN [18] represents a significant advancement by combining CNNs with attention mechanisms to analyze facial pain and AU relationships. On the UNBC dataset, it achieves 36.2% F1-score and 56.5% accuracy. While these metrics show improvement, they remain insufficient for clinical applications. The model’s fine-grained category scheme for pain levels fails to account for PSPI’s sensitivity to AU intensity. This limitation becomes apparent when considering cases such as intense AU4 (brow lowering) can occur without pain [44], leading to confusion between its *Weak Pain* ( $PSPI = 0$ ), *weak Pain* ( $PSPI = 1$ ), and *Mild Pain* ( $PSPI = 2$ ) categories. Furthermore, the black-box nature of the design limits clinical utility by obscuring how the estimated pain intensity is derived. These challenges highlight the need for more reliable and interpretable pain detection models, motivating our development of GraphAU-Pain. Our approach explicitly incorporates AUs through a graph-based modeling framework, providing enhanced interpretability through transparent AU-pain intensity relationships and improved accuracy via comprehensive modeling of AU interdependencies.

## 3. Material and Methods

### 3.1. Implementation Overview

The GraphAU-Pain model is designed to estimate pain intensity based on facial image data, and its training involves two key steps. First, the full-face and AU representation learning modules are trained for *AU occurrence prediction*. These modules are adapted from the AU Relationship-aware Node Feature Learning (ANFL) component of ME-GraphAU [20]. Because directly training ANFL for AU occurrence prediction on the UNBC dataset only yielded a 20% average F1-score, we developed a cross-dataset transfer learning strategy to improve performance. This strategy transfers AU prediction capabilities from the DISFA dataset model trained by Luo et al. [20] to the UNBC dataset. To address differences in AU labels between datasets, we created a relabeled UNBC+ dataset to align the AU annotations and used undersampling to address AU imbalance. Second, with the weights trained for AU prediction, GraphAU-Pain is then trained for *pain intensity estimation* with the full UNBC dataset. Facial images are first processed through the CNN backbone to extract pixel-wise full-face representations. These representations are then transformed into graph-based structures to learn AU-specific features. Finally, the full-face and AU-based features are combined to estimate pain intensity. The performance of the model is evaluated using two metrics: average F1-score and accuracy.

As illustrated in Fig. 1, the GraphAU-Pain model comprises three sequential modules: *Full-face Representation Learning* that extracts high-level facial features through pixel-based global representations, *AU Representation Learning* that captures both local and global AU using graph-based representations, and a *Pain Intensity Classifier* that maps the learned features to specific pain intensity levels.



**Figure 1:** Illustration of the GraphAU-Pain model architecture. A global CNN backbone (orange) extracts a full-face representation, which is split into three feature branches by the AU representation learning module (green) and processed by the pain level classification module (blue). *ANFL* denotes the AU Relationship-aware Node Feature Learning component, *FC* the fully-connected layer, and *MatMult* matrix multiplication. GraphAU-Pain is trained by first initializing weights of the representation learning modules with the AU occurrence prediction task and then train the full model for the pain intensity estimation task.

### 3.2. Model

**Full-face Representation Learning** GraphAU-Pain uses a ResNet-50 backbone for extracting a full-face representation from an input image. By inputting a face image  $\mathbf{x} \in \mathbb{R}^{172 \times 172 \times 3}$  to the backbone, we obtain  $\mathbf{h}_b \in \mathbb{R}^{36 \times 2048}$ , which represents 36 facial image features of length 2048, each corresponding to a position in the image.

**AU Representation Learning** After acquiring the backbone feature  $\mathbf{h}_b$ , the AU representation module learns two distinct embeddings:  $\mathbf{h}_a \in \mathbb{R}^{d_{AU}}$  ( $d_{AU} = 512$ ) encoding individual AU occurrences, and  $\mathbf{h}_g \in \mathbb{R}^{d_{AU}}$  encoding the complete AU relational graph structure. The module first transforms  $\mathbf{h}_b$  through  $n_{AU}$  fully connected layers to generate initial AU representations  $\mathbf{H}_a \in \mathbb{R}^{n_{AU} \times d_{AU}}$ , where each row represents one AU. These representations serve as node features in a graph where each node connects to its  $K = 3$  most similar nodes based on dot-product similarity. The graph structure is then processed through a graph convolutional layer:

$$\mathbf{H}'_a = \text{ReLU}\left(\mathbf{H}_a + \text{BN}\left(\mathbf{A}^\top \text{FC}_1(\mathbf{H}_a) + \text{FC}_2(\mathbf{H}_a)\right)\right) \in \mathbb{R}^{n_{AU} \times d_{AU}}, \quad (1)$$

where BN denotes batch normalization, FC represents fully connected layers, and  $\mathbf{A}$  is the normalized adjacency matrix. The graph representation  $\mathbf{h}_g \in \mathbb{R}^{d_{AU}}$  is obtained through global sum pooling of the node embeddings. Although we could also add an edge update module here [20], we omitted it to avoid overfitting on the UNBC dataset. For each AU, its occurrence probability  $p_i$  is computed as the cosine similarity between its representation  $h_{a,i}$  and a learnable vector  $s_i$ :

$$p_i = \frac{\text{ReLU}(h_{a,i})^\top \text{ReLU}(s_i)}{\|\text{ReLU}(h_{a,i})\|_2 \|\text{ReLU}(s_i)\|_2}, \quad (2)$$

where  $\|\cdot\|$  represents the L2 norm. This AU occurrence prediction can serve as either a pretraining task or an auxiliary training objective alongside the primary pain intensity estimation task. In this work, we use it as a pretraining task on an undersampled dataset to enforce the AU representation module to focus on minority classes. After obtaining these three features, three fully connected layers with ReLU

activation map each of them to a common dimension of 36, producing  $\mathbf{h}'_b$ ,  $\mathbf{h}'_a$ , and  $\mathbf{h}'_g$ . (For  $\mathbf{h}_b$ , FC is applied row-wise.) Finally, a feature-infusing step on  $\mathbf{h}'_a$  and  $\mathbf{h}'_b$  is performed:

$$\mathbf{h}_{ab} = \text{ReLU}(\mathbf{h}'_a{}^\top \mathbf{h}'_b) \in \mathbb{R}^{36}. \quad (3)$$

**Representation Classifier** The final pain intensity classification is get by concatenating the interaction scalar  $\mathbf{h}_{ab}$  with the  $\mathbf{h}'_g$  feature, then passing the result to an FC layer:

$$\hat{y} = \mathbf{W}_y [\mathbf{h}_{ab} \parallel \mathbf{h}'_g] + \mathbf{b}_y \in \mathbb{R}^{d_{\text{pain}}}, \quad (4)$$

where  $d_{\text{pain}} = 3$  for the one-hot encoding of the three-level pain intensity classification used in this work.

### 3.3. Loss Function

With over 80% of the frames demonstrating no expression of pain, the class imbalance in the UNBC dataset poses a significant challenge for deep learning models. To minimize this, we employ a weighted cross-entropy loss to prioritize underrepresented classes. It is calculated by

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C w_j y_{i,j} \log \max(p_{i,j}, \epsilon), \quad (5)$$

where  $N$  is the number of samples,  $C$  is the number of classes,  $p_{i,j}$  are the softmax output probabilities for the  $i$ -th sample belonging to the  $j$ -th class,  $y_{i,j}$  is a binary indicator (1 if the  $i$ -th sample belongs to the  $j$ -th class, otherwise 0),  $w_j$  are class weights, and  $\epsilon = 1e^{-8}$  prevents  $\log(0)$ . The weight of a class  $c_j$  is calculated by

$$w_j = C \cdot \frac{1/\text{occurrence\_rate}(c_j)}{\sum_{k=1}^C 1/\text{occurrence\_rate}(c_k)}. \quad (6)$$

In this work, the class weights are 0.07 for No Pain, 0.33 for Mild Pain, and 2.6 for Obvious Pain.

### 3.4. Transfer Learning

Our preliminary experimental results showed that directly training GraphAU-Pain on UNBC yielded unsatisfactory results, with 30–40% F1-score and 60–80% accuracy after trying several settings. To tackle this problem and improve AU prediction performance through transfer learning, in this work we initialized the weights of the full-face and AU representation modules with the weights pretrained on DISFA provided by Luo et al. [20]. We chose the DISFA dataset [22] for pretraining because it is three times larger than UNBC and provides high-quality AU annotations that align well with UNBC’s AU annotation scheme—sharing six out of eight AUs with UNBC and three of them are used in PSPI calculation—making this dataset suitable for transfer learning. However, since DISFA includes two additional AUs (AU1 and AU2) not present in UNBC’s original annotations, for fine-tuning the pretrained weights, we need to label these additional AUs for UNBC to ensure complete AU coverage. To do this, we pass UNBC’s facial images through the pretrained representation learning modules to predict all eight AU labels. We then create a hybrid dataset (UNBC+) by keeping UNBC’s original annotations for the six overlapping AUs while using the predicted values for AU1 and AU2. This relabeling process ensures that our model can learn from a complete set of AU activations while maintaining the reliability of UNBC’s original annotations where available.



**Table 1**

F1-scores and accuracies achieved for each AU by the fine-tuned representation learning module. The average F1-score and accuracy are calculated by averaging the F1-scores and accuracies of all AUs. While accuracy remains high across all AUs, F1-scores are notably lower for AU9 and AU26, likely due to class imbalance.

Metric	AU1	AU2	AU4	AU6	AU9	AU12	AU25	AU26	Avg
F1	63.09	88.65	70.74	88.46	31.94	92.12	63.93	6.63	63.20
Acc.	90.38	79.88	92.04	79.62	91.43	86.09	83.11	85.21	85.98

## 4. Experiments

### 4.1. Datasets and Labels

GraphAU-Pain was trained on the UNBC-McMaster Shoulder Pain Expression Archive Database [15]. The dataset contains 48,398 colored frames from 25 participants with shoulder problems, showing facial expressions during pain-inducing actions. The faces were detected with OpenCV’s haarcascade\_frontalface\_default classifier and cropped to  $172 \times 172$ . Each frame has 10 AU intensities (0–5) and PSPI scores (0–16), calculated as  $PSPI = AU4 + \max(AU6, AU7) + \max(AU9, AU10) + AU43$  [14]. PSPI pain intensity is categorized into ordinal levels: No Pain ( $PSPI = 0$ ), Mild Pain ( $PSPI \in [1, 4]$ ), and Obvious Pain ( $PSPI \geq 5$ ). The categories are distributed in a skewed way, with each category respectively consisting approximately 82%, 15%, and 3% of the UNBC dataset. This categorization is more clinically meaningful and interpretable compared to the method used by Wu et al. [18], as it mitigates the high sensitivity of PSPI scores to AU variations. Moreover, we replaced AU intensity with occurrence by capping the AU score at one.

### 4.2. Training and Evaluation Details

To prepare AU occurrence prediction before pain intensity estimation training, the supervised fine-tuning (SFT) process of GraphAU-Pain’s AU representation learning module was performed on the ANFL component with weights pretrained on DISFA for 20 epochs provided by Luo et al. [20]. Our relabeled UNBC+ dataset was used in the SFT process and undersampled to address data imbalance. The undersampling process involved randomly removing approximately 90% of facial images with  $PSPI = 0$  and excluding facial images without active AUs. The full list of frames included in this subset is made available in the code repository. We use all  $n_{AU} = 8$  AU labels in the UNBC+ dataset as listed in Table 1. We trained the module through SFT for 17 epochs with a learning rate of  $1e^{-5}$ , a batch size of 16, and an Adam optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and a weight decay of  $5e^{-4}$ . The AU representation learning module achieved remarkable results in AU detection, compared to state-of-the-art results, as detailed in Table 1. The training for pain intensity estimation used the same hyperparameters as the SFT process but was performed on the full original UNBC dataset.

The GraphAU-Pain model was trained using a learning rate of  $1e^{-4}$ , a batch size of 64, and an Adam optimizer with the same hyperparameters used in SFT. The representation learning module is set to connect edges between an AU node and its 3 most similar AU nodes. The weights learned through SFT on ANFL were used to initialize the ResNet backbone and the representation learning module of GraphAU-Pain. The model was trained on the full UNBC dataset for 8 epochs on an NVIDIA GeForce RTX 4070 GPU (8 GB) and an Intel i7-13900H CPU, with an estimated training time of about 3 minutes per epoch. The evaluation uses the same metrics as the state-of-the-art method GLA-CNN [18]: accuracy, average F1, average recall, and average precision, where all average values are unweighted.

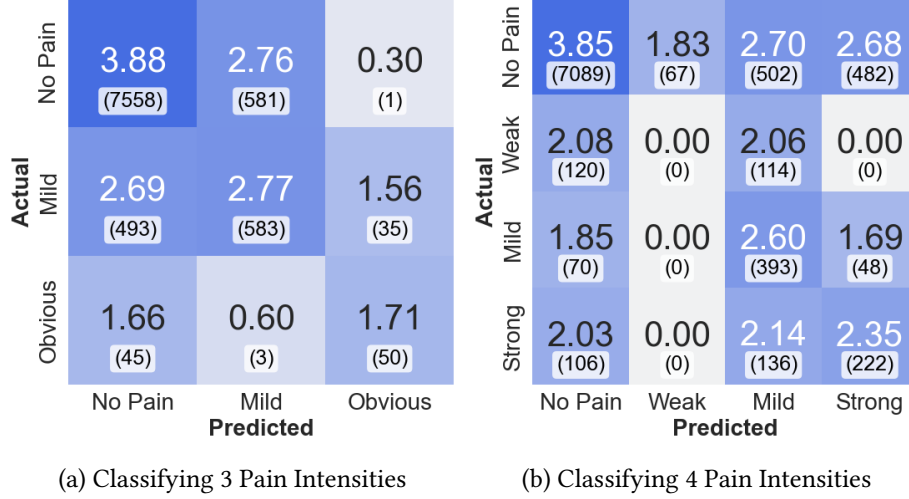
### 4.3. Model Performance

Overall, GraphAU-Pain achieves a commendable average F1-score of 66.21% and a high accuracy of 87.61%. Table 2 details the per-class results, showing strong performance for the No Pain category, which has an F1-score of 93.10%. However, performance declines for the Mild and Obvious categories,

**Table 2**

GraphAU-Pain’s performance statistics. Our model achieves strong performance for No Pain detection but lower and similar performance for Mild and Obvious pain categories, likely due to data imbalance. The overall F1/Recall/Precision/Accuracy are calculated by averaging the F1/Recall/Precision/Accuracy of all categories.

Metric	No Pain	Mild	Obvious	Overall
F1-score	93.10	51.19	54.35	66.21
Recall	92.85	52.48	51.02	65.45
Precision	93.35	49.96	58.14	67.15
Accuracy	-	-	-	87.61



**Figure 2:** Confusion matrix of GraphAU-Pain. The numbers are transformed in  $\log_{10}$  scale for better visualization; the original number is shown below the scaled one. The 3-category classification (left) is introduced in this work to mitigate the excessive sensitivity of the 4-category system (right) proposed by Wu et al. [18]. The 4-category system’s additional “Weak Pain” category ( $PSPI=1$ ) proves too sensitive to AU intensity variations and lacks clinical significance when separated from our “Mild Pain” category. Therefore, the lower performance in recognizing “Weak Pain” reflects the metric’s sensitivity to minor AU variations rather than model limitations.

reflecting F1-scores of 51.19% and 54.35%, respectively. This performance gap between No Pain and the Mild/Obvious categories is largely attributable to the dataset’s pronounced class imbalance. The log-scaled confusion matrix in Figure 2 further illustrates the distribution of predictions. While No Pain dominates the diagonal, indicating high accuracy there, some off-diagonal misclassifications occur between Mild and Obvious, and there is a noticeable bias toward predicting No Pain. Consequently, while the model is robust in detecting No Pain, additional strategies are needed to better distinguish between higher pain intensities.

#### 4.4. Comparison to SOTA

To the best of our knowledge, GLA-CNN [18] is the only other method that uses AUs for pain-intensity estimation on UNBC while focusing on cross-sectional facial image frames. No additional methods apply exactly the same categorization scheme, so we compare GraphAU-Pain with GLA-CNN and other models reported in [18]. To align labels, we reclassify pain intensities into four categories: No Pain ( $PSPI = 0$ ), Weak Pain ( $PSPI = 1$ ), Mild Pain ( $PSPI = 2$ ), and Strong Pain ( $PSPI \geq 3$ ). By altering only the model’s final layer from three to four outputs and keeping other settings unchanged, GraphAU-Pain shows substantial gains in both accuracy and average F1-score, as shown in Table 3 and Figure 2b. Note that GLA-CNN and the other compared models were trained and evaluated on an undersampled subset of UNBC [18] to deal with class imbalance, whereas GraphAU-Pain is trained on the full dataset. Therefore, their published performance might be higher than what would have been

**Table 3**

Comparison of GraphAU-Pain with other models on the UNBC dataset. We compare our results with those reported by Wu et al. [18], and we retrained our model with the 4-category system for fair comparison.

Model	Accuracy	F1-score	Recall	Precision
GraphAU-Pain (Ours)	82.40	43.83	52.96	39.96
GLA-CNN [18]	56.45	36.52	34.08	43.23
MA-NET [45]	55.16	-	-	-
Swin Transformer [46]	54.40	-	-	-
Attention CNN [47]	51.10	-	-	-
LBP [48]	34.70	-	-	-
Head Analysis [49]	28.10	-	-	-
LPQ [48]	23.40	-	-	-

**Table 4**

Ablation analysis for GraphAU-Pain. *w/o graph rep.* removes the graph representation from the model. *w/o GNN* removes the GNN layer. *Only ResNet* relies solely on CNNs. The highest value for each category and metric is **bold**; the second highest is underlined.

Model	F1				Precision				Recall			
	No Pain	Mild	Obvious	Mean	No Pain	Mild	Obvious	Mean	No Pain	Mild	Obvious	Mean
Full	<b>93.1</b>	<b>51.2</b>	<u>54.3</u>	<b>66.2</b>	93.4	<b>50.0</b>	<u>58.1</u>	<b>67.2</b>	<b>92.9</b>	52.5	<b>51.0</b>	<u>65.4</u>
w/o graph rep.	<u>87.7</u>	<u>46.4</u>	<b>55.2</b>	<u>63.1</u>	<b>95.9</b>	<u>33.9</u>	<b>60.2</b>	<u>63.4</u>	<u>80.8</u>	<b>73.4</b>	<b>51.0</b>	<b>68.4</b>
w/o GNN	86.4	30.7	4.0	40.3	<u>93.9</u>	23.8	2.5	40.1	79.9	43.2	10.2	44.4
Only ResNet	70.3	31.0	4.2	35.2	92.8	20.1	2.5	38.5	56.5	<u>68.3</u>	15.3	46.7

achieved on the full dataset.

#### 4.5. Ablation Analysis

The ablation analysis in Table 4 underscores the critical role of graph representation and GNN in the GraphAU-Pain model. Removing the graph representation (*w/o graph rep.*) reduces the mean F1-score from 66.2% to 63.1%, mainly due to the performance drops in No Pain and Mild Pain, highlighting the value of graph modeling for capturing AU relationships. Similarly, removing the GNN layer (*w/o GNN*) causes a significant drop to 40.3%, emphasizing the importance of graph-based interactions in learning AU features. The simplest setup (*Only ResNet*), relying solely on CNNs, achieves the lowest mean F1-score of 35.2%, demonstrating that CNNs alone fail to effectively model AU correlations for pain estimation. These results affirm the superiority of graph-based learning methods for pain estimation.

#### 4.6. Discussion and Future Work

GraphAU-Pain demonstrates significant potential for advancing automated pain assessment through several key contributions. By leveraging graph-based learning to model AU relationships, our approach achieves superior performance compared to existing methods, with an accuracy of 87.61% in the clinically meaningful three-category classification system. The model’s strong performance in detecting No Pain (93.10% F1-score) makes it particularly valuable for initial screening applications. Furthermore, the AU representation learning module provides a more interpretable framework for understanding how different facial expressions contribute to pain assessment. This could lead to more reliable and explainable automated pain monitoring systems in clinical settings, potentially reducing the burden on healthcare providers and improving patient care through continuous, objective pain assessment.

While the proposed method shows promising results in pain estimation, there remains room for improvement. Firstly, since PSPI only captures facial expressions, it may not reflect true subjective pain [50]. Future research could focus on finding alternative pain indicators. Secondly, aligning UNBC



with DISFA through UNBC+ removes three pain-related AUs and adds noise through predicted labels, potentially impacting performance. A promising direction is to design AU occurrence prediction models specifically for pain-oriented datasets like UNBC. Lastly, while the AU occurrence-based representation learning module provides satisfactory representation, AU intensity-based approaches (e.g., GRN [21]) could also be explored since AU intensity better relates to pain intensity. However, this direction may require more complex models and additional training data to mitigate the impact of data imbalance.

## 5. Conclusion

This paper presents GraphAU-Pain, a GNN-based model combining graph-based AU features with full-face representation for pain prediction. It surpasses the state-of-the-art methods while enabling AU-informed pain estimation for clinical transparency. GraphAU-Pain addresses challenges like limited data and class imbalance in the UNBC dataset through a novel transfer learning strategy. Key contributions include improved pain classification benchmarks, better interpretability through AU-based representations, and critical baselines for future AU-based pain estimation. Overall, this work demonstrates the potential of GNNs for accurate, clinically viable pain estimation solutions.

## Acknowledgments

Y. Liu's work was supported in part by the Finnish Cultural Foundation for North Ostrobothnia Regional Fund under Grant 60231712, and in part by the Instrumentarium Foundation under Grant 240016. Z. Wang was supported by the Churchill College Postgraduate Academic Travel Grant PAT0062 for conference participation.

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

- [1] T. Hassan, D. Seus, J. Wollenberg, K. Weitz, M. Kunz, S. Lautenbacher, J. U. Garbas, U. Schmid, Automatic Detection of Pain from Facial Expressions: A Survey, 2021. doi:10.1109/TPAMI.2019.2958341.
- [2] E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Automatic vs. Human recognition of pain intensity from facial expression on the x-ite pain database, *Sensors* 21 (2021). doi:10.3390/s21093273.
- [3] P. Rodriguez, G. Cucurull, J. Gonzalez, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, F. X. Roca, Deep Pain: Exploiting Long Short-Term Memory Networks for Facial Expression Classification, *IEEE Transactions on Cybernetics* 52 (2022) 3314–3324. URL: <https://ieeexplore.ieee.org/document/7849133/>. doi:10.1109/TCYB.2017.2662199.
- [4] S. El Morabit, A. Rivenq, M.-E.-n. Zighem, A. Hadid, A. Ouahabi, A. Taleb-Ahmed, Automatic Pain Estimation from Facial Expressions: A Comparative Analysis Using Off-the-Shelf CNN Architectures, *Electronics* 10 (2021) 1926. doi:10.3390/electronics10161926.
- [5] R. Yang, X. Hong, J. Peng, X. Feng, G. Zhao, Incorporating high-level and low-level cues for pain intensity estimation, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 3495–3500. doi:10.1109/ICPR.2018.8545244.
- [6] P. D. Barua, N. Baygin, S. Dogan, M. Baygin, N. Arunkumar, H. Fujita, T. Tuncer, R.-S. Tan, E. Palmer, M. M. B. Azizan, N. A. Kadri, U. R. Acharya, Automated detection of pain levels using deep feature extraction from shutter blinds-based dynamic-sized horizontal patches with facial images, *Scientific Reports* 12 (2022) 17297. doi:10.1038/s41598-022-21380-4.

- [7] H. Liu, H. Xu, J. Qiu, S. Wu, M. Liu, Hierarchical global and local transformer for pain estimation with facial expression videos, *Pattern Analysis and Applications* 27 (2024) 85. doi:10.1007/s10044-024-01302-y.
- [8] H. Xu, M. Liu, A Deep Attention Transformer Network for Pain Estimation with Facial Expression Video, 2021, pp. 112–119. doi:10.1007/978-3-030-86608-2\_13.
- [9] N. Ben Aoun, A Review of Automatic Pain Assessment from Facial Information Using Machine Learning, 2024. doi:10.3390/technologies12060092.
- [10] Z. Chen, R. Ansari, D. Wilkie, Automated Pain Detection from Facial Expressions using FACS: A Review (2018).
- [11] M. Feighelstein, L. Henze, S. Meller, I. Shimshoni, B. Hermoni, M. Berko, F. Twele, A. Schütter, N. Dorn, S. Kästner, L. Finka, S. P. Luna, D. S. Mills, H. A. Volk, A. Zamansky, Explainable automated pain recognition in cats, *Scientific Reports* 13 (2023). doi:10.1038/s41598-023-35846-6.
- [12] Y. Huang, L. Qing, S. Xu, L. Wang, Y. Peng, HybNet: a hybrid network structure for pain intensity estimation, *Visual Computer* 38 (2022) 871–882. doi:10.1007/s00371-021-02056-y.
- [13] P. Ekman, W. V. Friesen, Facial Action Coding System, 1978. URL: <https://doi.apa.org/doi/10.1037/t27734-000>. doi:10.1037/t27734-000.
- [14] K. M. Prkachin, P. E. Solomon, The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain, *Pain* 139 (2008) 267–274. URL: <https://journals.lww.com/00006396-200810150-00006>. doi:10.1016/j.pain.2008.04.010.
- [15] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, I. Matthews, Painful data: The UNBC-McMaster shoulder pain expression archive database, in: *Face and Gesture 2011*, IEEE, 2011, pp. 57–64. URL: <http://ieeexplore.ieee.org/document/5771462/>. doi:10.1109/FG.2011.5771462.
- [16] Z. Zafar, N. A. Khan, Pain Intensity Evaluation through Facial Action Units, in: *2014 22nd International Conference on Pattern Recognition*, IEEE, 2014, pp. 4696–4701. URL: <http://ieeexplore.ieee.org/document/6977516/>. doi:10.1109/ICPR.2014.803.
- [17] W. Guo, Z. Xu, Z. Guo, L. Mao, Y. Hou, Z. Huang, Pain Assessment Using Facial Action Units and Bayesian Network, in: *Chinese Control Conference, CCC*, volume 2021-July, IEEE Computer Society, 2021, pp. 4665–4670. doi:10.23919/CCC52363.2021.9550304.
- [18] J. Wu, Y. Shi, S. Yan, H.-M. Yan, Global-local combined features to detect pain intensity from facial expression images with attention mechanism, *Journal of Electronic Science and Technology* 22 (2024) 100260. URL: <https://www.sciencedirect.com/science/article/pii/S1674862X24000284>. doi:https://doi.org/10.1016/j.jnlest.2024.100260.
- [19] Z. Chen, R. Ansari, D. J. Wilkie, Learning Pain from Action Unit Combinations: A Weakly Supervised Approach via Multiple Instance Learning, *IEEE Transactions on Affective Computing* 13 (2022) 135–146. URL: <https://ieeexplore.ieee.org/document/8887215/>. doi:10.1109/TAFFC.2019.2949314.
- [20] C. Luo, S. Song, W. Xie, L. Shen, H. Gunes, Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition, in: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, California, 2022, pp. 1239–1246. URL: <https://www.ijcai.org/proceedings/2022/173>. doi:10.24963/ijcai.2022/173.
- [21] C. Wang, F. Jiang, R. Shen, Facial action units intensity estimation via graph relation network, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12533 LNCS, Springer Science and Business Media Deutschland GmbH, 2020, pp. 345–356. doi:10.1007/978-3-030-63833-7\_29.
- [22] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, J. F. Cohn, DISFA: A Spontaneous Facial Action Intensity Database, *IEEE Transactions on Affective Computing* 4 (2013) 151–160. URL: <http://ieeexplore.ieee.org/document/6475933/>. doi:10.1109/T-AFFC.2013.4.
- [23] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, J. M. Girard, BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database, *Image and Vision Computing* 32 (2014) 692–706. doi:10.1016/J.IMAVIS.2014.06.002.
- [24] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *2016 IEEE*

- Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.
- [25] S. Liu, W. Deng, Very deep convolutional neural network based image classification using small training sample size, in: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015, pp. 730–734. doi:10.1109/ACPR.2015.7486599.
  - [26] Y. Liu, X. Zhang, Y. Li, J. Zhou, X. Li, G. Zhao, Graph-based facial affect analysis: A review, *IEEE Transactions on Affective Computing* 14 (2022) 2657–2677.
  - [27] Y. Liu, X. Zhang, J. Kauttonen, G. Zhao, Uncertain label correction via auxiliary action unit graphs for facial expression recognition, in: 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 777–783.
  - [28] Y. Liu, X. Zhang, J. Kauttonen, G. Zhao, Uncertain facial expression recognition via multi-task assisted correction, *IEEE Transactions on Multimedia* 26 (2024) 2531–2543.
  - [29] T. Baltrusaitis, A. Zadeh, Y. C. Lim, L. P. Morency, OpenFace 2.0: Facial behavior analysis toolkit, in: Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, 2018. doi:10.1109/FG.2018.00019.
  - [30] T. Baltrusaitis, M. Mahmoud, P. Robinson, Cross-dataset learning and person-specific normalisation for automatic Action Unit detection, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), volume 2015-January, IEEE, 2015, pp. 1–6. URL: <http://ieeexplore.ieee.org/document/7284869/>. doi:10.1109/FG.2015.7284869.
  - [31] K. Zhao, W.-S. Chu, H. Zhang, Deep Region and Multi-label Learning for Facial Action Unit Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016, pp. 3391–3399. URL: <http://ieeexplore.ieee.org/document/7780738/>. doi:10.1109/CVPR.2016.369.
  - [32] Z. Shao, Z. Liu, J. Cai, Y. Wu, L. Ma, Facial Action Unit Detection Using Attention and Relation Learning, *IEEE Transactions on Affective Computing* 13 (2022) 1274–1289. URL: <https://ieeexplore.ieee.org/document/8880512/>. doi:10.1109/TAFFC.2019.2948635.
  - [33] G. Miriam Jacob, B. Stenger, Facial Action Unit Detection With Transformers, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2021, pp. 7676–7685. URL: <https://ieeexplore.ieee.org/document/9577264/>. doi:10.1109/CVPR46437.2021.00759.
  - [34] K. Zuo, Y. Jiang, F. Mo, P. Lio, KG4Diagnosis: A hierarchical multi-agent LLM framework with knowledge graph enhancement for medical diagnosis, in: Proceedings of The First AAAI Bridge Program on AI for Medicine and Healthcare, ????
  - [35] S. Wu, F. Sun, W. Zhang, X. Xie, B. Cui, Graph Neural Networks in Recommender Systems: A Survey, *ACM Computing Surveys* 55 (2022). doi:10.1145/3535101.
  - [36] J. He, X. Yu, B. Sun, L. Yu, Facial expression and action unit recognition augmented by their dependencies on graph convolutional networks, *Journal on Multimodal User Interfaces* 15 (2021) 429–440. doi:10.1007/s12193-020-00363-7.
  - [37] X. Jia, S. Xu, Y. Zhou, L. Wang, W. Li, A novel dual-channel graph convolutional neural network for facial action unit recognition, *Pattern Recognition Letters* 166 (2023) 61–68. doi:10.1016/j.patrec.2023.01.001.
  - [38] Xing Zhang, Lijun Yin, J. F. Cohn, Three dimensional binary edge feature representation for pain expression analysis, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2015, pp. 1–7. URL: <http://ieeexplore.ieee.org/document/7163107/>. doi:10.1109/FG.2015.7163107.
  - [39] F. Wang, X. Xiang, C. Liu, T. D. Tran, A. Reiter, G. D. Hager, H. Quon, J. Cheng, A. L. Yuille, Regularizing face verification nets for pain intensity regression, in: 2017 IEEE International Conference on Image Processing (ICIP), volume 2017-September, IEEE, 2017, pp. 1087–1091. URL: <http://ieeexplore.ieee.org/document/8296449/>. doi:10.1109/ICIP.2017.8296449.
  - [40] J. Zhou, X. Hong, F. Su, G. Zhao, Recurrent Convolutional Neural Network Regression for Continuous Pain Intensity Estimation in Video, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2016, pp. 1535–1543. URL: <http://ieeexplore.ieee.org/document/7789681/>. doi:10.1109/CVPRW.2016.191.

- [41] E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, S. Walter, Classification networks for continuous automatic pain intensity monitoring in video using facial expression on the X-ITE Pain Database, *Journal of Visual Communication and Image Representation* 91 (2023). doi:10.1016/j.jvcir.2022.103743.
- [42] S. Gruss, M. Geiger, P. Werner, O. Wilhelm, H. C. Traue, A. Al-Hamadi, S. Walter, Multi-Modal Signals for Analyzing Pain Responses to Thermal and Electrical Stimuli, *Journal of Visualized Experiments* 2019 (2019). URL: <https://app.jove.com/t/59057>. doi:10.3791/59057.
- [43] K. Feghoul, M. Bouazizi, D. Santana, D. Santana Maia, Facial Action Unit Detection using 3D Face Land-marks for Pain Detection, Technical Report, 2023. URL: <https://hal.science/hal-04320516v1>.
- [44] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, S. Gruss, H. C. Traue, Automatic Pain Assessment with Facial Activity Descriptors, *IEEE Transactions on Affective Computing* 8 (2017). doi:10.1109/TAFFC.2016.2537327.
- [45] Z. Zhao, Q. Liu, S. Wang, Learning deep global multi-scale and local attention features for facial expression recognition in the wild, *IEEE Transactions on Image Processing* 30 (2021) 6544–6556. doi:10.1109/TIP.2021.3093397.
- [46] Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, H. Hu, Video swin transformer, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 3192–3201. doi:10.1109/CVPR52688.2022.00320.
- [47] X. Xin, X. Lin, S. Yang, X. Zheng, Pain intensity estimation based on a spatial transformation and attention cnn, *PLOS ONE* 15 (2020) 1–15. URL: <https://doi.org/10.1371/journal.pone.0232412>. doi:10.1371/journal.pone.0232412.
- [48] R. Yang, S. Tong, M. Bordallo, E. Boutellaa, J. Peng, X. Feng, A. Hadid, On pain assessment from facial videos using spatio-temporal local descriptors, in: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016, pp. 1–6. doi:10.1109/IPTA.2016.7820930.
- [49] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. C. Traue, P. Werner, A. Al-Hamadi, S. Crawcour, A. O. Andrade, G. Moreira da Silva, The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system, in: 2013 IEEE International Conference on Cybernetics (CYBCO), 2013, pp. 128–131. doi:10.1109/CYBConf.2013.6617456.
- [50] G. D. De Sario, C. R. Haider, K. C. Maita, R. A. Torres-Guzman, O. S. Emam, F. R. Avila, J. P. Garcia, S. Borna, C. J. McLeod, C. J. Bruce, R. E. Carter, A. J. Forte, Using AI to Detect Pain through Facial Expressions: A Review, 2023. doi:10.3390/bioengineering10050548.