Original Research Article

A Multimodal Approach To Pain Detection Using Facial Action Units and Convolutional Neural Networks

Aaryan Verma[1], Anvitha Kakarlapudi[1], Tamiika Hurst-Darby[2]

[1]William Mason High School, 6100 Mason Montgomery Rd, Mason, OH 45040, United States;
[2]Independent researcher, Mason, OH 45040, United States

ABSTRACT

Optimal chronic pain detection requires objective and timely communication. However, patients are often unable to convey the intensity and length of pain, leading to frequent miscommunication in clinical settings. To address this issue, a machine learning model was developed to detect chronic pain in a timely, less subjective manner. A manually labeled facial image dataset was used for formal model training and held-out testing. Separately, locally collected volunteer images, including a consented image used as an illustrative example, were used only for exploratory demonstration. Three versions were applied, an Action Unit (AU) based model, a Convolutional Neural Network (CNN), and a hybrid model that consists of both AU and CNN features. The hybrid model achieved a testing accuracy of 0.91, outperforming the individual models. Though the small sample size and limited formal chronic pain representation imposed restrictions, this study demonstrates the possibility for AI-based pain detection to be a reliable, on-demand method that can be used alongside clinical examination to improve pain assessment for patients with limited communication.

Keywords: Computer vision; AI; Machine-learning; Pain-assessment; Healthcare technology; Action Unit; Nonverbal pain indicators; Facial action coding system

INTRODUCTION

Chronic pain is a constant, recurring ache that indicates a long-term weakness in certain parts of the human body. It is characterized as pain that lasts longer than three months, and can be expressed through various symptoms, including throbbing, burning, and shooting (1). More broadly, pain can be categorized as acute, chronic, or simulated pain. Among these categories, chronic pain research is limited due to the variability of pain communication, as social and psychological factors influence how people express pain. Studies show that pain communication tends to be masked or exaggerated due to emotional states, cultural norms, social presence, and depression (2). Limitations in pain communication indicate why an estimated 30% of patients are misdiagnosed (3).

Because of the difficulties surrounding chronic pain diagnosis, a variety of detection methods have been developed to assess pain levels. The approaches vary, depending on the patient, and intensity of the condition.

The most common pain assessment, self-report, is heavily influenced by external factors, often unrelated to the pain (4). These influences, discussed previously, contribute to unnecessary bias and inconsistent evaluation of pain. Another common pain evaluation method,

Corresponding author: Aaryan Verma, E-mail: aav221009@gmail.com.

Accepted June 11, 2026
https://doi.org/10.70251/HYJR2348.43531540

clinician analysis, is sometimes subjective as well, since perception can sway. It also isn't a feasible method, and pain episodes won't get the on-demand attention they need. These current detection systems introduce a common problem: patients who have limited communication, or impaired recognition of pain are unable to apprehend their own pain during critical pain episodes.

To address these challenges, pain assessment has progressed through technology, enabling precise, timely monitoring of pain. Through technological innovation, specific muscles in brow, eye, and mouth regions can be detected to further diagnose chronic pain. Minute contractions in these facial muscles are associated with action units (AUs) (5), which are associated with small changes in facial expression, and are used in the Facial Action Coding System (FACS). FACS is a system that utilizes these AUs to measure facial muscle movements (6). Combining FACS with a convolutional neural network (CNN) (7), which classifies images, allows both the processing and analysis of facial images. This can enable automated facial expression recognition, which can provide on-demand, consistent evaluations of pain, and can be less subjective, compared to controlled clinician or lab settings. Some of the most common practices, their pros and cons, and their percent success are listed in Table 1.

Interest in chronic pain detection was prompted by years of observing an individual experiencing shooting pain attacks from a chronic condition, Trigeminal Neuralgia. During these episodes, the individual was unable to communicate that they were in pain. Though clinician examination of the patient's pain helped determine the intensity of pain, observing facial patterns during shooting attacks offered insight into when exactly the pain was occurring. The approach in this study was to use these facial clues to determine when these shooting attacks started and ended, enabling more precise identification of when nerve compression occurred and the duration of each pain episode.

This machine learning system provides a less subjective approach compared to methods like self-report, as it is not influenced by stress, emotions, or mood. It also offers a low-risk, non-invasive evaluation of pain, with practical, on-demand applicability. When combined with clinical interpretation, assessment can be taken further and provide more accurate evaluations

Table 1. *Overview of current pain evaluation methods, including the pros and cons of six common approaches and their reported success rates. The limitations of these methods highlight the potential of AI-based pain recognition as an improved alternative. Percent success ranges were compiled from the cited sources for each method category and are provided as approximate comparative ranges.*

Existing Method	Pros	Cons	Percent Success
Self-Report	Sensitive to perceived pain intensity (8) Commonly used in clinical monitoring (9)	Influenced by external factors, creating bias (9) Not reliably measurable (10)	69-76% (11)
Clinical Assessment	Reliable, based on trained professional evaluation (12) Detects underlying pathology (13)	Bias from evaluator (14) Pain is inferred, not measured (15)	71-89% (16)
Quantitative Sensory Testing	Directly evaluates nerve function (17) Provides measurable pain data (18)	Requires high-cost equipment (19) Not specific to all chronic pain disorders (15)	69-78% (20)
Neurophysiological Testing	Directly evaluates nerve/brain activity (21) Provides measurable neural data (22)	Not always correlated with pain (23) Influenced by external factors, creating bias (24)	54-80% (25)
Wearable Sensors	Non-invasive monitoring (22) Provides measurable data (26)	Doesn't always correlate with pain (27) Influenced by external factors, creating bias (24)	65-85% (28)
Medical Imaging	Precise diagnosis and visual evidence of abnormalities (29) Provides measurable pain data (21)	Requires high-cost equipment (21) Doesn't always recognize cause of pain (30)	69-87% (31)

of chronic pain. Additionally, AI systems contribute empirical comparison of Action Unit (AU)-only, Convolutional Neural Network (CNN)-based, and hybrid fusion (both AU and CNN) models (2), enhancing pain detection while blending quantitative evaluation with objective analysis.

For patients who cannot communicate or identify their pain, an AI model proposes an advantage to other detection methods, because it offers real-time, immediate outcomes that other machine learning models do not provide.

This study aims to develop a machine learning model combining AUs and a CNN to classify pain, making quantification objective and accessible. The analysis evaluated whether AI-based facial analysis could reliably detect pain and whether models incorporating facial action units (AUs) achieved higher accuracy than approaches that did not use AU-based features.

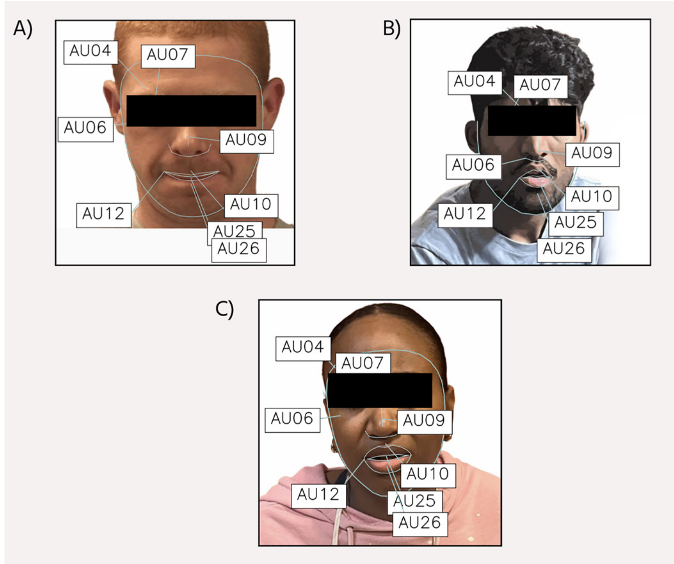

Figure 1. Examples of facial action unit (AU) overlays across different image sources. (A) Original dataset image with annotated AU regions. The overlays show facial regions associated with pain-related expressions, including brow lowering (AU4), eye tightening (AU7), cheek raising (AU6), nose wrinkling (AU9), and mouth-related movements (AU10, AU12, AU25, AU26). (B) Sample image of individual from Chacing Beauty Med Spa with annotated AU regions. (C) Sample image of an individual from Chacing Beauty Med Spa experiencing Trigeminal Neuralgia. Sample images B and C were used anonymously with participant consent for academic publications, in addition to deidentifying them with black bars covering their eyes.

METHODS AND MATERIALS

Study Design and Dataset Source

This study was designed as a supervised machine learning classification study using a retrospective facial image dataset to classify facial images as pain or nonpain. The data was originally sourced from the UNBC-McMaster Shoulder Pain Expression Archive Database, a popular benchmark dataset for pain related classification and modeling (32, 33).

Dataset Description and Manual Labeling

The dataset consisted of only pre-extracted facial images which were unlabeled. Labels therefore were manually assigned to each image within the dataset in order to result in a binary classification (pain vs nonpain). Manual labeling was done by a single evaluator using visible facial expressions and cues coupled with established indicators of pain derived from the Facial Action Coding System (FACS) (6, 34). The features mainly used were brow lowering, mouth deformation, and eye tightening. The images that exhibited clearly defined pain related facial indicators were manually labeled as pain, while the rest were labeled as nonpain due to them being neutral or non-expressive. Ambiguous images were excluded from the labeled dataset in order to maintain consistency and quality of data. Images were considered ambiguous when pain-related facial cues were unclear, conflicting, partially obstructed,affected by poor image quality, or insufficient for confident classification as either pain or nonpain.

The manual labelling process allowed for the creation of a usable binary classified dataset from an unlabeled image dataset. From this point, the labeled image dataset will be referred to as Set A. Set A is made up of 1,166 manually labeled facial images. 133 images out of Set A were separated to be used as a held-out test set for evaluating models, which had 74 nonpain images and 59 pain images. The remaining images out of Set A were used for training models. Only Set A was used for formal model training and final testing. Images from locally collected volunteer sources were not included in training or the final test set.

Image Preprocessing and Facial Action Unit Extraction

To process Set A, the facial images from it were used as static inputs. Facial action unit (AU) data was extracted using Py-Feat architecture and used for the

AU-based model, while raw image data was used for the CNN model (7, 35). Both models were combined to make a hybrid AU-CNN model. All facial images used in this study from Set A were processed through the Py-feat framework (7, 35). Those images were first standardized to a specific resolution and orientation to ensure that results remained consistent across various inputs. Facial detection and landmarking was done by using a RetinaFace-based detection system and a MobileFaceNet landmark model (36, 37). After detection, facial action units (AUs) were extracted from the Set A images through a gradient-boosted model found within Py-feat (7).

A predefined batch of AU intensities was used as input features. These values were normalized before model training. For deep-learning processing, the Set A images were resized and cropped to 224 by 224 pixels by ImageNet standards (38), and were subsequently used as input for the convolutional neural network (CNN) model (Figure 2).

Model Development: AU, CNN, and Hybrid Models

The system uses three distinct modeling approaches within the same pipeline. The three approaches are: AU-only, CNN only, and a AU-CNN hybrid weighted model.

For the AU-only model, facial action unit intensities were used as input features. The features were passed through a machine learning pipeline using a gradient boosting classifier and feature scaling. This model was the baseline for performance since it was interpretable and its dependency on established facial action units (AUs), which is a dependable metric for facial expression (6, 34).

The convolutional neural network (CNN) uses raw image data to learn visual patterns and representations directly from the facial image itself. This model was based on ResNet-18 architecture (39), classifying images by pain or nonpain categories. Image inputs were processed using standard normalization.

The hybrid model combines both the AU and CNN models' embeddings, which results in the hybrid model using both facial features and deep visual patterns in predicting pain. Used as a feature extractor, the CNN model generated a fixed-length embedding vector for each image. Those embeddings were then concatenated with the respective facial action unit (AU) based vector to output a combined representation. The final vector was ultimately passed into a gradient boosting classifier. Hyperparameters were optimized (learning rate = 0.03, 200 estimators, max depth = 3) using grid search before the model utilized a tuned decision threshold to improve classification and maximize the weighted F1 score.

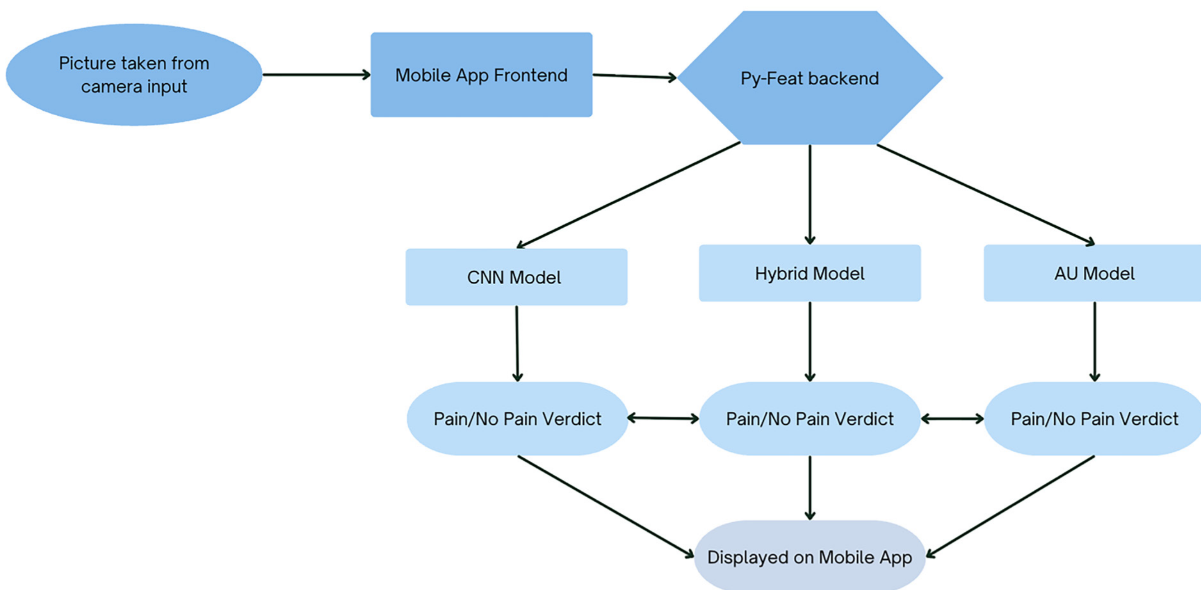

Figure 2. Overview of the mobile pain detection pipeline. An image is captured from the mobile application using a camera, sent to the backend where facial action units (AUs) are extracted. That data is then evaluated by three models (AU-only, CNN only, and an AU-CNN hybrid model). Each model outputs a pain or no pain verdict which is then returned and displayed on the mobile application.

Training, Validation, and Optimization Workflow

The main dataset was split into training and held-out test sets using stratified sampling before model optimization. All preprocessing decisions, normalization, hyperparameter optimization, and threshold selection were performed using the training data only. Scaling parameters were fitted on the training set and then applied to the held-out test set. The held-out test set was not used during model training, preprocessing, hyperparameter selection, or threshold optimization. It was solely used for final model evaluation. Stratified sampling was used to maintain class balance between the pain and nonpain categories. The final performance assessment of the AU-only and AU-CNN hybrid models was performed on the 133-image held-out Set A test set.

Mobile Application Architecture

The proposed pain detection system employs a client-to-server architecture, enabling the real-time analysis of facial images. The mobile client, operated on iOS, is responsible for taking facial photos through the mobile phone's camera and transmitting them to the server backend (40). Processing and analysis of the received images is handled by the backend architecture. This separation allows the overall application to function seamlessly without lag, and also allows the computational analysis to occur on a computer that requires intensive resources. After the image is processed, the results are sent to the mobile client and displayed as pain metrics and classifications to the user. The end-to-end pipeline begins with image capture, followed by processing and model usage on the backend, and resulting in an output on the mobile frontend (Figure 3).

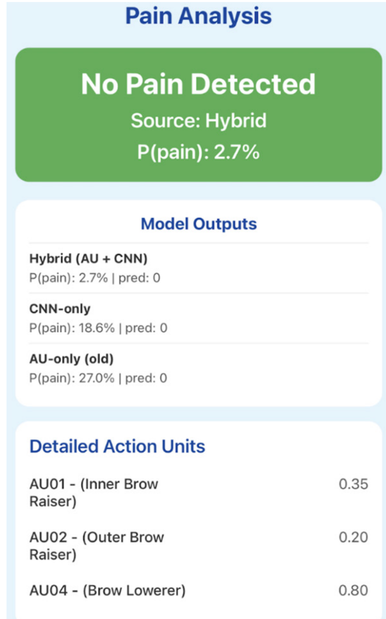

Figure 3. *Interface screenshot of the classification of pain or nonpain following image capture on the mobile application. The primary green panel displays the prediction, with the model source and a P(pain) probability. A secondary section below compares the prediction of all three models (AU-only, CNN-only, and AU-CNN hybrid), which allows for a direct comparison of all three models. The lowest panel shows Facial Action Unit (AU) intensities, which provides insight into the facial features which contribute to the prediction.*

Exploratory Volunteer Datasets: Sets B and C

Alongside the primary dataset, facial images were collected from local volunteer environments as exploratory demonstration datasets. From this point, the local environment datasets will be referred to as Sets B and C. Sets B and C were collected in January 2026 at Chacing Beauty Med Spa, where patients volunteered to provide pain and nonpain pictures to observe whether the trained system could process real-world facial images. Sets B and C were not used for model training, hyperparameter tuning, threshold optimization, or final validation; however, selected consented images from Sets B and C were used as illustrative examples within Figure 1. Performance metrics for Sets B and C were therefore not reported as model results. Volunteers who were experiencing acute pain agreed to having pain and nonpain pictures taken of them. Pain pictures were taken when patients indicated so. Nonpain pictures were taken ten minutes after pain episodes occurred, with confirmation from patients that they were not in pain.

Volunteers who were not experiencing acute or chronic pain were asked to mimic facial patterns correlative to pain, basing it on a level six using standard pain scale. This included grimacing, scrunching eyebrows, squinting, and frowning. Example images were given to patients to give a baseline of what a pain level of six looks like. They were then asked to formulate a neutral expression, and a happier expression, not based on any scale.

At least three images per expression were acquired per volunteer to reduce random variation and increase consistency during exploratory system demonstration. Patients were also asked to remove any eyewear in order for the model to distinctly analyze facial units, as well as effectively calibrate the model for each patient.

Patients who were a part of this study provided a diverse selection of photos, contributing simulated and acute pain, pain and nonpain photos, and a variety of pain expressions. These images from Sets B and C were only used to test the trained system on real-world volunteer images and to observe whether the application could process facial images outside Set A. They were not used to calculate formal accuracy, precision, recall, F1 score, or AUC values reported in the Results section.

Ethical Considerations and Participant Privacy

Prior to image collection from Sets B and C, all participants were provided with a printed consent form, emphasizing that participation is voluntary and that they can withdraw at any time without any repercussions. It also informed them that their data will be stored securely and anonymously. All volunteers independently read through and signed the consent before beginning image collection. Volunteer image collection was conducted with written participant consent for exploratory system demonstration. No names or directly identifying information were included in the analysis datasets or manuscript. Each participant was assigned a unique code number, and facial images were stored securely with access limited to the study team. Because facial images are inherently identifiable, images from Sets B and C were not publicly released as datasets. Any images from Sets B and C included in the manuscript were used anonymously with participant consent for academic publications, in addition to deidentifying them with black bars covering their eyes.

Formal IRB approval was not obtained because Sets B and C were only used for exploratory demonstration and were not used for model training, validation, or final performance evaluation.

RESULTS

Three models were used: AU-only, CNN only, and an AU-CNN hybrid model. The performance of all three models was assessed on the held-out test set taken from Set A using accuracy, precision, recall, F-1 score, and confusion matrices, all on the test dataset. The AU-only model had an accuracy of 0.87 (Table 2).

The AU-only model was balanced in results across both the pain and nonpain classes. The CNN-only model achieved an accuracy score of 0.78. The CNN-only evaluation shown in Table 3 was based on a 54-image evaluable subset rather than the full 133-image held-out test set. CNN-only results should be interpreted as a subset evaluation and not as directly equivalent in sample size to the AU-only and AU-CNN hybrid evaluations. The AU-CNN hybrid model initially had an accuracy of 0.83 with a high number of false positives (Table 3).

Post-optimization, the AU-CNN hybrid model had an accuracy score of 0.91. This multimodal approach hybrid model achieved a precision of 0.94 and a recall of 0.85 for the pain class. This model also had an AUC of 0.956 (Table 4).

Table 2. *This is the classification report for the AU-only facial action unit–based pain detection model evaluated on the test dataset. The model achieved an overall accuracy of 0.87, with precision and recall values of 0.89 and 0.88 for the no-pain class and 0.85 and 0.86 for the pain class.*

	Precision	Recall	F1-Score	Support
No Pain	0.89	0.88	0.88	74
Pain	0.85	0.86	0.86	59
Accuracy	-	-	0.87	133
Macro Avg	0.87	0.87	0.87	133
Weighted Avg	0.87	0.87	0.87	133

Table 3. *This is the classification report for the CNN-only pain detection model evaluated on a 54-image evaluable subset of the test dataset. The model achieved an overall accuracy of 0.78, with precision and recall values of 0.83 and 0.83 for the no-pain class and 0.68 and 0.68 for the pain class.*

	Precision	Recall	F1-Score	Support
No Pain	0.83	0.83	0.83	35
Pain	0.68	0.68	0.68	19
Accuracy	-	-	0.78	54
Macro Avg	0.76	0.76	0.76	54
Weighted Avg	0.78	0.78	0.78	54

Table 4. *This is the classification report for the optimized AU-CNN hybrid multimodal model evaluated on the test dataset. The model achieved an overall accuracy of 0.91, with precision and recall values of 0.89 and 0.96 for the no-pain class and 0.94 and 0.85 for the pain class.*

	Precision	Recall	F1-Score	Support
No Pain	0.89	0.96	0.92	74
Pain	0.94	0.85	0.89	59
Accuracy	-	-	0.91	133
Macro Avg	0.92	0.90	0.91	133
Weighted Avg	0.91	0.91	0.91	133

DISCUSSION

The results demonstrate that each individual approach captures different aspects of facial pain expression, with the AU-CNN hybrid model achieving the strongest performance overall. Model performance was evaluated using accuracy, precision, recall, and F1 score. Accuracy was used to provide a measure of correct predictions, along with precision and recall for an evaluation of class-specific performance. Recall was prioritized after overall accuracy since false negatives (missed pain cases) have more significant consequences than false positives. The F1 score is included in the analysis since it balances precision and recall into one metric that reflects the model's overall effectiveness in classification of pain and nonpain. Confusion matrices were utilized to visualize and analyze performance among pain and nonpain, identifying false negative and positive rates. The visual of a confusion matrix gave a more detailed overview of the behavior of each model. Furthermore, the area under the receiver operating characteristic curve (AUC-ROC) was used to assess the model's capability to differentiate between pain and nonpain classes across different thresholds. Support was included to show the number of samples in each class, which provides context for the reliability of the performance of all three models (Figure 4).

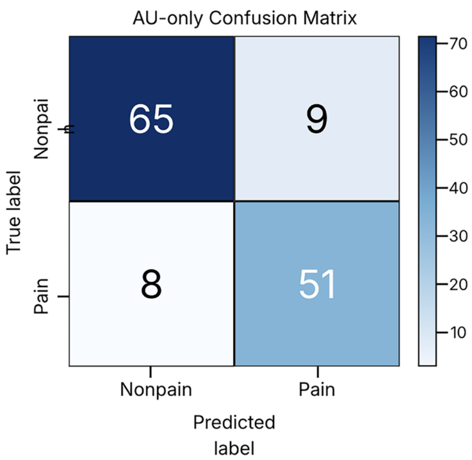

Figure 4. *Confusion matrix for the AU-only facial action unit–based pain detection model evaluated on the test dataset. The model achieved 65 true negatives and 51 true positives, with 9 false positives and 8 false negatives. These results indicate relatively balanced performance across the no-pain and pain classes, establishing the AU-only model as a stable baseline for comparison with the CNN-only and hybrid models.*

The AU-only model provided a strong baseline, with its balanced performance across both pain and nonpain. The results of this model support the existing robust information and research behind the FACS (Facial Action Coding System), which is a promising method for quantifying facial muscles (6, 34).

The CNN-only model showed lower performance than the AU-only and AU-CNN hybrid models in this evaluation. CNN models are extremely capable of learning complex visual patterns, but their performance is dependent on the quality of data and the quantity (38, 39) (Figure 5).

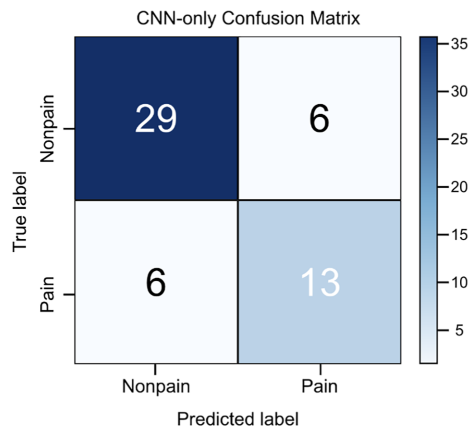

Figure 5. *Confusion matrix for the CNN-only model evaluated on a 54-image evaluable subset of the test dataset. The model achieved 29 true negatives and 13 true positives, with 6 false positives and 6 false negatives. These results indicate that the model demonstrates moderate performance in identifying both no-pain cases and pain cases.*

The AU-CNN hybrid model achieved the highest performance post-optimization (accuracy of 0.91) by combining the features from both AU and CNN models, using both facial action units and deep visual patterns. These results indicate that combining features can enhance model performance compared to single modality systems (41, 42) (Figure 6).

An observation to note for the results of this study is the tradeoff between sensitivity and specificity. Since the mobile app is built to be used for practical applications in real-world scenarios, especially in healthcare, minimizing missed pain cases (false negatives) is more critical than minimizing missed nonpain cases (false positives) (28, 43). The hybrid model is more suitable

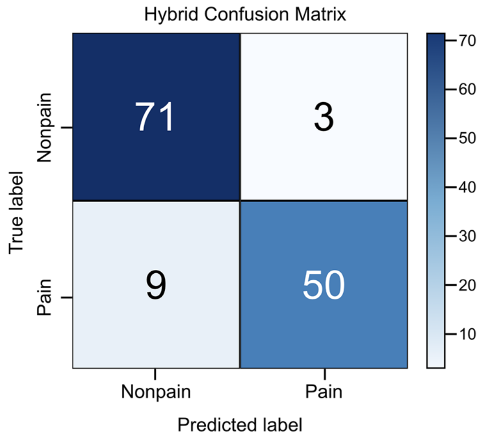

Figure 6. *Confusion matrix for the AU-CNN hybrid model evaluated on the test dataset. The model achieved 71 true negatives and 50 true positives, with 3 false positives and 9 false negatives. These results indicate that the model demonstrates strong performance in identifying both no-pain cases and pain cases, with a low false positive rate and improved classification accuracy compared to the other two models.*

for this purpose, where an accurate pain assessment is essential and has impact.

Beyond overall performance, it is also relevant to consider whether action unit (AU) profiles remain consistent across different types of pain. Consistency in AU feature patterns suggests that certain facial muscle movements can be significant in the faces of individuals with chronic and acute pain. Specifically, this implies that AU-based features can be used as a generalized pain representation, which allows for machine learning models to detect pain independent of the type of pain itself.

Data quality influenced the results of all three models. The main dataset used (Set A) was sourced from the UNBC-McMaster Shoulder Pain Expression Archive Database (32), which required manual labeling due to missing annotations. Although FACS-informed criteria were used during manual labeling, the process itself introduced subjectivity and may have affected model performance (34). Because labeling was performed by a single evaluator, some borderline facial expressions may reflect evaluator-specific judgment rather than a fully generalizable labeling standard. This may limit generalizability because another evaluator could classify subtle or unclear expressions differently. Future work should use multiple independent evaluators and report inter-rater reliability to improve label reliability. Variations in lighting, facial structure, orientation and other real-world factors can affect model performance. In addition, the held-out test set contained 133 images, which limits the stability of the reported performance estimates and may increase the risk that model performance reflects dataset-specific patterns. The CNN-only model was evaluated on a smaller 54-image subset, which limits direct comparison between the CNN-only model and the AU-only and AU-CNN hybrid models. Although the hybrid model achieved the highest accuracy in this study, external validation on a larger dataset is needed to better assess overfitting risk and real-world reliability. Specifically, overfitting may have occurred if the model learned patterns specific to the limited images in Set A rather than generalizable pain-related facial expressions.

Future improvements on the facial pain app could include increasing the size of the manually labelled dataset and incorporating more diverse subject facial pain photos. CNN model performance may improve with a larger and more diverse facial image dataset. Future work could also evaluate more advanced model architectures and a more precise labeling scheme beyond binary pain and nonpain classification. Overall, these findings support the hypothesis that AI-based facial analysis can reliably detect pain, and that models incorporating AU-based features have improved performance over approaches that rely only on raw image data.

CONCLUSION

This study's aim was to construct a machine learning model to classify facial images as pain or nonpain, making quantification objective and more accessible. Although precision, recall, and F1 score were also evaluated, this study prioritized accuracy because the main goal was to measure overall classification correctness rather than focusing on minimizing a specific type of misclassification. Through deliberate investigation, the hybrid model provided the highest accuracy compared to the other two models, achieving an accuracy of 0.91 on the held-out test set. This indicates that the model can accurately distinguish pain 91% of the time, which is advantageous for patients unable to distinguish pain themselves.

Although the hybrid model achieved high performance, the study has limitations. The models were trained on data labeled by a single evaluator, which may introduce subjectivity into the training data. In addition, the relatively small held-out test set may increase overfitting risk and limit the generalizability of the findings to broader real-world settings.

Future studies regarding pain-detection could be more accurate by gathering more pictures from patients experiencing chronic pain. This would strengthen the model's ability to identify this specific type of pain. Furthermore, developing the detection model to identify both the location and timing of pain would improve the app's ability to assess chronic pain and overall reliability. By utilizing action units and a convolutional neural network, this AI trained pain-recognition model demonstrates reliability and provides a relatively objective and feasible method that enhances chronic pain assessment, potentially aiding real-world clinical settings.

FUNDING SOURCES

No external funding was provided for the research or writing of this article.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to this work.

REFERENCES

1. Cleveland Clinic. Chronic pain: what it is, symptoms, treatment. Available from: https://my.clevelandclinic.org/health/diseases/4798-chronic-pain (accessed on 2026-6-9).
2. Rowbotham S, Wardy AJ, Lloyd DM, Wearden A, Holler J. Increased pain intensity is associated with greater verbal communication difficulty and increased production of speech and co-speech gestures. *PLoS One.* 2014; 9 (10). doi:10.1371/journal.pone.0110779.
3. Kramer Dillof Livingston & Moore. Lack of communication causes devastating hospital errors. Available from: https://kdlm.com/lack-of-communication-leads-to-devastating-hospital-errors/ (accessed on 2026-6-9).
4. Dutta P, M N. Facial pain expression recognition in real-time videos. *J Healthc Eng.* 2018; 2018: 7961427. doi:10.1155/2018/7961427.
5. Tian Y, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. *IEEE Trans Pattern Anal Mach Intell.* 2001; 23 (2): 97-115. doi:10.1109/34.908962.
6. Ekman P, Friesen WV, Hager JC. Facial Action Coding System: the manual. Salt Lake City (UT): Research Nexus; 2002. ISBN: 978-0-931835-01-8.
7. Cheong JH, Jolly E, Xie T, Byrne S, Kenney M, Chang LJ. Py-Feat: Python Facial Expression Analysis Toolbox. *Affective Sci.* 2023; 4 (4). doi:10.1007/s42761-023-00191-4.
8. International Association for the Study of Pain. IASP announces revised definition of pain. Available from: https://www.iasp-pain.org/publications/iasp-news/iasp-announces-revised-definition-of-pain/ (accessed on 2025-6-18).
9. Price AK, Bruceta M, Donovan J, *et al*. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *Pain*. 2020; 161 (9): 1976-1982. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7884155/ (accessed on 2025-6-18).
10. Suri P, Heagerty PJ, Timmons A, Jensen MP. Description and initial validation of a novel measure of pain intensity: the Numeric Rating Scale of Underlying Pain without Concurrent Analgesic Use. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11189761/ (accessed on 2026-6-18).
11. Krebs EE, Carey TS, Weinberger M. Accuracy of the pain numeric rating scale as a screening test in primary care. *J Gen Intern Med*. 2007; 22 (10): 1453-1458. doi:10.1007/s11606-007-0321-2.
12. Stretanski MF, Stinocher S, Grandhe S. Pain assessment. Available from: https://www.ncbi.nlm.nih.gov/books/NBK556098/ (accessed on 2026-1-3).
13. Funkhouser WK. Pathology: the clinical description of human disease. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7380242/ (accessed on 2026-2-2).
14. Ahluwalia SC, Giannitrapani KF, Dobscha SK, *et al*. Sometimes you wonder, is this really true?: clinician assessment of patients' subjective experience of pain. Available from: https://pubmed.ncbi.nlm.nih.gov/31680385/ (accessed on 2025-8-13).
15. Mackey S, Greely HT, Martucci KT. Neuroimaging-based pain biomarkers: definitions, clinical and research applications, and evaluation frameworks to achieve personalized pain medicine. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6727999/ (accessed on 2025-7-21).
16. Chanques G, Pohlman A, Kress JP, Molinari N, de Jong A, Jaber S, *et al*. Psychometric comparison of three behavioural scales for the assessment of pain in critically ill patients unable to self-report. *Crit Care*. 2014; 18 (5): R160. https://doi.org/10.1186/cc14000
17. Krumova EK, Geber C, Westermann A, *et al*. Neuropathic pain: is quantitative sensory testing helpful? Available from: https://pubmed.ncbi.nlm.nih.gov/22623149/ (accessed on 2025-7-15).
18. Mücke M, Cuhls H, Radbruch L, *et al*. Quantitative sensory testing (QST). Available from: https://pubmed.ncbi.nlm.nih.gov/26826097/ (accessed on 2025-7-20).
19. van Driel MEC, Huygen FJPM, Rijsdijk M, *et al*.

Quantitative sensory testing: a practical guide and clinical applications. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11368601/ (accessed on 2025-7-20).

20. Ponirakis G, Odriozola MN, Odriozola S, Petropoulos IN, Azmi S, Ferdousi M, *et al*. NerveCheck for the detection of sensory loss and neuropathic pain in diabetes. *Diabetes Technol Ther*. 2016; 18 (12): 800-805. doi:10.1089/dia.2016.0279.
21. Morton DL, Sandhu JS, Jones AKP. Brain imaging of pain: state of the art. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5019436/ (accessed on 2025-7-19).
22. Zhang LB, Chen YX, Li ZJ, *et al*. Advances and challenges in neuroimaging-based pain biomarkers. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11513815/ (accessed on 2025-7-21).
23. Robinson ME, Staud R, Price DD. Pain measurement and brain activity: will neuroimages replace pain ratings? Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3790328/ (accessed on 2025-7-19).
24. Mouraux A, Iannetti GD. The search for pain biomarkers in the human brain. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6262221/ (accessed on 2025-7-28).
25. Zhu W, Chen Y, Zhang H, *et al*. Physiological sensor modality sensitivity test for pain intensity classification in quantitative sensory testing. *Sensors (Basel)*. 2025; 25 (7): 2086. https://doi.org/10.3390/s25072086
26. Semiz B, Hancioglu OK, Sahin RS. Pain assessment and determination methods with wearable sensors: a scoping review. Available from: https://pubmed.ncbi.nlm.nih.gov/40987964/ (accessed on 2025-7-19).
27. Ayena JC, Bouayed A, Ben Arous M, *et al*. Predicting chronic pain using wearable devices: a scoping review of sensor capabilities, data security, and standards compliance. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC12137249/ (accessed on 2026-3-26).
28. Chen J, Abbod M, Shieh JS. Pain and stress detection using wearable sensors and devices: a review. *Sensors (Basel)*. 2021; 21 (4): 1030. doi:10.3390/s21041030.
29. Martucci KT, Mackey SC. Neuroimaging of pain: human evidence and clinical relevance of central nervous system processes and modulation. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5953782/ (accessed on 2025-7-19).
30. Moayedi M, Salomons TV, Atlas LY. Pain neuroimaging in humans: a primer for beginners and non-imagers. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6192705/ (accessed on 2025-7-26).
31. Santana AN, Cifre I, Novaes de Santana C, Montoya P. Using deep learning and resting-state fMRI to classify chronic pain conditions. *Front Neurosci*. 2019; 13: 1313. doi:10.3389/fnins.2019.01313.
32. Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I. Painful data: the UNBC-McMaster Shoulder Pain Expression Archive Database. In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011). 2011; p.57-64. https://doi.org/10.1109/FG.2011.5771462
33. Coder98. Emotion pain dataset. Available from: https://www.kaggle.com/datasets/coder98/emotionpain (accessed on 2025-6-12).
34. Prkachin KM. Assessing pain by facial expression: facial expression as nexus. *Pain Res Manag*. 2009; 14 (1): 53-58. https://doi.org/10.1155/2009/542964
35. Py-Feat documentation. Available from: https://py-feat.org/pages/intro.html (accessed on 2026-3-26).
36. Deng J, Guo J, Zhou Y, Yu J, Kotsia I, Zafeiriou S. RetinaFace: single-stage dense face localisation in the wild. arXiv [Preprint]. 2019:arXiv:1905.00641. Available from: https://arxiv.org/abs/1905.00641 (accessed on 2026-3-26).
37. Chen S, Liu Y, Gao X, Han Z. MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices. arXiv [Preprint]. 2018:arXiv:1804.07573. Available from: https://arxiv.org/abs/1804.07573 (accessed on 2025-6-22).
38. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, *et al*. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015; 115: 211-252. doi:10.1007/s11263-015-0816-y.
39. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016; p.770-778. https://doi.org/10.1109/CVPR.2016.90
40. Expo Documentation. Available from: https://docs.expo.dev/ (accessed on 2025-5-16).
41. Chen Z, Ansari R, Wilkie DJ. Learning pain from action unit combinations: a weakly supervised approach via multiple instance learning. *IEEE Trans Affect Comput*. 2022; 13 (1): 135-146. doi:10.1109/TAFFC.2019.2949314.
42. Ben Aoun N. A review of automatic pain assessment from facial information using machine learning. *Technologies (Basel)*. 2024; 12 (6): 92. doi:10.3390/technologies12060092.
43. Helfand M, Freeman M. Assessment and management of acute pain in adult medical inpatients: a systematic review. Washington (DC): Department of Veterans Affairs (US); 2008. Available from: https://www.ncbi.nlm.nih.gov/books/NBK56602/ (accessed on 2026-2-12).