

**DESARROLLO DE UN SISTEMA DE VISIÓN ARTIFICIAL PARA LA DETECCIÓN DE
MICROEXPRESIONES FACIALES ASOCIADAS AL DOLOR
DEVELOPMENT OF AN ARTIFICIAL VISION SYSTEM FOR THE DETECTION OF
FACIAL MICROEXPRESSIONS ASSOCIATED WITH PAIN**

Autores: ¹Yadira Abigail Tutasig Ushiña y ²Luis Geovanny Romero Mejía.

¹ORCID ID: <https://orcid.org/0009-0004-8443-9199>

²ORCID ID: <https://orcid.org/0009-0006-0037-5955>

¹E-mail de contacto: yadirabigail1234@outlook.com

²E-mail de contacto: lromerom@ups.edu.ec

Afiliación: ¹²Universidad Politécnica Salesiana, (Ecuador).

Artículo recibido: 15 de Enero del 2026

Artículo revisado: 28 de Enero del 2026

Artículo aprobado: 06 de Febrero del 2026

¹Egresada de la carrera de Ingeniería Biomédica de la Universidad Politécnica Salesiana, (Ecuador).

²Profesor de la Universidad Politécnica Salesiana, Quito, Ecuador, y miembro del Grupo de Investigación en Biomecatrónica y Bioingeniería (GIBYB). Es Ingeniero en Electrónica y Control, Tecnólogo en Electromecánica y Magíster en Mecatrónica y Robótica.

Resumen

El dolor se evalúa con frecuencia mediante escalas subjetivas, lo que reduce la precisión y aumenta la variabilidad clínica. En este estudio se desarrolló y evaluó un sistema de visión artificial para detectar microexpresiones faciales asociadas al dolor en pacientes de 20 a 68 años utilizando el dataset Pain Expressions Multimodal Framework (PEMF). El flujo incluyó detección y recorte facial, normalización espacial y temporal de los frames y extracción de rasgos espaciotemporales con una red MoViNet preentrenada, complementada con descriptores geométricos basados en Unidades de Acción. Se compararon dos enfoques: (i) modelo híbrido MoViNet + clasificadores supervisados (MLP y XGBoost) y (ii) modelo end-to-end I3D entrenado directamente sobre secuencias de video. El desempeño se midió como clasificación binaria dolor/no dolor con F1-macro y balanced accuracy, por el desbalance del conjunto de prueba (223 clips). Los resultados globales mostraron mejor rendimiento para los modelos híbridos (MoViNet+XGBoost: accuracy 0.97, F1-macro 0.96, balanced accuracy 0.9486; MoViNet+MLP: accuracy 0.96, F1-macro 0.94) frente a I3D (accuracy 0.93, F1-macro 0.91), además de menor riesgo de omitir casos con dolor. Se concluye que el enfoque híbrido ofrece un soporte objetivo, interpretable y

ajustable para la valoración clínica del dolor, con potencial de integración en una interfaz web casi en tiempo real en entornos hospitalarios y de telemedicina

Palabras clave: Visión artificial, Microexpresiones faciales, Dolor, MoViNet, I3D, Aprendizaje profundo.

Abstract

Pain is often assessed with subjective scales, which lowers accuracy and increases inter-observer variability. This study developed and evaluated a computer-vision system to detect facial microexpressions associated with pain in patients aged 20–68 years using the Pain Expressions Multimodal Framework (PEMF) dataset. The pipeline included face detection and cropping, spatial and temporal normalization of frames, and spatiotemporal feature extraction with a pretrained MoViNet network, complemented with geometric descriptors derived from Facial Action Units. Two strategies were compared: (i) a hybrid model combining MoViNet features with supervised classifiers (MLP and XGBoost) and (ii) an end-to-end deep learning model based on the I3D architecture trained directly on video sequences. Performance was framed as binary pain/no-pain classification and evaluated with macro F1-score and balanced accuracy due to class imbalance in the test set (223 clips). Overall results favored the hybrid approaches (MoViNet+XGBoost:

accuracy 0.97, macro F1 0.96, balanced accuracy 0.9486; MoViNet+MLP: accuracy 0.96, macro F1 0.94) over I3D (accuracy 0.93, macro F1 0.91), with a lower risk of missing pain cases. In addition, the work highlights the value of prioritizing clinically critical errors and selecting thresholds that balance sensitivity and false alarms carefully

Keywords: Computer vision, Facial microexpressions, Pain detection, MoViNet, I3D, Deep learning.

Sumário

A avaliação da dor na prática clínica depende com frequência de escalas subjetivas, o que reduz a precisão e aumenta a variabilidade entre observadores. Este estudo desenvolveu e avaliou um sistema de visão computacional para detectar microexpressões faciais associadas à dor em pacientes de 20 a 68 anos, utilizando o dataset Pain Expressions Multimodal Framework (PEMF). O pipeline incluiu detecção e recorte do rosto, normalização espacial e temporal dos frames e extração de características espaço-temporais com uma rede MoViNet pré-treinada, complementada por descritores geométricos derivados de Unidades de Ação faciais. Foram comparadas duas estratégias: (i) um modelo híbrido que combina atributos extraídos por MoViNet com classificadores supervisionados (MLP e XGBoost) e (ii) um modelo end-to-end baseado na arquitetura I3D treinado diretamente em sequências de vídeo. O desempenho foi avaliado como classificação binária dor/sem dor com F1-macro e balanced accuracy, considerando o desbalanceamento do conjunto de teste (223 cliques). Os resultados favoreceram os modelos híbridos (MoViNet+XGBoost: acurácia 0.97, F1-macro 0.96, balanced accuracy 0.9486; MoViNet+MLP: acurácia 0.96, F1-macro 0.94) em relação ao I3D (acurácia 0.93, F1-macro 0.91), com menor risco de omitir casos com dor. Conclui-se que a abordagem híbrida é promissora como suporte objetivo, interpretável e ajustável para a avaliação clínica, com integração quase em tempo real em contextos biomédicos sensíveis.

Palavras-chave: Visão computacional, Microexpressões faciais, Dor, MoViNet, I3D, Aprendizado profundo.

Introducción

El dolor representa un estímulo sensorial y emocional complejo en el estado fisiológico humano. La International Association for the Study of Pain (IASP) lo define como una experiencia desagradable asociada a un daño tisular real o potencial (Raja et al., 2020). En el entorno clínico, el dolor es considerado un signo vital, aumentando la necesidad de evaluarlo con precisión para garantizar una atención médica segura y eficaz (Fang et al., 2025). Los métodos tradicionales de evaluación, como la Escala Visual Analógica (EVA) y la Escala Numérica del Dolor (NRS), presentan limitaciones significativas debido a su dependencia del estado cognitivo y comunicativo del paciente, así como a la variabilidad entre observadores (Ministerio de Salud Pública del Ecuador, 2024). Estas limitaciones se incrementan en poblaciones vulnerables, como pacientes geriátricos, sedados o con trastornos neurológicos.

Las microexpresiones faciales han sido propuestas como un biomarcador objetivo para la evaluación del dolor. El Facial Action Coding System (FACS), desarrollado por Ekman y Friesen, ofrece un marco sistemático para la codificación de la actividad muscular facial mediante unidades de acción (AUs), algunas vinculadas consistentemente con la expresión del dolor (Chaisiriprasert y Patchsuwan, 2025; Ekman y Friesen, 1978). Los avances en inteligencia artificial y visión por computadora han permitido el desarrollo de sistemas automáticos para analizar expresiones faciales a partir de secuencias de video. Las redes neuronales convolucionales tridimensionales (CNN 3D) han demostrado desempeño destacado en la modelación conjunta de

información espacial y temporal (Chen et al., 2024; Sabater-Gárriz et al., 2024). Arquitecturas como MoViNet han sido diseñadas para capturar dinámicas faciales sutiles de corta duración con alta eficiencia computacional (Hussain et al., 2023). En Latinoamérica y Ecuador existe limitada incorporación de tecnologías de visión artificial para la evaluación objetiva del dolor. La práctica clínica regional continúa apoyándose en métodos subjetivos tradicionales, evidenciando una brecha tecnológica relevante. El presente estudio desarrolló y evaluó un sistema de visión artificial para la detección automática de microexpresiones faciales asociadas al dolor en pacientes de 20 a 68 años, utilizando el conjunto de datos PEMF. Se compararon dos estrategias: un enfoque híbrido basado en MoViNet y clasificadores supervisados, y un enfoque end-to-end basado en I3D, evaluados mediante métricas de clasificación binaria.

Materiales y Métodos

El estudio se llevó a cabo bajo un enfoque computacional y experimental, orientado al análisis del desempeño de modelos de visión artificial en condiciones controladas. La investigación es de tipo aplicada con diseño experimental comparativo. Se utilizó el conjunto de datos Pain Expressions Multimodal Framework (PEMF), disponible públicamente en Open Science Framework (<https://osf.io/3hgca>). El dataset incorpora secuencias de video facial, codificaciones FACS y evaluaciones subjetivas de dolor. La muestra incluyó 68 participantes anónimos, cada uno evaluado bajo cuatro condiciones experimentales: Algometer, Laser, Neutral y Posed. El conjunto original contiene 272 clips de video en formato MP4, con duraciones entre 1 y 3 segundos y frecuencia de muestreo de 24 fps. Mediante técnicas de aumento de datos

(volteo horizontal, rotaciones y ajustes de brillo), se generaron 1,360 clips.

El dataset fue particionado a nivel de sujeto en subconjuntos de entrenamiento (70%), validación (15%) y prueba (15%), garantizando que cada participante esté presente en un único subconjunto para evitar filtración de información. Se implementó un pipeline automatizado para detección facial, recorte y normalización espacial. Se utilizó el algoritmo Multi-task Cascaded Convolutional Networks (MTCNN) sobre el primer frame de cada clip. Todos los frames fueron transformados a formato RGB de tres canales y redimensionados a 172×172 píxeles, preservando la secuencia temporal original a 24 fps. El enfoque híbrido utiliza MoViNet como extractor fijo de características espaciotemporales profundas, con pesos preentrenados en Kinetics-600. Cada clip fue representado mediante 32 frames RGB uniformemente distribuidos. Las características extraídas se utilizaron como entrada para dos clasificadores supervisados:

- Perceptrón multicapa (MLP): entrenado con optimizador Adam, regularización L2, dropout del 10%, y ponderación de clases para mitigar desbalance. Se implementó early stopping con paciencia de 10 épocas.
- Árboles de decisión (XGBoost): entrenados directamente sobre vectores de características, con validación cruzada estratificada (k=10 folds) y ponderación de clases.

El modelo I3D aprende directamente desde frames hasta predicción final. Se entrenó con optimizador Adam (learning rate 1×10^{-5}), regularización L2 ($\lambda = 1 \times 10^{-4}$), y programador StepLR. Se empleó focal loss ($\gamma = 2$, $\alpha = 0.25$) para mitigar desbalance, con early stopping basado en F1-macro de validación. El

desempeño se evaluó mediante métricas de clasificación binaria: precisión, sensibilidad (recall), F1-score macro, balanced accuracy y área bajo la curva ROC (ROC-AUC). Se priorizó el F1-score macro por el desbalance de datos. Se analizaron matrices de confusión para identificar tipos de errores clínicamente relevantes. Se implementó validación cruzada estratificada (k=10 folds) para modelos híbridos. Las curvas ROC analizaron capacidad discriminativa ante distintos umbrales de decisión.

Resultados y Discusión

Modelo híbrido MoViNet y MLP

El clasificador MLP obtuvo accuracy de 0.96, F1-score macro de 0.94 y balanced accuracy de 0.91. El umbral de decisión se ajustó a 0.4, priorizando sensibilidad. La curva ROC indicó ROC-AUC de 0.9972, mostrando alta capacidad discriminativa.

Tabla 1. Resultados por clase MoViNet + MLP (threshold = 0.4)

Clase	Precision	Recall	F1-score
No-Dolor	1.00	0.82	0.90
Dolor	0.94	1.00	0.97

Fuente: Elaboración propia

La matriz de confusión mostró 168 verdaderos positivos, 45 verdaderos negativos, 0 falsos negativos y 10 falsos positivos. El modelo presenta alta sensibilidad para detectar dolor, con tasa mínima de casos positivos omitidos.

Modelo híbrido MoViNet y XGBoost

El modelo basado en árboles alcanzó accuracy de 0.97, F1-score macro de 0.95 y balanced accuracy de 0.9486. El umbral óptimo se fijó en 0.69. El ROC-AUC de 0.9943 confirma capacidad de discriminación entre clases. La matriz de confusión mostró 166 verdaderos positivos, 50 verdaderos negativos, 2 falsos negativos y 5 falsos positivos. Este patrón

evidencia reducción notable de falsas alarmas respecto al MLP, manteniendo alto nivel de detección.

Tabla 2. Resultados por clase MoViNet + XGBoost (threshold = 0.69)

Clase	Precisión	Recall	F1-score
No-Dolor	0.96	0.91	0.93
Dolor	0.97	0.99	0.98

Fuente: Elaboración propia

Modelo I3D end-to-end

El modelo I3D alcanzó accuracy de 0.93, F1-score macro de 0.91 y balanced accuracy de 0.95. El ROC-AUC de 0.9694 y PR-AUC de 0.9913 muestran habilidad discriminativa adecuada.

Tabla 3. Resultados por clase modelo I3D end-to-end

Clase	Precision	Recall	F1-score
No-Dolor	0.78	0.98	0.87
Dolor	0.99	0.91	0.95

Fuente: Elaboración propia

La matriz de confusión reportó 153 verdaderos positivos, 54 verdaderos negativos, 15 falsos negativos y 1 falso positivo. El incremento de falsos negativos implica mayor omisión de casos con dolor.

Comparación global de modelos

Tabla 4. Comparación global de desempeño entre modelos evaluados

Modelo	Accuracy	Precision	Recall	F1-Macro	Bal. Acc.	ROC-AUC
MoViNet+MLP	0.96	0.97	0.91	0.94	0.91	0.9972
MoViNet+XGBoost	0.97	0.97	0.95	0.96	0.9486	0.9943
I3D end-to-end	0.93	0.89	0.95	0.91	0.95	0.9694

Fuente: Elaboración propia

Interfaz web

Se desarrolló una interfaz web interactiva basada en Streamlit para evaluación visual del sistema. El tiempo total de respuesta se mantiene bajo el segundo (793 ms promedio), permitiendo interacción casi en tiempo real. La etapa de mayor consumo temporal corresponde a extracción de características mediante MoViNet (558 ms).

Tabla 5. *Tiempos medios de respuesta de la plataforma web (n=25 ejecuciones)*

Etapas	Tiempo (ms)	Desv. estándar (ms)
Registro video	140	18
Inferencia IA	558	42
Resultados	95	12
Tiempo total	793	52

Fuente: Elaboración propia

Los modelos híbridos basados en MoViNet demostraron mayor capacidad para capturar microvariaciones espaciotemporales en regiones faciales relevantes. La separación explícita entre extracción de características y clasificación facilita el aprendizaje de patrones sutiles asociados al dolor, reduciendo riesgo de sobreajuste cuando el conjunto de datos es limitado. El comportamiento diferenciado entre MLP y XGBoost evidencia cómo el clasificador final influye directamente en el tipo de error cometido. Mientras el MLP favorece detección más sensible del dolor, los árboles de decisión introducen mecanismo más conservador, reduciendo falsas alarmas. El modelo I3D, aunque presenta arquitectura más integrada, mostró mayor dependencia de la distribución del conjunto de entrenamiento, reflejado en incremento de falsos negativos. Bajo condiciones de datos limitados, el aprendizaje conjunto de características y clasificación puede priorizar patrones dominantes y omitir señales faciales de baja intensidad. Desde perspectiva clínica, estos hallazgos sugieren que

interpretabilidad y control del proceso de decisión son factores tan relevantes como el desempeño global. La capacidad de ajustar el sistema según tipo de error aceptable resulta crítica, posicionando enfoques híbridos como alternativa más flexible y segura para detección automática del dolor.

Conclusiones

El presente trabajo abordó la detección automática del dolor facial mediante análisis de microexpresiones, utilizando información espaciotemporal extraída de secuencias de video. Los resultados permiten afirmar que el enfoque propuesto identifica patrones faciales asociados al dolor de manera consistente, respaldando la viabilidad de técnicas de visión artificial como apoyo al análisis automático. El F1-score macro adquiere mayor relevancia al evaluar de forma equilibrada el desempeño sobre clases Dolor y No-Dolor, especialmente en presencia de desbalance de datos. Los hallazgos se entienden a la luz del diseño metodológico adoptado, apoyado en conjunto de datos experimental y evaluación exploratoria.

Los modelos híbridos presentaron desempeño más estable y confiable, reduciendo errores clínicamente críticos. El modelo MoViNet+XGBoost logró mejor balance entre sensibilidad y control de falsas alarmas (F1-macro=0.96, ROC-AUC=0.9943). Este trabajo aporta evidencia técnica preliminar sobre uso de modelos de visión artificial para análisis automático del dolor facial, estableciendo base metodológica sólida para investigaciones futuras. La ampliación del conjunto de datos, incorporación de escenarios clínicos más diversos y empleo de diseños experimentales de mayor alcance permitirán evaluar con mayor profundidad la aplicabilidad del enfoque propuesto en contextos reales.

Referencias Bibliográficas

- Chen, W. (2024). Decoding pain through facial expressions: A study of patients with migraine. *The Journal of Headache and Pain*, 25. <https://pubmed.ncbi.nlm.nih.gov/38462615/>
- Chaisiriprasert, P. (2025). A weighted facial expression analysis for pain level estimation. *Journal of Imaging*, 11(5), 151. <https://www.mdpi.com/2313-433X/11/5/151>
- Dhar, M. (2025). A novel 3D convolutional neural network-based deep learning model for spatiotemporal feature mapping for video analysis. *Journal of Imaging*, 11(7), 243.
- Ekman, P. (1978). Facial Action Coding System (FACS): A technique for the measurement of facial movement. Consulting Psychologists Press.
- Fang, R. (2025). Survey on pain detection using machine learning models: Narrative review. *JMIR AI*, 4.
- Hammal, Z. (2012). Automatic detection of pain intensity. Proceedings of the ACM International Conference on Multimodal Interaction, 47.
- Hussain, T. (2023). EMO-MoviNet: Enhancing action recognition in videos with EvoNorm, Mish activation, and optimal frame selection for efficient mobile deployment. *Sensors*, 23(19), 8106.
- Javaid, M. (2024). A systematic review for classification and selection of deep learning methods. *Decision Analytics Journal*, 12, 100489.
- Kondratyuk, D. (2021). Movinets: Mobile video networks for efficient video recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 16015–16025.
- Lecun, Y. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Martin, P. (2022). 3D convolutional networks for action recognition: Application to sport gesture recognition. arXiv preprint.
- Ministerio de Salud Pública del Ecuador. (2024). El ministerio de salud pública ejerce la rectoría del sistema nacional de salud. <https://www.salud.gob.ec/>
- Raja, S. (2020). The revised international association for the study of pain definition of pain: Concepts, challenges, and compromises. *Pain*, 161(9), 1976–1982.
- Sabater-Gárriz, Á. (2024). Automated facial recognition system using deep learning for pain assessment in adults with cerebral palsy. *Digital Health*, 10.
- Talluri, K. (2022). Deep 3D convolutional neural network for facial micro-expression analysis from video images. *Applied Sciences*.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional. Copyright © Yadira Abigail Tutasig Ushiña y Luis Geovanny Romero Mejía.

