

Detección en tiempo real de mala postura y signos de dolor utilizando la cámara de una computadora

Alejandro Román Telli
aroman.telli@gmail.com

Jerónimo Grinovero
jeronimogrinovero@gmail.com

Pablo Mariño
pabloomarino56@gmail.com

Resumen—Este trabajo presenta un sistema automático para detectar expresiones faciales asociadas al dolor y evaluar la postura del usuario frente a la computadora, utilizando únicamente la cámara de la computadora y técnicas de visión por computadora en tiempo real. Se emplean las bibliotecas OpenCV, MediaPipe y modelos de aprendizaje automático supervisado entrenados con datos individuales.

I. INTRODUCCIÓN

A. Expresiones faciales y dolor

El **Facial Action Coding System (FACS)** es un sistema desarrollado para codificar las expresiones faciales humanas mediante **Unidades de Acción (AUs)**, cada una correspondiente a la activación de músculos específicos del rostro. Investigaciones como las de Prkachin [1] y Chen et al. [2] han demostrado que ciertas AUs están fuertemente correlacionadas con expresiones faciales asociadas al dolor físico, lo que permite utilizarlas como indicadores objetivos en sistemas automáticos de detección.

Basado en el estudio de Chen et al. [2], este trabajo emplea seis AUs relacionadas con contracciones y relajaciones musculares típicas en situaciones de dolor para desarrollar un algoritmo de detección automática. Las AUs seleccionadas son las siguientes:

- **AU4:** descenso de cejas, común en expresiones de dolor.
- **AU6 y AU7:** elevación de mejillas y tensión de párpados, asociadas con incomodidad o angustia.
- **AU9 y AU10:** arruga nasal y elevación del labio superior, indicadores de aversión o malestar.
- **AU43:** cierre de ojos, característico del dolor extremo o sufrimiento intenso.

B. Postura del cuerpo

En el contexto del uso de computadoras, la postura del usuario puede influir significativamente en su bienestar físico. Si bien existen ciertos indicadores visuales que permiten inferir una postura adecuada o inadecuada, estos son altamente dependientes del entorno físico y los hábitos individuales del usuario. Factores como el ángulo de apertura de la computadora portátil, la cercanía a la cámara, la altura del escritorio y la silla, entre otros, introducen un grado de subjetividad que dificulta la estandarización de criterios.

Para abordar esta variabilidad, se toman como referencia estudios previos en ergonomía y visión por computadora [3], lo que permite establecer parámetros flexibles que guíen el desarrollo de un sistema de detección de postura basado en

imágenes capturadas por la cámara frontal del dispositivo. Este sistema busca reconocer patrones corporales que, de forma general, pueden asociarse con una postura ergonómicamente aceptable o con desviaciones potencialmente perjudiciales.

II. DISEÑO Y DESARROLLO

El sistema desarrollado realiza la detección de expresiones faciales asociadas al dolor y la evaluación de la postura del usuario en tiempo real. Para ambos casos, se adopta un enfoque basado en visión por computadora e inteligencia artificial, cuyo funcionamiento puede describirse mediante las siguientes etapas:

- 1) **Captura de video en tiempo real** desde la cámara del dispositivo, permitiendo la adquisición continua de datos visuales del usuario.
- 2) **Preprocesamiento de la imagen** con el objetivo de reducir el ruido, estandarizar las entradas y aumentar la robustez del algoritmo frente a variaciones individuales o ambientales.
- 3) **Detección de puntos clave** de las partes del cuerpo y la cara con MediaPipe.
- 4) **Extracción de puntos clave relevantes:**
 - **Unidades de Acción (AUs)** para el análisis facial.
 - Posición de los **hombros, ojos, orejas y boca** para el análisis de postura.
- 5) **Tratamiento de los datos**, para obtener posiciones relativas y normalizadas de los puntos clave, que provee de mayor versatilidad al algoritmo.
- 6) **Evaluación la posición del rostro** para verificar si los datos obtenidos son lo suficientemente precisos para el análisis.
- 7) **Entrenamiento supervisado** de modelos de inteligencia artificial utilizando patrones generados por el usuario de la computadora. Estos modelos aprenden a reconocer patrones vinculados al dolor facial o a posturas adecuadas/inadecuadas.
- 8) **Clasificación en tiempo real**, mediante la cual el sistema aplica los modelos previamente entrenados para determinar la presencia de signos de dolor o una postura no ergonómica.
- 9) **Visualización de resultados** en la interfaz gráfica, ofreciendo retroalimentación inmediata al usuario.

A. Captura de video en tiempo real y preprocesamiento de la imagen

Para capturar la imagen de la cámara de la computadora se utiliza la biblioteca **OpenCV**. Luego, cada fotograma se procesa para mitigar condiciones de imagen que podrían dificultar una detección robusta de rasgos faciales, tales como desenfoque e iluminación inadecuada.

En primer lugar, se evalúa el nivel de nitidez de la imagen mediante el cálculo de la varianza del operador de Laplaciano aplicado sobre la versión en escala de grises del fotograma. Si dicha varianza se encuentra por debajo de un umbral determinado, se considera que la imagen presenta desenfoque. En ese caso, se aplica un filtro de realce para acentuar bordes y detalles.

Posteriormente, se analiza la iluminación global de la imagen a través del valor medio del canal de brillo (V) en el espacio de color HSV. Este valor permite clasificar la iluminación en tres categorías: oscura, normal o sobre expuesta.

- Si la imagen es clasificada como **oscura**, se aplica una corrección gamma con un coeficiente mayor que uno $\gamma > 1$ para aumentar el brillo general.
- En cambio, si se detecta una **sobre exposición**, se utiliza una corrección gamma con un coeficiente menor que uno $\gamma < 1$ para reducir la intensidad lumínica.

Este procedimiento contribuye a garantizar condiciones visuales óptimas para una correcta detección facial, disminuyendo así la probabilidad de errores causados por imágenes borrosas o con iluminación deficiente.

B. Detección de puntos clave y extracción de datos relevantes

Cada fotograma es luego procesado utilizando la biblioteca **MediaPipe**, que permite detectar y extraer en tiempo real los puntos de referencia faciales y corporales del usuario.

En el caso del rostro, MediaPipe genera una malla detallada compuesta por 468 puntos clave (*keypoints*), cada uno de los cuales cuenta con cuatro atributos: las coordenadas espaciales (x, y, z) y un valor de visibilidad v que indica la confiabilidad de la detección. Sin embargo, para este trabajo sólo se consideran aquellos puntos relevantes para las Unidades de Acción (AUs) asociadas a expresiones de dolor. Por lo tanto, únicamente se almacenan los *keypoints* correspondientes a estas regiones faciales específicas, descartando el resto con el objetivo de optimizar el procesamiento.

Para el cuerpo, MediaPipe Pose genera un conjunto de 33 *keypoints* que representan las principales articulaciones y partes del cuerpo humano. Al igual que en el caso facial, cada *keypoint* corporal incluye las coordenadas espaciales (x, y, z) y un valor de visibilidad v . Para una detección primaria de la postura, es suficiente analizar los *keypoints* que se encuentran por encima del torso, que son los siguientes: nariz, ojos, orejas, boca y hombros.

La extracción selectiva de *keypoints* relevantes permite una mayor eficiencia en las etapas posteriores de análisis, mejorando el rendimiento del sistema sin comprometer la calidad

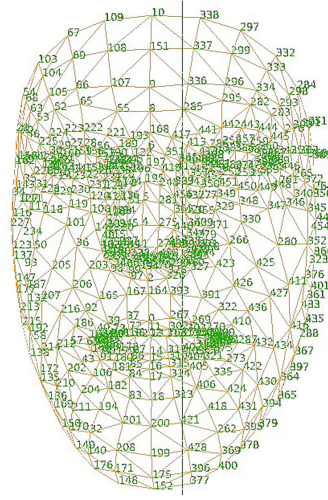


Fig. 1. Keypoints faciales de MediaPipe [4]

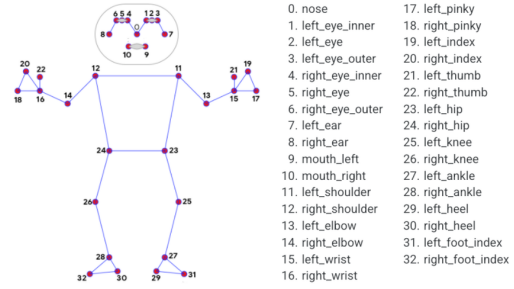


Fig. 2. Keypoints corporales de MediaPipe [5]

de la información esencial para las tareas de clasificación de postura y expresión facial.

C. Tratamiento de los datos

Para garantizar la independencia de escala, orientación y distancia a la cámara, se aplica un método de normalización a los puntos clave detectados, tanto para el rostro como para el cuerpo.

El procedimiento general consiste en:

- 1) Elegir dos **keypoints de referencia** a y b para calcular la posición X_{med} del punto medio entre los dos.
 - Para el análisis del rostro, se toman las sienes izquierda y derecha.
 - Para el análisis del cuerpo, se toman los hombros izquierdo y derecho.

$$X_{med} = \frac{X_a + X_b}{2}$$

- 2) Calcular la **posición relativa** X' de cada *keypoint* k con respecto a X_{med}

$$X'_k = X_k - X_{med}$$

- 3) Calcular la **distancia de referencia** entre X_a y X_b

$$D = ||X_a - X_b||$$

- 4) Calcular la **posición normalizada** X'' del *keypoint* k

$$X''_k = \frac{X'_k}{D}$$

D. Evaluación de la orientación del rostro

Con el fin de garantizar la validez del procesamiento facial, se evalúa la orientación del rostro del usuario respecto a la cámara. Este mecanismo se basa en el cálculo del ángulo entre el plano definido por tres puntos clave del rostro (el centroide facial, y los extremos izquierdo y derecho del rostro) y un vector de referencia perpendicular al plano de la cámara (por defecto, el vector $(0, 0, -1)$).

El procedimiento comienza con la construcción de dos vectores entre el centroide facial y cada uno de los extremos izquierdo y derecho del rostro. A partir de estos, se obtiene la normal del plano facial mediante el cálculo del producto vectorial. El ángulo entre esta normal y el vector de referencia se utiliza como indicador de la orientación del rostro con respecto al eje frontal de la cámara.

Si el ángulo resultante excede un umbral predefinido, se interpreta que el usuario no está mirando directamente a la cámara. En tal caso, es probable que algunos puntos clave faciales no se detecten correctamente, afectando negativamente las etapas subsiguientes del análisis. Por lo tanto, el sistema solicita al usuario que reoriente su rostro hacia la cámara antes de continuar.

Este proceso de verificación permite asegurar la calidad y completitud de los datos faciales capturados, reduciendo significativamente el riesgo de errores en la detección de expresiones o emociones cuando el rostro se encuentra desalineado.

E. Entrenamiento supervisado de modelos de inteligencia artificial

Se utiliza la biblioteca Scikit-Learn para entrenar dos modelos clasificadores de *gradient boosting*, uno para el análisis del rostro y otro para el análisis de la postura. Para el entrenamiento, se realiza primero una generación de patrones a partir de la captura de imágenes del usuario.

Inicialmente, el usuario debe ubicarse con una postura correcta frente a la cámara, haciendo movimientos leves para un entrenamiento generalizado. Luego, el usuario debe ubicarse con una postura incorrecta frente a la cámara, haciendo movimientos y mostrando diferentes posturas incorrectas para un entrenamiento generalizado. El programa captura los *keyframes* de 250 fotogramas para cada clase.

Para la recopilación de los datos faciales de dolor, el mecanismo es el mismo. El usuario debe mostrar muecas de dolor, así como también mostrar posturas faciales normales, para que el programa capture los *keyframes* de 500 fotogramas para cada clase. Se utilizan más frames para el análisis facial debido a que este es mucho más fino y detallado, lo que requiere un mayor volumen de datos para un entrenamiento robusto.

Los *keyframes* obtenidos son normalizados utilizando el método descrito en la sección II.C, y se almacenan en archivos

.csv, cuyo nombre depende de la clasificación. Con los datos guardados, se entrena el modelo de *gradient boosting* y se lo almacena en un archivo .pkl.

F. Clasificación en tiempo real

Una vez entrenados los modelos, se procede a la clasificación en tiempo real, tanto de la postura corporal como de las expresiones faciales del usuario. Se realiza el siguiente proceso:

- 1) Se captura la imagen de la cámara con OpenCV.
- 2) Se procesa la imagen para mejorar la calidad en la obtención de los *keypoints*.
- 3) Se clasifica la postura corporal.
 - a) Se capturan los 33 *keypoints* corporales utilizando la biblioteca MediaPipe.
 - b) Se filtran los 8 *keypoints* relevantes que se encuentran por encima del torso (nariz, ojos, orejas, boca, hombros) y se extraen las coordenadas tridimensionales (x, y, z) de estos puntos.
 - c) Se normalizan las posiciones utilizando como referencia el punto medio entre los hombros y la distancia entre ellos.
 - d) Se genera un vector de características que se introduce en el modelo entrenado.
 - e) El sistema clasifica la postura corporal como buena o mala.
- 4) Se detecta dolor a partir del análisis facial.
 - a) Se capturan los *keypoints* faciales utilizando la biblioteca MediaPipe.
 - b) Se filtran los *keypoints* relevantes que corresponden a las AUs y se extraen las coordenadas tridimensionales (x, y, z) de estos puntos.
 - c) Se normalizan las posiciones utilizando como referencia el punto medio entre las sienes y la distancia entre ellos.
 - d) Se genera un vector de características que se introduce en el modelo entrenado.
 - e) El sistema clasifica la expresión facial como dolor o no dolor.
- 5) Se muestra la información en pantalla con colores, texto explicativo y visualización de las AUs activadas.

En ambos casos, si el sistema detecta que la persona no está posicionada adecuadamente (por ejemplo, no mirando de frente), se omite el análisis para evitar errores.

III. RESULTADOS

El algoritmo desarrollado presenta buenos resultados a la hora de detectar si el usuario se encuentra en una buena o mala postura, y si está sufriendo dolor o no.

Para las pruebas, se utiliza una computadora con un procesador AMD Ryzen 5500U, una tarjeta gráfica AMD Radeon integrada, 8GB de RAM, y cuya cámara tiene una resolución de 720p y una tasa de 30 fotogramas por segundo.

Durante el entrenamiento del modelo clasificador, se separa un 80% de los patrones recopilados para entrenar, y un 20%

para las pruebas. En el reporte de clasificación de las pruebas, se obtuvo un porcentaje de precisión de 100% para ambos modelos.

La detección en tiempo real funciona oscilando los 15 fotogramas por segundo.

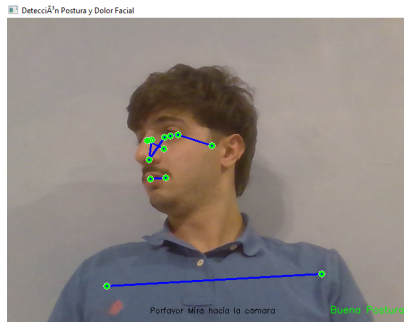


Fig. 3. Rotación de la cara

En la Fig. 3 se observa que el usuario no se encuentra mirando a la cámara, lo que impide una detección fiable de los *keypoints* faciales. Por lo tanto, se le solicita que rote su cabeza, y mientras no lo haga, no se realiza el procesamiento para la detección de dolor.

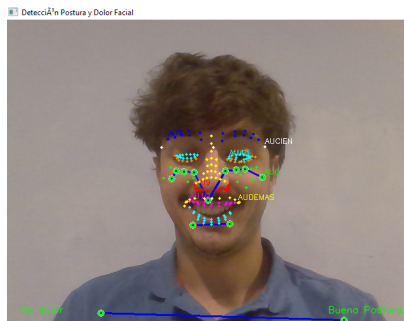


Fig. 4. Buena postura sin dolor

En la Fig. 4, el usuario se encuentra sonriendo y en una buena postura. El sistema entonces detecta correctamente la ausencia de dolor y clasifica correctamente la postura como buena.

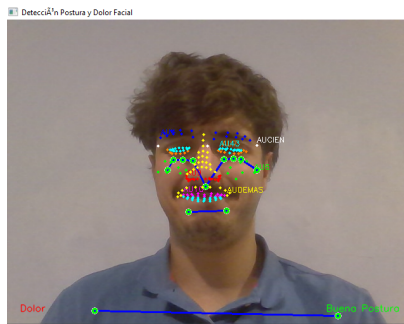


Fig. 5. Buena postura con dolor

En la Fig. 5, el usuario se encuentra en una buena postura, pero su cara muestra dolor. El sistema entonces detecta cor-

rectamente la presencia de dolor y clasifica correctamente la postura como buena.

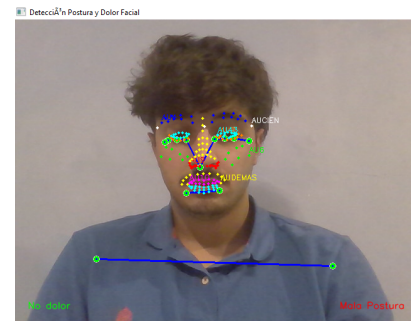


Fig. 6. Mala postura sin dolor

En la Fig. 6, el usuario se encuentra serio pero con su columna encorvada y sus hombros caídos. El sistema entonces detecta correctamente la ausencia de dolor y clasifica correctamente la postura como mala.

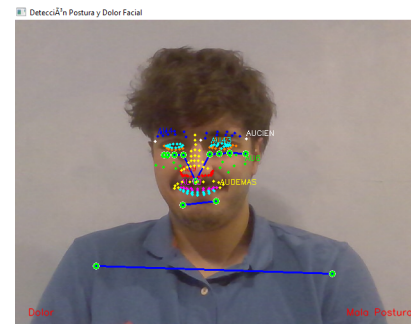


Fig. 7. Mala postura con dolor

Por último, en la Fig. 7, el usuario se encuentra con su columna encorvada y sus hombros caídos, y además, su cara muestra dolor. El sistema entonces detecta correctamente la presencia de dolor y clasifica correctamente la postura como mala.

IV. CONCLUSIONES

Se logró el objetivo planteado, incorporando valiosos conocimientos sobre la detección de dolor a partir del análisis facial y adquiriendo competencia en el uso de las bibliotecas OpenCV y MediaPipe.

Se obtuvo un modelo robusto en líneas generales, pero susceptible a una mala calibración. La precisión en el entrenamiento del modelo de predicción fue perfecta (100%), pero depende de la calidad de los datos ingresados por el usuario. Además, su capacidad para realizar la detección en tiempo real depende de los recursos de *hardware* disponibles donde se ejecuta el *software*.

En resumen, se verificó la gran utilidad del procesamiento digital de imágenes para el desarrollo de herramientas que permitan mejorar la salud y la calidad de vida de los seres humanos.

BIBLIOGRAFÍA

- [1] Z. Chen, R. Ansari, D. J. Wilkie, "Automated Pain Detection from Facial Expressions using FACS: A Review," 2018.
- [2] K. M. Prkachin "The consistency of facial expressions of pain: a comparison across modalities," 1992.
- [3] E. Weber y M. Nadeem, "PosturePal: Real-Time Posture Classification with a Laptop Webcam," [Online]. Disponible: <https://ethanweber.me/documents/posturepal.pdf>
- [4] B. Pereira, B. Cunha, P. Viana, M. Lopes, A. S. C. Melo, y A. S. P. Sousa, "A Machine Learning App for Monitoring Physical Therapy at Home," unpublished. Sensors, 2024, 24, 158.
- [5] A. H. Khaleel, T. H. Abbas, y A. S. Ibrahim, "Best low-cost methods for real-time detection of the eye and gaze tracking," 2024.