



UNIVERSIDAD DE VALLADOLID

FACULTAD DE MEDICINA

ESCUELA DE INGENIERÍAS INDUSTRIALES

TRABAJO DE FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

Valoración clínica de la parálisis facial mediante aprendizaje profundo

Autor/a:

D.^a Alicia del Hierro Diez

Tutor/a:

D. Mario Martínez Zarzuela

D. Luis Ángel Vallejo Valdezate

Valladolid, 6 de julio de 2025

TÍTULO: Valoración clínica de la parálisis facial
mediante aprendizaje profundo.

AUTOR/A: D. ^a Alicia del Hierro Diez

TUTOR/A: D. Mario Martínez Zarzuela
D. Luis Ángel Vallejo Valdezate

DEPARTAMENTO: Departamento de Teoría de la Señal y
Comunicaciones e Ingeniería Telemática
(TSCIT)

TRIBUNAL

PRESIDENTE: D. ^a María García Gadañón

SECRETARIO: D. Mario Martínez Zarzuela

VOCAL: D. Gonzalo César Gutiérrez Tobal

SUPLENTE 1: D. Carlos Gómez Peña

SUPLENTE 2: D. Jesús Poza Crespo

FECHA:
CALIFICACIÓN:

Agradecimientos

En primer lugar, quiero expresar mi agradecimiento a mis tutores, Mario y Luis Ángel, por su orientación, paciencia y apoyo a lo largo de la realización de este trabajo. Sus consejos y conocimientos han sido fundamentales para el desarrollo de este estudio.

También quiero dar las gracias a los pacientes que han querido participar en este estudio, permitiéndonos grabarles para la obtención de datos. Su colaboración desinteresada ha sido clave para llevar a cabo esta investigación.

A mi familia, en especial a mis padres, mis abuelos y mis hermanos, por ser una fuente de inspiración, por su apoyo incondicional y por animarme en cada paso del camino. Gracias por creer en mí incluso en los momentos en los que yo misma dudaba.

A mis amigos, por acompañarme en este proceso, por sus palabras de ánimo y por estar ahí en los momentos de estrés y dificultades. Su compañía ha sido un gran impulso para seguir adelante.

Y, por último, a mis compañeros de prácticas, por hacer de esta experiencia algo mucho más enriquecedor, por su ayuda cuando encontraba obstáculos y por convertir la estancia en un entorno de aprendizaje y crecimiento personal.

ÍNDICE

Abreviaturas	10
Resumen	11
Abstract.....	12
1. INTRODUCCIÓN	13
1.1 Contexto y motivación.....	13
1.2 Trabajos previos	14
1.3 Hipótesis y objetivos	15
1.4 Recursos utilizados	15
1.5 Organización de la memoria	16
2. PARÁLISIS FACIAL.....	17
2.1 Etiología.....	17
2.2 Manifestaciones clínicas	18
2.3 Clasificación	19
2.4 Escalas para la evaluación facial	19
2.5 Limitaciones de la evaluación clínica tradicional.....	22
3. ANÁLISIS COMPUTACIONAL FACIAL	23
3.1 Introducción al análisis automático.....	23
3.2 Landmarks faciales: concepto y aplicación.....	23
3.3 Herramientas existentes para la colocación de landmarks	23
3.4 Relación entre signos de parálisis y movimiento de los landmarks.....	30
4. MATERIALES Y MÉTODOS	31
4.1 Interfaz de usuario: diseño y funcionalidades.....	31
4.2 Bases de datos utilizadas.....	32
4.3 Extracción de puntos de referencia faciales con MediaPipe.....	33
4.4 Métricas: definición, cálculo e interpretación.....	33
4.5 Criterios de clasificación del grado de severidad.....	37
4.6 Flujo del análisis automático y salida de resultados.....	43
5. RESULTADOS Y DISCUSIÓN.....	45
5.1 Análisis de concordancia en las valoraciones	45
5.2 Primer intento de estimación del grado de afectación	47
5.3 Estimación final del grado de afectación	48
6. LIMITACIONES.....	50
6.1 Utilidad clínica y aportación del modelo	50
6.2 Limitaciones técnicas y clínicas del sistema.....	50
7. CONCLUSIONES Y LÍNEAS FUTURAS	52
7.1 Cumplimiento de los objetivos.....	52

7.2 Conclusiones.....	53
7.3 Líneas futuras.....	53
8. BIBLIOGRAFÍA.....	55

ÍNDICE DE FIGURAS

Figura 1. Escala Sunnybrook (Lassaletta et al., 2020).....	20
Figura 2. Escala eFace (Lassaletta et al., 2020).....	21
Figura 3. Dataset 300W con 68 puntos de referencia faciales (mmpose/docs/es/dataset_zoo/2d_face_keypoint.md en main · open-mmlab/mmpose · GitHub, s. f.)	24
Figura 4. Dataset WFLW con 95 puntos de referencia faciales (mmpose/docs/es/dataset_zoo/2d_face_keypoint.md en main · open-mmlab/mmpose · GitHub, s. f.)	24
Figura 5. Dataset COFW con 29 puntos de referencia faciales (mmpose/docs/es/dataset_zoo/2d_face_keypoint.md en main · open-mmlab/mmpose · GitHub, s. f.)	25
Figura 6. Dataset COCO-WholeBody-Face (mmpose/docs/es/dataset_zoo/2d_face_keypoint.md en main · open-mmlab/mmpose · GitHub, s. f.)	25
Figura 7. Detección de landmarks faciales en imagen de dataset público mediante MMPose.	26
Figura 8. Detección de landmarks faciales en imagen de dataset público mediante MMPose.	26
Figura 9. Detección de landmarks faciales en una imagen de dataset público mediante Face-Alignment.....	27
Figura 10. Detección de landmarks faciales en una imagen de dataset público mediante Face-Alignment.....	27
Figura 11. Detección de landmarks faciales en una imagen propia mediante Face-Alignment.	28
Figura 12. Malla con 478 puntos de referencia tridimensionales de MediaPipe (Pillai, 2024).	29
Figura 13. Malla con 478 puntos de referencia tridimensionales de MediaPipe (Guía de detección de puntos de referencia faciales Google AI Edge, s. f.).	29
Figura 14. Detección de landmarks faciales en una imagen propia mediante MediaPipe.....	30
Figura 15. Interfaz de usuario, desplegable con acciones y botón de confirmación de análisis.	31
Figura 16. Resultado del análisis que sale por pantalla.	32
Figura 17. Contenido del archivo .txt generado.....	32
Figura 18. Distancia entre landmarks seleccionados para valoración de cierre de ojos suave/fuerte.....	34
Figura 19. Distancias lineales que utiliza Nottingham System para la evaluación de parálisis	

facial (Quesada Marín et al., 2010).....	35
Figura 20. Distancia entre landmarks seleccionados para valoración de levantamiento de cejas.	35
Figura 21. Distancia entre landmarks seleccionados para valoración de levantamiento de comisuras al sonreír.	36
Figura 22. Puntos de referencia faciales que forman los labios en MediaPipe (google-ai-edge, s. f.).....	36
Figura 23. Cuestionario realizado por profesionales médicos para la evaluación clínica tradicional de la parálisis facial de 4 vídeos.	38
Figura 24. Interpretación de los valores de Fleiss' Kappa (Landis & Koch, 1977).....	45
Figura 25. Diagrama de barras de variabilidad en el diagnóstico del vídeo 1.	46
Figura 26. Diagrama de barras de variabilidad en el diagnóstico del vídeo 2.	46
Figura 27. Diagrama de barras de variabilidad en el diagnóstico del vídeo 3.	47
Figura 28. Diagrama de barras de variabilidad en el diagnóstico del vídeo 4.	47

ÍNDICE DE TABLAS

Tabla 1. Patrones faciales principales en sujetos con parálisis facial en estado de reposo.....	18
Tabla 2. Patrones principales de movimientos faciales en sujetos con parálisis facial tras distintos movimientos	18
Tabla 3. Escala de House Brackmann (Lassaletta et al., 2020).....	19
Tabla 4. Escala Yanagihara (Quesada Marín et al., 2010).....	22
Tabla 5. Conjunto de expresiones útiles para la evaluación de función facial (Lassaletta et al., 2020).....	22
Tabla 6. Resultados de la valoración del vídeo 1 por parte de los especialistas.	38
Tabla 7. Resultados de la valoración del vídeo 2 por parte de los especialistas.	39
Tabla 8. Resultados de la valoración del vídeo 3 por parte de los especialistas.	39
Tabla 9. Resultados de la valoración del vídeo 4 por parte de los especialistas.	40
Tabla 10. Matriz de frecuencias de valoraciones para cada paciente según la escala House-Brackmann.....	40
Tabla 11. Resultados de los porcentajes calculados para el vídeo 1 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	41
Tabla 12. Resultados de los porcentajes calculados para el vídeo 2 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	41
Tabla 13. Resultados de los porcentajes calculados para el vídeo 3 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	41
Tabla 14. Resultados de los porcentajes calculados para el vídeo 4 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	42
Tabla 15. Valoración conjunta de los especialistas para cada paciente según escala House-Brackmann.....	43
Tabla 16. Resultados de los porcentajes calculados para el vídeo 5 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	43
Tabla 17. Resultados de los porcentajes calculados para el vídeo 6 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.	43
Tabla 18. Resultados del primer ajuste de mínimos cuadrados.	48
Tabla 19. Resultados del ajuste de mínimos cuadrados final.	48

Abreviaturas

PF: Parálisis facial.

VIH: Virus de la inmunodeficiencia humana.

CNN: Redes neuronales convolucionales.

IA: Inteligencia artificial.

MSE: Error cuadrático medio.

Resumen

La parálisis facial (PF) es una alteración neuromuscular provocada por una lesión del VII par craneal, el nervio facial, que afecta el movimiento de los músculos del rostro, la secreción de lágrimas y saliva, y el sentido del gusto. La evaluación del grado de afectación es esencial para establecer un tratamiento adecuado, y suele realizarse mediante escalas clínicas como House-Brackman o Sunnybrook. No obstante, estas herramientas dependen de la interpretación subjetiva del profesional, lo que puede ocasionar diagnósticos inconsistentes.

En los últimos años, se han desarrollado herramientas automáticas basadas en inteligencia artificial (IA) y aprendizaje profundo para realizar esta evaluación de forma más objetiva. Muchos de estos sistemas utilizan tecnología de reconocimiento facial en tres dimensiones (3D), implementada en dispositivos móviles como el iPhone a través de su sistema de cámaras TrueDepth. Esta tecnología permite analizar la asimetría facial con gran precisión y da como resultado un porcentaje de afectación, como ocurre en la aplicación *FacialPalsyZero*. Sin embargo, estas aplicaciones presentan importantes limitaciones como el requerimiento de un hardware específico y la falta de traducción de resultados a escalas médicas estándar.

Por ello, en este estudio se ha desarrollado un sistema de evaluación basado en aprendizaje profundo que funcione con cámaras estándar presentes en la mayoría de los dispositivos. Para el desarrollo del sistema, se implementó una interfaz gráfica para la selección del vídeo y de los fragmentos del vídeo que se pretenden analizar. Tras esta selección se colocan de manera automática puntos anatómicos faciales mediante el modelo de MediaPipe y posteriormente se calculan distancias entre algunos de estos puntos que se cuantifican en forma de porcentajes. Estos porcentajes se compararon con valoraciones clínicas reales para establecer una relación entre los porcentajes y los grados de la escala de House-Brackmann. A partir de esta relación, se ajustó una ecuación que permite al sistema estimar el grado de severidad.

Los resultados obtenidos mostraron que el modelo desarrollado permite la estimación del grado de afectación según la escala House-Brackmann a partir de medidas de asimetría calculadas en imágenes 2D. Para el entrenamiento del modelo se utilizaron las valoraciones consensuadas de los especialistas para mejorar la coherencia de los diagnósticos y se obtuvo un error cuadrático medio (MSE) de 1.4587. Se comprobó que los casos en los que hay mayor diferencia entre la etiqueta real y la estimada coinciden con los vídeos en los que no se capturaron todas las acciones requeridas. Estos resultados destacan la utilidad del sistema como herramienta de apoyo para el diagnóstico siempre y cuando se siga el protocolo de adquisición de los vídeos. A pesar de que el sistema no ha podido ser validado en nuevos pacientes debido a la escasez de datos, los resultados en esta fase inicial son prometedores y sientan una buena base para su futura aplicación clínica.

PALABRAS CLAVE

Parálisis facial, marcadores faciales, aprendizaje profundo.

Abstract

Facial paralysis (FP) is a neuromuscular disorder caused by a lesion of the seventh cranial nerve, the facial nerve, which affects facial muscle movement, secretion of tears and saliva, and the sense of taste. Evaluating the severity of FP is essential for determining appropriate treatment, and is usually performed using clinical scales such as House-Brackman or Sunnybrook. However, these tools depend on the subjective interpretation of the professional, which can lead to inconsistent diagnoses.

In recent years, automatic evaluation systems based on artificial intelligence (AI) and deep learning have emerged to provide more objective assessments. Many of these systems use three-dimensional (3D) facial recognition technology, implemented in mobile devices such as the iPhone through its TrueDepth camera system. This technology allows facial asymmetry to be analyzed with great precision and results in a percentage of impairment, as is the case in the FacialPalsyZero application. However, these applications present important limitations such as the requirement of specific hardware, which restricts their accessibility, and, in addition, they do not translate the results to standard medical scales, which makes their interpretation difficult.

Therefore, this study has developed a deep learning-based evaluation system that does not rely on 3D technology, but works with standard cameras present in most devices. For the development of the system, a graphical interface was implemented for the selection of the video and video fragments to be analyzed. After this selection, the facial anatomical points are automatically placed using the MediaPipe model and then distances between some of these points are calculated and quantified in the form of percentages. These percentages were compared with real clinical assessments to establish a relationship between the percentages and the House-Brackmann scale grades. From this relationship, an equation was fitted that allows the system to estimate the degree of severity.

The results obtained showed that the model developed allows estimation of the degree of severity from asymmetry measurements calculated on 2D images. For training the model we used the consensus assessments of the doctors and we obtained a mean squared error (MSE) of 1.4587. It was found that the cases in which there is the highest difference between the actual and estimated label coincide with the videos in which not all required actions were captured. Although the system could not be validated in new patients due to the limited amount of data, the results at this early stage are promising and lay a good foundation for future clinical application.

KEYWORDS

Facial paralysis, facial *landmarks*, deep learning.

1. INTRODUCCIÓN

1.1 Contexto y motivación

La parálisis facial (PF) es una alteración parcial o completa de movimiento de todos o algunos músculos faciales, se trata de un problema relativamente frecuente que puede deberse a diferentes causas. Se puede distinguir entre PF periféricas, en las que la lesión se encuentra en el nervio facial y PF centrales, donde la lesión se encuentra en el cerebro o tronco cerebral. En las PF centrales únicamente se afecta la porción inferior de la musculatura de la cara, mientras que en las periféricas se ven afectadas todas las ramas del nervio involucrando, de este modo, a la motilidad de una hemicara al completo.

La causa más común de PF periférica es la denominada “idiopática” o parálisis de Bell, con una incidencia anual aproximada de 20-30 casos por cada 100 000 habitantes, suponiendo un 60-75% de las PF unilaterales (Lassaletta et al., 2020).

Las PF periféricas son las más comunes y tienen un cuadro muy característico. Debido a que el nervio facial contiene tanto fibras motoras como sensitivas y sensoriales, podemos encontrar signos de afectación motora como: dificultad para cerrar el ojo, boca desviada hacia un lado, surco nasogeniano borrado, etc. y signos de afectación sensitiva: alteraciones en la secreción lagrimal, en la secreción salival y del gusto en la hemilengua homolateral a la lesión motora (Garrido Calvo et al., 2000; Lassaletta et al., 2020).

Es de gran importancia realizar un correcto diagnóstico ya que es necesario determinar el tratamiento lo antes posible para evitar secuelas permanentes. Para evaluar la PF en la actualidad existen diferentes escalas, entre ellas destacamos las más utilizadas: la escala de House-Brackmann (Lassaletta et al., 2020) y la escala de Sunnybrook (Lassaletta et al., 2020). La escala de House-Brackmann clasifica la función facial en seis grados, desde la normalidad (grado I) hasta la parálisis completa (grado VI). Es ampliamente utilizada debido a su simplicidad, pero tiene una resolución limitada y no permite la detección de sincinesias. Por otro lado, la escala de Sunnybrook realiza una evaluación más detallada, combinando una puntuación de simetría en reposo, movimiento voluntario y presencia de sincinesias, y proporciona como resultado una puntuación total de 0 a 100. No obstante, ambas presentan la problemática de que dependen de la valoración subjetiva de los especialistas, lo que puede derivar en diagnósticos inconsistentes debido a la variabilidad interobservador y a la experiencia clínica del evaluador (Lassaletta et al., 2020).

La falta de objetividad ha impulsado en los últimos años el desarrollo de sistemas automáticos de evaluación facial mediante tecnologías digitales, con el objetivo de estandarizar el proceso diagnóstico y reducir el margen de error. Diversos estudios han desarrollado sistemas que utilizan las cámaras 3D de los teléfonos de última generación para la evaluación automática de la PF. En estos estudios destaca la creación de aplicaciones que utilizan la cámara TrueDepth de Apple ® para capturar y analizar los movimientos faciales con alta precisión, permitiendo detectar asimetrías faciales mediante medición fotográfica tridimensional. Estos sistemas representan un gran avance ya que ofrecen herramientas accesibles para la realización de diagnósticos objetivos de la patología. Sin embargo, presentan varias limitaciones importantes: requieren obligatoriamente de dispositivos con cámaras 3D, lo cual restringe su disponibilidad y, en segundo lugar, los resultados que proporcionan no se expresan en escalas médicas estandarizadas, lo que dificulta la interpretación de los resultados por los profesionales sanitarios. Por tanto, es necesario desarrollar herramientas más accesibles y clínicamente relevantes que permitan diagnósticos objetivos (Andrews et al., 2023; Hasebe et al., 2024; Taeger et al., 2021).

En este trabajo se propone un modelo computacional para estimar el grado de severidad de la

parálisis facial a partir de videos faciales convencionales. A diferencia de otros, este modelo no requiere un hardware especializado, su funcionamiento se basa en la utilización de videos 2D sin profundidad, que pueden ser obtenidos con cámaras comunes, como las de los dispositivos móviles, lo que facilita su uso en entornos clínicos. El modelo analiza las estructuras faciales relevantes y proporciona como resultado una estimación porcentual sobre el grado de afectación según la escala House-Brackmann. De este modo, el modelo propuesto trata de ser una herramienta útil, precisa y accesible para la evaluación objetiva de la gravedad de personas afectadas de parálisis facial.

1.2 Trabajos previos

El estudio del movimiento facial es realmente importante, ya que una evaluación adecuada es clave para determinar la gravedad de trastornos como la parálisis facial. Esto, a su vez, permite diseñar tratamientos personalizados que pueden mejorar la calidad de vida de los pacientes.

El estudio realizado por Gómez Cuarán e Higuera Castro (2019) presenta un metaanálisis de la documentación y las aplicaciones disponibles relacionadas con la PF. En este estudio, se revisaron 11 artículos que trataban sobre aplicaciones con fines diagnósticos, informativos y de rehabilitación. De esos once artículos, solo dos se enfocaban en el diagnóstico de la enfermedad, mientras que la mayoría se centraba en aspectos de rehabilitación. Además, los autores hicieron una búsqueda en Google Play y App Store, encontrando 13 aplicaciones disponibles; sin embargo, solo tres de ellas estaban destinadas al diagnóstico de la PF, lo que pone de manifiesto la escasez de herramientas para realizar diagnósticos automáticos.

Siguiendo con el desarrollo de aplicaciones que faciliten el diagnóstico de enfermedades, Hasebe et al. (2024) presentaron una aplicación móvil diseñada para evaluar la PF. Esta aplicación utilizaba el reconocimiento facial del iPhone y técnicas de aprendizaje automático para ofrecer una medición objetiva de la función motora facial. La aplicación analizaba los videos de los pacientes y generaba una puntuación automática basada en distintos parámetros faciales. Después, el estudio realizaba una correlación con algunas escalas medicas convencionales (Yanagihara y House-Brackmann) para comprobar la precisión de sus resultados, los cuales fueron prometedores.

Taeger et al. (2021) también presentaron el desarrollo y evaluación de una aplicación para la cuantificación rápida y objetiva de asimetrías faciales. Esta aplicación utilizaba, al igual que el anterior estudio, la cámara tridimensional de los dispositivos iPhone para la evaluación de la simetría facial. La aplicación proporcionaba un resultado numérico en forma de porcentaje para cada mitad facial. No obstante, a pesar de que los resultados obtenidos fueron destacables, la aplicación no fue probada en pacientes con parálisis, por lo que sugirieron una evaluación que incluyese a pacientes con la enfermedad para validar la utilidad de la aplicación en contextos clínicos reales.

Katsumi et al. (2015) realizaron un estudio en el que destacaban limitaciones de las mediciones en 2D a la hora de diagnosticar de forma automática la PF. Por ello realizaron un sistema de medición tridimensional cuyos resultados fueron comparados posteriormente con escalas tradicionales. Los resultados mostraron que los sistemas 2D y 3D eran bastante similares en áreas de los ojos y cejas, pero el sistema 3D conseguía mayor precisión en la zona bucal. Aunque los hallazgos fueron significativos, el sistema propuesto necesitaba un hardware especializado, lo que limita mucho su accesibilidad.

Quesada Marín et al. (2010) realizaron un análisis sobre las escalas de valoración de la PF, donde distinguieron entre escalas subjetivas y escalas objetivas. Entre las escalas subjetivas están la escala House Brackmann, la escala Sunnybrook, la escala Yanagihara y la escala Peitersen, todas ellas con problemas de variabilidad interobservador. Y entre las escalas que

son más objetivas destacaron la escala Nottingham, que se basa en la medida de distancias entre distintos puntos faciales. Esta escala propone una manera interesante y alternativa para mejorar la objetividad de los diagnósticos, sin embargo, no es un sistema automatizado.

Teniendo en cuenta la revisión de los trabajos anteriores, observamos que existe una necesidad de herramientas de evaluación de PF que sean capaces de combinar tanto precisión como facilidad de acceso. El presente Trabajo Fin de Grado (TFG) desarrolla un software que, a través del modelo de puntos de referencia faciales de la librería MediaPipe (Lugaresi et al., 2019), es capaz de analizar vídeos 2D adquiridos con cualquier dispositivo. Este sistema busca la integración de la accesibilidad y la fiabilidad diagnóstica basándose en la escala House-Brackmann, sin necesidad de un hardware especializado o cámaras 3D usadas en otros estudios.

1.3 Hipótesis y objetivos

En este TFG se plantea la hipótesis de que es posible realizar un análisis facial mediante dispositivos de bajo coste y herramientas de software públicas que tenga suficiente relevancia clínica como para ser utilizado en ayuda al diagnóstico de la parálisis facial reduciendo así la subjetividad de los profesionales.

El objetivo principal de este proyecto es desarrollar un sistema que permita la evaluación semiautomática de la parálisis facial según escalas médicas estandarizadas. Esta herramienta debiera conseguir reducir la variabilidad interobservador y también disminuir el número de diagnósticos erróneos.

Con el fin de alcanzar este objetivo se ha dividido este Trabajo Fin de Grado en subobjetivos específicos. Estos son los siguientes:

- i. Realizar una búsqueda bibliográfica para conseguir una completa documentación acerca del análisis facial tradicional y automático en el panorama actual de la parálisis facial.
- ii. Crear una interfaz gráfica en Python que permita seleccionar y recorrer los *frames* de un vídeo del paciente y contestar a ciertas preguntas.
- iii. Crear una herramienta en Python que permita tomar medidas a partir de los puntos de referencia faciales marcados por el modelo de MediaPipe y estime en base a ellas la severidad de la enfermedad.
- iv. Analizar y comparar los resultados alcanzados con los que se obtendrían mediante una evaluación tradicional.
- v. Extraer conclusiones basadas en los datos obtenidos, considerando las limitaciones y proponiendo posibles líneas futuras.

1.4 Recursos utilizados

Durante el desarrollo de este trabajo se hizo uso de los siguientes recursos de hardware y software:

Software:

- PyCharm 2024.3.5.
- Mediapipe 0.10.21, en específico el modelo Face Landmarker v0.4.0.
- Python 3.11.

Hardware:

- Ordenador con procesador i9, 32 GB de memoria y una tarjeta gráfica NVIDIA GeForce RTX 2080 Ti de 11GB.
- Ordenador con procesador i5 y 8GB de memoria.

1.5 Organización de la memoria

Este documento está dividido en 7 capítulos. A lo largo del primer capítulo introductorio se han tratado los conceptos básicos sobre los que se sustenta este TFG, junto con el contexto actual y el problema que se aborda, la hipótesis planteada, los objetivos y los recursos empleados. A continuación, se ofrece un resumen del contenido de los capítulos restantes:

En el **Capítulo 2 “Parálisis facial”** presenta las principales características de esta enfermedad, resaltando su etiología, síntomas, clasificación según diferentes criterios, las principales escalas para su evaluación y las limitaciones de estas escalas.

En el **Capítulo 3 “Análisis computacional facial”** se realiza un análisis de los modelos existentes para la colocación de puntos de referencia faciales y como estos varían según la presencia o no de parálisis facial.

En el **Capítulo 4 “Materiales y métodos”** se describe el modelo de colocación de puntos y la base de datos utilizadas para este trabajo. También se detalla la interfaz creada y las métricas extraídas de los puntos de referencia faciales.

En el **Capítulo 5 “Resultados y discusión”** se exponen los resultados obtenidos para todas las pruebas realizadas y se interpretan.

En el **Capítulo 6 “Limitaciones”** en este punto se analizan la aplicabilidad clínica del estudio, se analizan las limitaciones e inconvenientes encontrados durante el trabajo, así como la identificación de los puntos débiles del sistema desarrollado.

En el **Capítulo 7 “Conclusión y líneas futuras”** se presentan las conclusiones extraídas y se sugieren las posibles líneas de mejora futuras que darían continuidad al estudio.

2. PARÁLISIS FACIAL

La parálisis facial (PF) se caracteriza por la afectación del VII par craneal, responsable de la innervación motora, sensorial y autonómica de la región facial. Habitualmente suele aparecer en único lado de la cara, aunque podría producirse en ambos. Para su abordaje clínico es fundamental la diferenciación de su causa, manifestaciones clínicas y la correcta aplicación de las técnicas de diagnóstico.

2.1 Etiología

La PF puede deberse a múltiples causas, para facilitar su comprensión vamos a dividirlas en dos grupos: congénitas y adquiridas.

Congénitas

Son aquellas parálisis que están presentes desde el momento del nacimiento, es decir no se adquieren de forma posterior al nacimiento. En este grupo podemos destacar diferentes causas:

- Idiopáticas: aquellas de las que no sabemos la causa exacta.
- Sindromáticas: Aquellas asociadas a síndromes genéticos. Destacamos el síndrome de Moebius, la microsomía hemifacial y el síndrome de Goldenhar.
- No sindromáticas: No asociadas a síndromes genéticos, como la parálisis congénita del labio inferior.
- Traumáticas: trauma obstétrico.

Adquiridas

Son aquellas parálisis faciales que aparecen de forma posterior al nacimiento, están provocadas por sucesos que ocurren después del nacimiento y provocan PF. Podemos dividir estas causas en varios grupos:

- Infecciosas: pueden deberse a virus o a bacterias. En el caso de las víricas destacamos la parálisis de Bell, el herpes zóster, VIH, y la hepatitis. En las bacterianas destaca la enfermedad de Lyme.
- Tumorales: tumores del hueso temporal, tumores intrínsecos del nervio facial, tumor de parótida, metástasis, meningioma...
- Iatrogénicas: aquellas parálisis causadas por intervenciones quirúrgicas o médicas. Pueden deberse a cirugías de oído, de parótida o del ángulo pontocerebeloso.
- Traumáticas: parálisis ocasionadas como consecuencia de lesiones o accidentes. Fracturas del hueso temporal o herida facial.
- Neurológicas: debidas a enfermedades del sistema nervioso. Se destaca la esclerosis múltiple, miastenia gravis, síndrome de Guillain-Barré, accidentes cerebrovasculares y la distrofia miotónica.
- Otógenas: aquellas debidas a infecciones o lesiones del oído. Pueden causarlas las otitis y los colesteatomas.
- Enfermedades generales: como la diabetes, sarcoidosis, amiloidosis...

A pesar de que hay numerosas causas que pueden provocar PF, la más común es la parálisis de Bell, ocasiona un 60-75% de las parálisis faciales unilaterales. Otras de las causas más comunes después de la parálisis de Bell son las causas iatrogénicas, el herpes zoster, el cáncer de cuello y cabeza y los traumatismos (Benítez S. et al., 2016; Lassaletta et al., 2020).

2.2 Manifestaciones clínicas

El nervio facial tiene fibras motoras, sensitivas y autónomas, por ello tras su afectación nos encontraremos con diversas manifestaciones.

En la mayoría de los casos de parálisis facial periférica el síntoma inicial es dolor en la zona posterior del oído y tras un periodo de 2 a 3 días se desarrollan el resto de los síntomas entre los que destacamos (Lassaletta et al., 2020; National Library of Medicine, s. f.; Rubin, 2023):

- Síntomas motores: sensación de entumecimiento de la cara y debilidad muscular del lado afectado, lo que produce ausencia del surco nasogeniano y de arrugas en la frente. Dificultad para cerrar el ojo paralizado, al intentar cerrarlo se puede producir una desviación del globo ocular hacia arriba y hacia dentro. Caída del párpado inferior y su eversion. Problemas para sonreír, gesticular o comer.
- Síntomas sensoriales: Pérdida del sentido del gusto.
- Síntomas autonómicos: Alteraciones en la secreción salival y lagrimal, lo que ocasiona resequedad en el ojo afectado.

A continuación, se presenta un resumen de los principales signos observables, para ello se realiza una comparación entre lo que se espera observar en el lado enfermo y en el lado sano en situaciones de reposo y de movimiento. Esta comparación será útil para una correcta comprensión de las medidas que se proponen posteriormente para la realización del diagnóstico.

REPOSO	Lado sano	Lado enfermo
Ojo	Ojo más abierto	Ojo más cerrado Párpado inferior caído
Ceja	Ceja más alta	Ceja más baja
Pliegue nasogeniano	Muy marcado	Poco marcado
Boca	Comisura más alta e inclinada hacia este lado	Comisura un poco más baja

Tabla 1. Patrones faciales principales en sujetos con parálisis facial en estado de reposo.

MOVIMIENTO	Lado sano	Lado enfermo
Cierre ocular suave	Cierre total del ojo	Puede no producirse el cierre total Cierre del ojo de forma más lenta Desviación del globo ocular hacia dentro y hacia arriba
Cierre ocular fuerte	Descenso de la ceja Elevación de la comisura bucal Ojo fuertemente cerrado	No hay especial movimiento en ceja o boca Ojo ligeramente más cerrado que al cerrar suave
Elevación de cejas	Ascensión de la ceja Aparición de arrugas en la frente	Ceja permanece casi en el mismo sitio No aparecen arrugas en la frente
Sonreír con dientes	Comisura bucal se eleva Ojo se cierra ligeramente	Comisura casi no se mueve No se cierra ojo No se consigue enseñar dientes
Abrir la boca	Apertura normal Boca se inclina hacia este lado	No se consigue la apertura de la boca

Tabla 2. Patrones principales de movimientos faciales en sujetos con parálisis facial tras distintos movimientos.

2.3 Clasificación

Podemos clasificar la parálisis facial en base a distintos criterios (Benítez S. et al., 2016; *Características de los grados de afectación del nervio facial y la discapacidad facial en pacientes con parálisis facial periférica de un año de evolución.*, s. f.; Garrido Calvo et al., 2000; Lagarde & Améri, 2012; Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello, 2022):

- Según su localización: periféricas, en las que la lesión se encuentra en el nervio facial, y centrales, la lesión se encuentra en el sistema nervioso central. En las PF centrales únicamente se afecta la porción inferior de la musculatura de la cara, mientras que en las periféricas se ven afectadas todas las ramas del nervio.
- Según su lateralidad: unilateral, si solo se ve afectado un lado de la cara, o bilateral, si están afectados ambos lados.
- Según su gravedad: esta clasificación depende de la escala con la que se evalúe, se explicarán más adelante.
- Según su duración: aguda, empeora entre las primeras 24 y 72 horas, recuperándose por completo en un periodo inferior a 6 meses; subaguda, el periodo de recuperación puede prolongarse hasta un año; y crónica, aquellas parálisis que, pasado un año, no se han recuperado.

2.4 Escalas para la evaluación facial

La evaluación de la función facial se puede realizar a través de varias escalas, estas tratan de evaluar la función estática y dinámica, pero tienen diferencias entre ellas.

Escala House-Brackmann

Es una de las escalas más usadas debido a su simplicidad. Evalúa la función facial en estado de reposo y alterado y establece 6 categorías según su gravedad:

Reposo	Grado	Disfunción	Cierre ocular y función frontal
Normal	I	Normal	Cierre ocular completo y función frontal
	II	Leve	Debilidad muscular
	III	Moderada	Asimetría facial no desfigurante
	IV	Moderada-grave	Cierre ocular incompleto y sin función frontal
Alterado	V	Grave	Escaso movimiento facial
	VI	Completa	Sin movimiento facial

Tabla 3. Escala de House Brackmann (Lassaletta et al., 2020).

El problema de esta escala es que no es sensible a sincinesias ni a cambios producidos en la rehabilitación del paciente.

Escala Sunnybrook

En contraste con la escala anterior, esta sí que es capaz de detectar los cambios en la recuperación debido a que al ser una escala continua es bastante más sensible, por ello, es muy usada por los médicos rehabilitadores. Analiza la simetría en reposo, en movimiento y las sincinesias de forma independiente y da una puntuación entre 0 y 100, siendo 0 la pérdida total de función y 100 una función normal.

ESCALA DE SUNNYBROOK														
SIMETRÍA EN REPOSO			SIMETRÍA EN MOVIMIENTO						SINCINESIAS					
Comparada con el lado sano			Grado de excursión muscular comparado con el lado sano						Grado de contracción muscular involuntaria asociada a cada expresión					
Ojo (elegir solo una puntuación)			EXPRESIONES ESTÁNDAR	Sin movimiento	Inicio discreto de movimiento	Inicio movimiento, amplitud leve	Movimiento casi normal	Movimiento normal		No	Leve	Moderada no desfigurante	Severa desfigurante, contracción en masa	
Normal	0													
Abierto	1													
Cerrado	1													
Cirugía del párpado	1													
Pliegue nasogeniano			FRO	1	2	3	4	5		0	1	2	3	
Normal	0		OCS	1	2	3	4	5		0	1	2	3	
Ausente	2		ZYG / RIS	1	2	3	4	5		0	1	2	3	
Menos pronunciado	1		LLA / LLS	1	2	3	4	5		0	1	2	3	
Más pronunciado	1		OOS / OOI	1	2	3	4	5		0	1	2	3	
Boca				Asimetría grave	Asimetría severa	Asimetría moderada	Asimetría leve	Simetría normal						
Normal	0													
Comisura caída	1													
Com. elevada/tracción	1													
TOTAL =			TOTAL =						TOTAL =					
Puntuación de la simetría en reposo total x 5 =			Puntuación del movimiento voluntario total x 4 =						Puntuación de las sincinesias total x 4 =					
PUNTUACIÓN FINAL														
Movimiento voluntario – Simetría reposo – Sincinesias =														
Expresiones estándar: FhRO: arrugar la frente, OCS: cierre ocular fuerte, ZYG / RIS: abrir la boca y sonreír, LLA / LLS: arrugar la nariz, Gruñir, OOS / OOI: poner «morritos» /beso														

Figura 1. Escala Sunnybrook (Lassalletta et al., 2020)

Escala eFace

Esta escala es de las más recientes. Se trata de una escala analógica-visual en formato digital que tiene 15 medidas, 4 estáticas, 7 dinámicas y 4 para sincinesias. A cada uno de los parámetros se le da una puntuación entre 0 y 100, siendo 0 parálisis total y 100 función normal. Si la patología es por exceso de movimiento la escala va de 0 a 200, esto es útil por ejemplo para evaluar la altura de la ceja en reposo. Es posible evaluar los resultados de esta escala tanto mediante números como a través de gráficos, examinando cada sección por separado y también obteniendo una calificación que englobe todo.

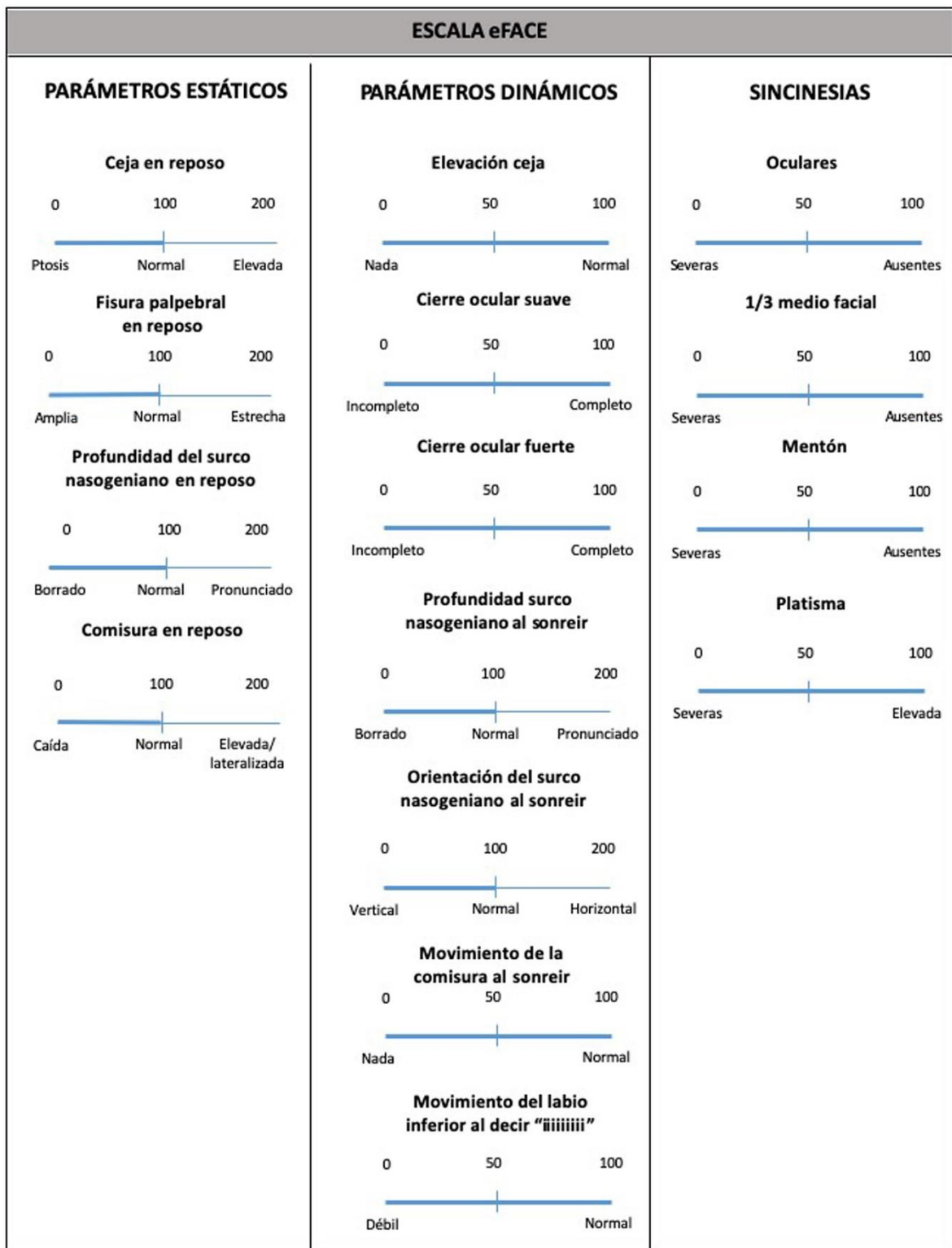


Figura 2. Escala eFace (Lassaletta et al., 2020).

Escala Yanagihara

Es la escala más usada en Japón, esta escala realiza la valoración de diferentes áreas de la cara y les asigna una puntuación que será 0 si la función es normal, 2 si hay cierto grado de parálisis o 4 si no se produce movimiento, y al sumarlas se obtiene un resultado global.

	Normal	Parálisis parcial/debilidad	No hay movimiento
En reposo	4	2	0
Fruncir el ceño	4	2	0
Arrugar zona entre las cejas	4	2	0
Cerrar ojos suave	4	2	0
Cerrar ojos fuerte	4	2	0
Cerrar ojo del lado patológico	4	2	0
Inflar carrillos	4	2	0
Silbar	4	2	0
Sonreír	4	2	0
Bajar labio inferior	4	2	0

Tabla 4. Escala Yanagihara (Quesada Marín et al., 2010).

Independientemente de la escala utilizada es importante que la evaluación se realice de manera sistemática. Por ello, Lassaletta et al. proponen un total de 10 expresiones estándar a partir de las cuales se podría realizar la evaluación de la función facial.

Expresiones estándar para valorar la función facial	
1. Reposo	2. Elevar cejas
3. Cerrar ojos suavemente	4. Cerrar ojos fuerte
5. Arrugar nariz	6. Sonrisa boca cerrada
7. Sonrisa enseñando dientes	8. Fruncir labios (poner morritos)
9. Decir “iiiiii” con dientes cerrados	10. Vista nasal inferior

Tabla 5. Conjunto de expresiones útiles para la evaluación de función facial (Lassaletta et al., 2020).

Teniendo en cuenta estas expresiones, para nuestro sistema realizamos la adquisición de vídeos en los que se pidió a los pacientes que realizaran los siguientes movimientos o acciones: reposo, levantar cejas, cerrar ojos de forma suave, cerrar ojos de forma fuerte y sonreír con dientes.

2.5 Limitaciones de la evaluación clínica tradicional

La utilización de las escalas mencionadas previamente requiere gran experiencia y formación para poder realizar diagnósticos lo más precisos posibles, esto puede limitar el número de profesionales que puedan evaluar la función facial.

Además, el principal problema que tienen es la variabilidad entre observadores. Estas escalas son subjetivas, el profesional clínico asignará un valor a cada movimiento según su criterio y experiencia, el problema es que en muchos casos estos criterios no coinciden entre distintos evaluadores, lo que puede dar lugar a distintos diagnósticos y por tanto producir diagnósticos erróneos y retrasos en el inicio de tratamientos (Quesada Marín et al., 2010).

3. ANÁLISIS COMPUTACIONAL FACIAL

Durante este capítulo se describen los fundamentos del análisis automático de la simetría facial. Se realiza una introducción sobre el concepto de marcador de punto anatómico facial (*landmark*) y se analizan las herramientas encontradas para su detección y colocación. Por último, se describe el desplazamiento que harán los *landmarks* en presencia de parálisis facial, lo que sienta las bases para el desarrollo del sistema de evaluación objetiva.

3.1 Introducción al análisis automático

El análisis automático permite detectar, registrar y evaluar cuantitativamente distintas métricas mediante algoritmos de visión por computadora e inteligencia artificial. Estos algoritmos de aprendizaje profundo se basan especialmente en redes neuronales convolucionales (CNN) capaces de aprender características cada vez más complejas y abstractas a medida que la imagen avanza por la red.

Recientemente, se han producido grandes avances en análisis de imágenes, lo que combinado con el continuo avance de dispositivos capaces de capturar imágenes o vídeos de alta resolución a dado como resultado el desarrollo de técnicas y algoritmos simples y eficientes para el ámbito médico. Varias enfermedades han utilizado estos algoritmos para asistir a los profesionales sanitarios a la hora de realizar diagnósticos debido a que permiten una evaluación objetiva, reproducible y coherente (Su et al., 2021; Vrochidou et al., 2023).

En el caso de la parálisis facial, el análisis automático solucionaría las limitaciones mencionadas previamente sobre los métodos tradicionales. Se ha demostrado la eficacia de las CNN para la detección de parálisis facial en varios estudios basándose en el análisis de la simetría facial. Muchos de estos estudios utilizan la colocación de puntos de referencia faciales para la valoración de simetría (Ten Harkel et al., 2023; Vrochidou et al., 2023). También destacan los estudios que utilizan tecnologías tridimensionales para el análisis de vídeos. Estos sistemas consiguen bastante precisión a la hora de detectar estructuras faciales, sin embargo, su uso se restringe a aquellos dispositivos que dispongan de cámaras tridimensionales, es decir, requieren un hardware específico (Andrews et al., 2023; Hasebe et al., 2024; Taeger et al., 2021).

3.2 Landmarks faciales: concepto y aplicación

Los *landmarks* faciales son puntos de referencia anatómicos que sirven para la localización de estructuras relevantes del rostro, como los ojos, las cejas, los labios y la nariz. Estos puntos son clave para poder realizar análisis cuantitativos del movimiento y la simetría facial.

En el ámbito médico, en especial en el diagnóstico de la PF por computadora, la colocación automática de los puntos de referencia faciales permite detectar cambios sutiles en las expresiones faciales, analizar la simetría facial y cuantificar la diferencia de movimiento de ambos lados de la cara. Esto es posible gracias a que disponer de la posición de los *landmarks* nos permite medir distancias, ángulos y áreas entre ellos que posteriormente serán utilizados para comparar entre el lado sano y el lado enfermo y tomar decisiones sobre el grado de severidad de la enfermedad (*Guía de detección de puntos de referencia faciales* | Google AI Edge | Google AI for Developers, s. f.; Rosebrock, 2017).

3.3 Herramientas existentes para la colocación de landmarks

Existen varios modelos públicos y preentrenados para la colocación automática de *landmarks* faciales. A lo largo de este punto se describen tres de ellos.

MMPose

Se trata de un kit de herramientas de código abierto que permite la estimación de poses basado en PyTorch. Está formado por un conjunto de algoritmos que permiten la estimación de poses de múltiples personas en 2D, estimación de puntos de referencia faciales en 2D, estimación de poses de manos en 2D, estimación de poses humanas de cuerpo entero, estimación de poses de animales, detección de puntos de referencia de moda. MMPose tiene una arquitectura modular, por lo que permite seleccionar distintos modelos como base, así como distintos datasets de entrenamiento.

La configuración de MMPose como modelo de puntos de referencia faciales permite la colocación automática de *landmarks* en imágenes o vídeos, como se observa en la Figura 3. Esta configuración tiene una serie de datasets compatibles y dependiendo de cuál se elija se colocarán un número de *landmarks* específico, distribuidos entre la ceja, ojos, nariz, mandíbula y boca, observar en las Figuras 4, 5 y 6. Cuando el modelo se prueba con una imagen propia, se colocan por defecto 68 puntos de referencia faciales.

Los datasets compatibles con detección facial son:

- 300W (300 imágenes y 68 puntos de referencia faciales)
- 300VW (para vídeo)
- WFLW (10000 imágenes y 98 puntos de referencia faciales)
- AFLW (25000 imágenes y 21 puntos de referencia faciales)
- COFW (1000 imágenes y 29 puntos de referencia faciales)
- COCO-WholeBody-Face
- LaPa (8000 imágenes y 98 puntos de referencia faciales)

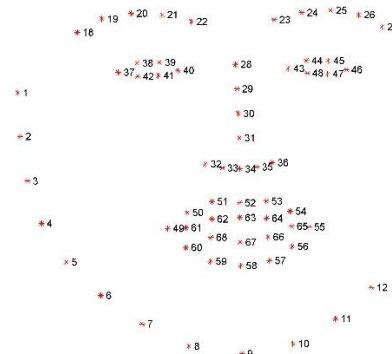


Figura 3. Dataset 300W con 68 puntos de referencia faciales
([mmpose/docs/es/dataset_zoo/2d_face_keypoint.md](#) en main · open-mmlab/mmpose · GitHub, s. f.)

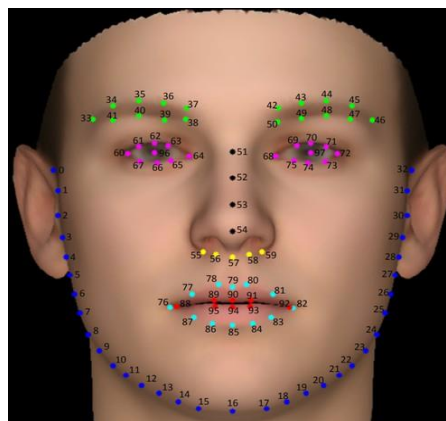


Figura 4. Dataset WFLW con 95 puntos de referencia faciales
([mmpose/docs/es/dataset_zoo/2d_face_keypoint.md](#) en main · open-mmlab/mmpose · GitHub, s. f.)

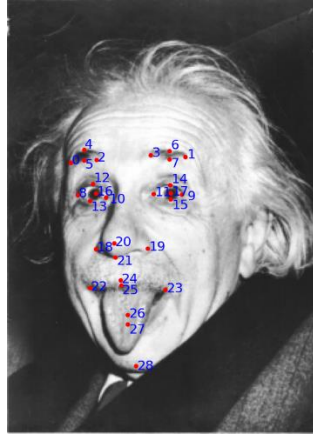


Figura 5. Dataset COFW con 29 puntos de referencia faciales
([mmpose/docs/es/dataset_zoo/2d_face_keypoint.md](#) en main · open-mmlab/mmpose · GitHub, s. f.)

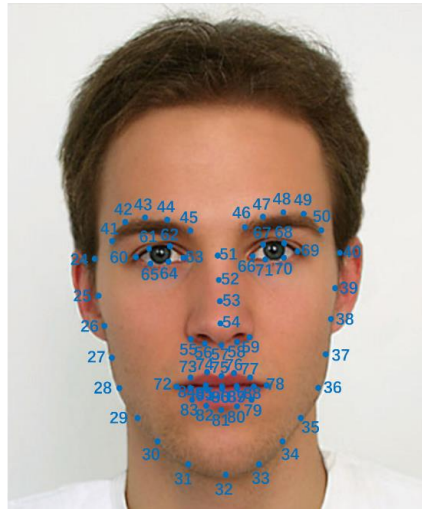


Figura 6. Dataset COCO-WholeBody-Face
([mmpose/docs/es/dataset_zoo/2d_face_keypoint.md](#) en main · open-mmlab/mmpose · GitHub, s. f.)

MMPose permite la visualización de resultados, además dibuja un mapa de calor que da información sobre la precisión con la que han sido colocados los *landmarks* (*Descripción general: documentación de MMPose 1.3.2, s. f.; mmpose/docs/es/dataset_zoo/2d_face_keypoint.md* en main · open-mmlab/mmpose · GitHub, s. f.).

A continuación, se muestra algún ejemplo de la aplicación de esta herramienta para la colocación de puntos faciales en distintas imágenes.

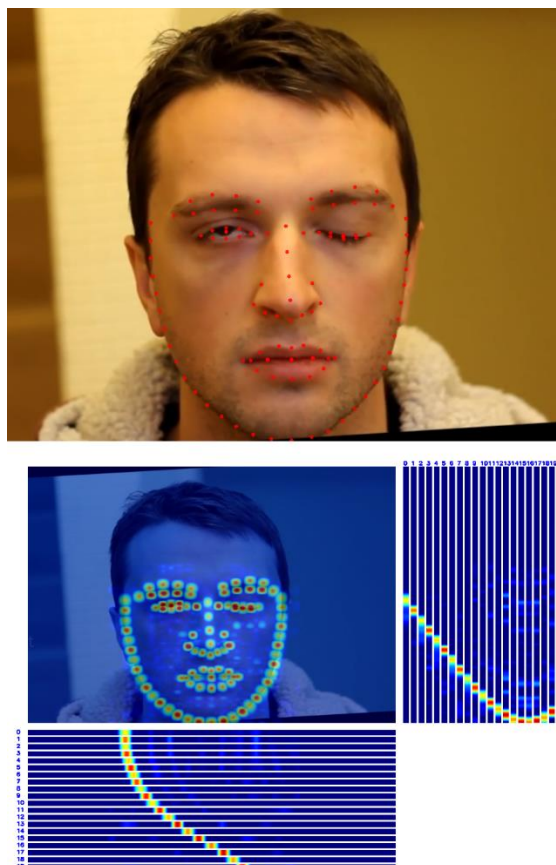


Figura 7. Detección de landmarks faciales en imagen de dataset público mediante MMPose.

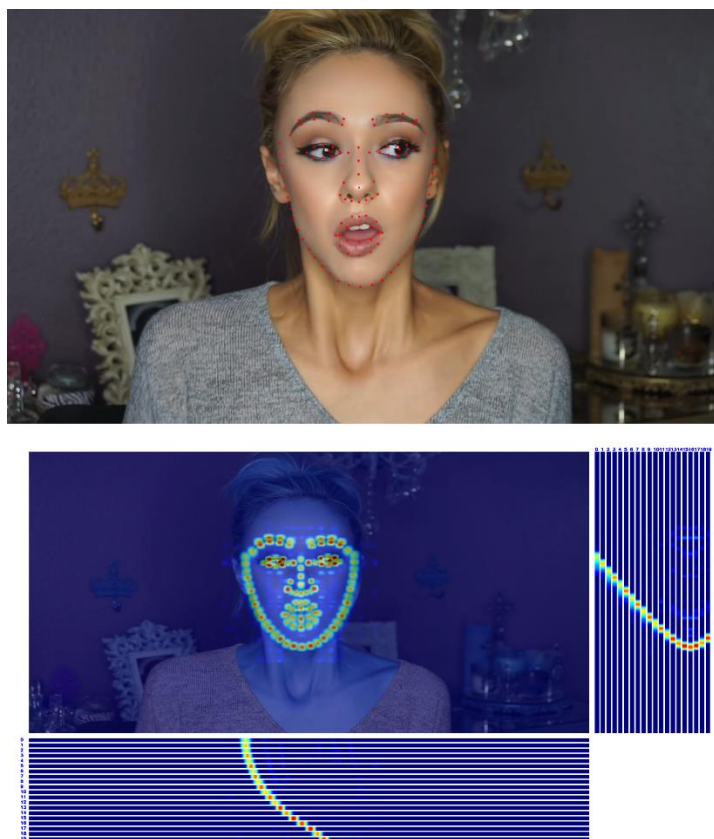


Figura 8. Detección de landmarks faciales en imagen de dataset público mediante MMPose.

Podemos observar que el modelo es capaz de colocar de forma bastante precisa las estructuras faciales principales (ojos, cejas, nariz, boca y mandíbula). Sin embargo, tras probarlo en nuestros sujetos se ha comprobado que en las situaciones en las que la boca se encuentra abierta, la detección de los labios no es muy exacta, especialmente en el interior de los labios. Esto puede deberse a que el modelo ha sido entrenado en personas sanas y, al probarlo con enfermos que tienen un grado más elevado de parálisis, no es capaz de colocar los puntos de referencia de forma correcta.

Face- Alignment

Otra herramienta para la detección de puntos de referencia faciales a través de Python es capaz de estimar *landmarks* faciales en 2D y 3D a partir de imágenes o de vídeos.

Este modelo está basado en redes neuronales profundas y permite detectar 68 puntos de referencia, distribuidos en las cejas, ojos, labios, nariz y contorno facial.

Al igual que el modelo MMPose, nos permite visualizar los resultados, pero, además, este modelo permite obtener un gráfico 3D interactivo con los *landmarks* colocados (*GitHub - 1adrianb/face-alignment: :fire: Biblioteca de alineación de rostros 2D y 3D creada con PyTorch*, s. f.).

A continuación, se muestran varios ejemplos en los que se utiliza este modelo.

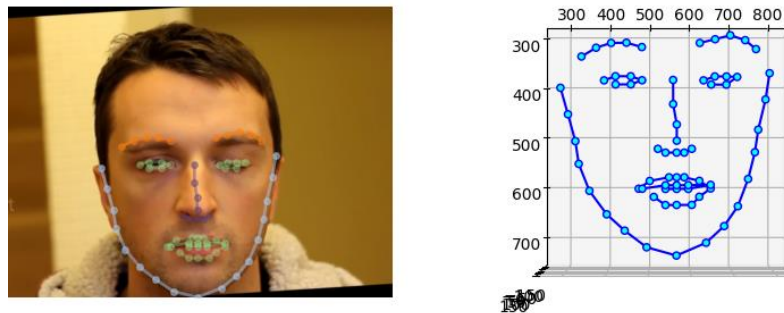


Figura 9. Detección de landmarks faciales en una imagen de dataset público mediante Face-Alignment.



Figura 10. Detección de landmarks faciales en una imagen de dataset público mediante Face-Alignment.

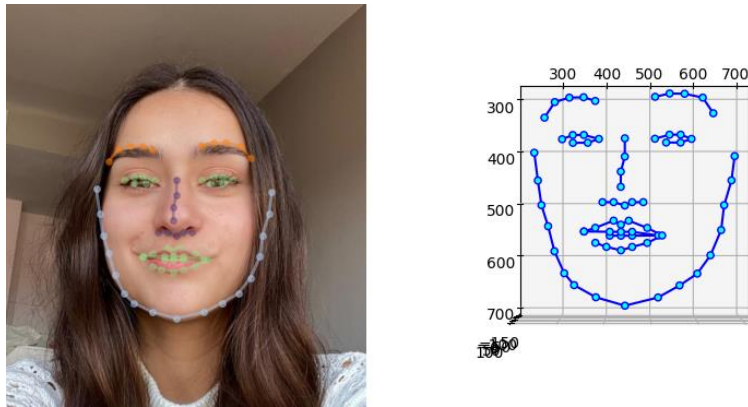


Figura 11. Detección de landmarks faciales en una imagen propia mediante Face-Alignment.

En comparación con el modelo MMPose este modelo es menos preciso. En la mayoría de las imágenes podemos ver que la detección de las cejas no es del todo correcta, ya que suele situarlas más arriba de lo que realmente están. Además, la detección de los ojos la realiza de forma bastante errónea, en el primer ejemplo el enfermo tiene un ojo cerrado y el modelo lo detecta como si estuviese abierto. Por otro lado, al probar estos modelos con las imágenes de nuestros pacientes, hemos comprobado que en situaciones de apertura de boca el modelo no es capaz de detectar el contorno de la cara y produce una deformación de todos los *landmarks*.

MediaPipe Solutions

Se trata de un paquete de bibliotecas y herramientas de código abierto desarrollado por Google y utilizado para aplicar técnicas de inteligencia artificial y aprendizaje automático de forma rápida. Las principales soluciones que tiene MediaPipe son para el cuerpo, la cara, manos, audio y lenguaje. De todas ellas destacamos para nuestro objetivo la tarea disponible de detección de puntos de referencia faciales (Face Landmarker). Esta tarea usa 3 modelos de aprendizaje automático, el primer modelo es el encargado de detectar rostros, el segundo modelo coloca una malla con 478 puntos de referencia tridimensionales en cada rostro y por último, el tercer modelo recibe los resultados de la malla y calcula 52 puntuaciones que representan las expresiones faciales. Puede funcionar con imágenes individuales, fotogramas de vídeo o un vídeo en vivo (*Guía de detección de puntos de referencia faciales | Google AI Edge, s. f.*).

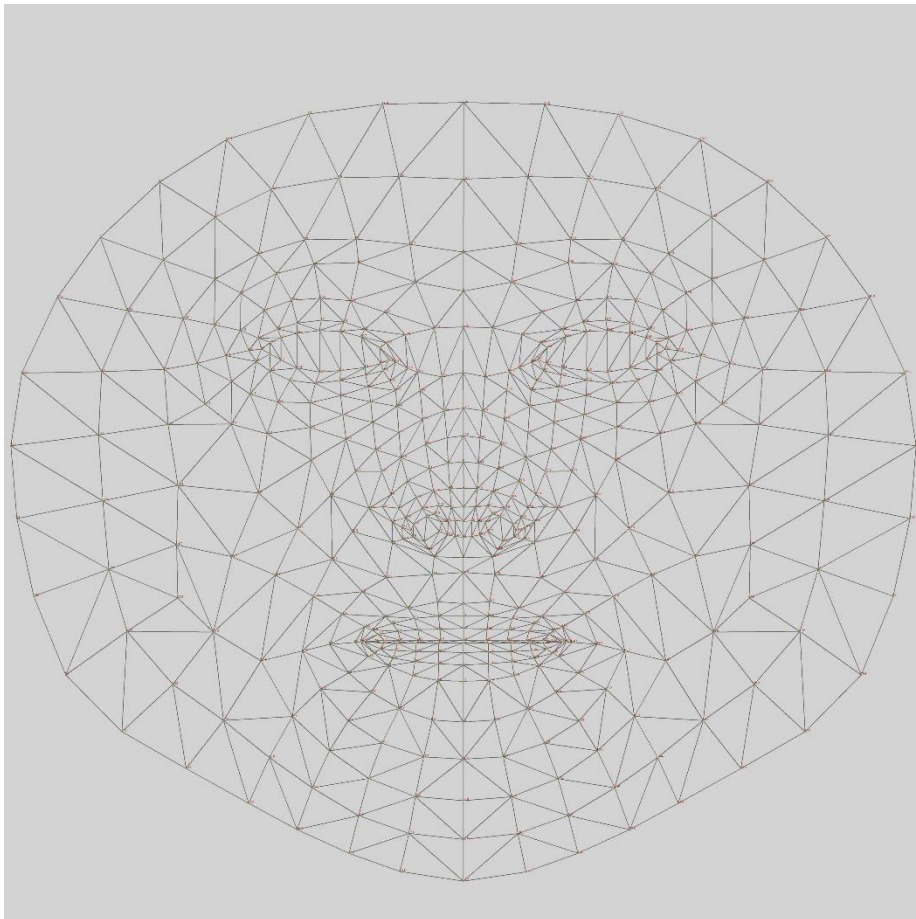


Figura 12. Malla con 478 puntos de referencia tridimensionales de MediaPipe (Pillai, 2024).

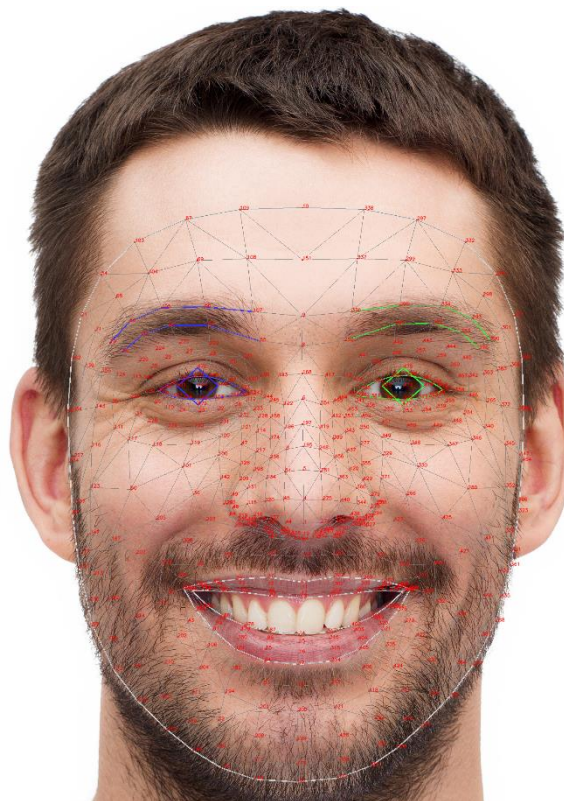


Figura 13. Malla con 478 puntos de referencia tridimensionales de MediaPipe (Guía de detección de puntos de referencia faciales | Google AI Edge, s. f.).

A continuación se muestra un ejemplo de utilización de este modelo.

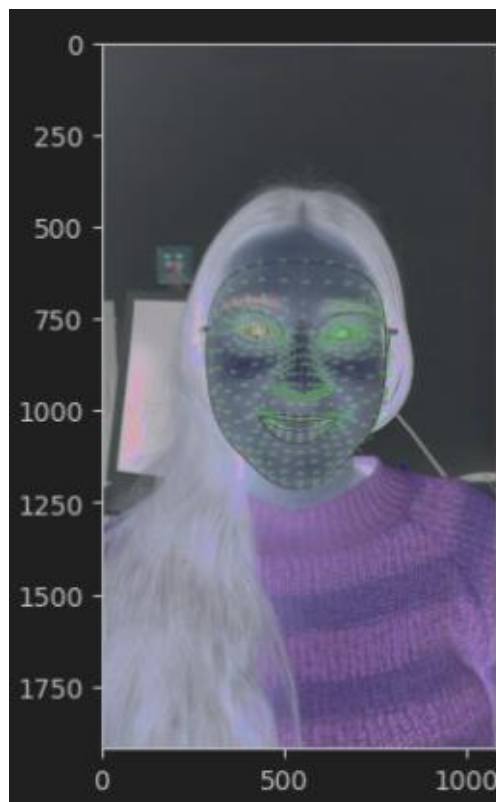


Figura 14. Detección de landmarks faciales en una imagen propia mediante MediaPipe.

En comparación con los modelos previos, este modelo tiene mayor número puntos de referencia por lo que es bastante más preciso a la hora de detectar las estructuras faciales de interés (ojos, cejas, nariz y boca), podríamos decir que es el que mejor resultado ofrece.

3.4 Relación entre signos de parálisis y movimiento de los landmarks

Teniendo en cuenta las manifestaciones clínicas de la parálisis facial mencionadas previamente podemos intuir como se moverán los *landmarks* de cada zona del rostro.

Landmarks oculares

Debido a la imposibilidad de cerrar el ojo paralizado, los *landmarks* del párpado superior e inferior no se llegan a juntar por lo que se va a producir una distancia interpalpebral aumentada en el ojo con parálisis.

Landmarks bucales

Al sonreír, la comisura del lado sano se eleva más que del lado enfermo, por lo que los *landmarks* del lado sano estarán más arriba que los del lado enfermo. Se puede cuantificar esta diferencia en el movimiento de las comisuras mediante distancias entre varios *landmarks*.

Landmarks de las cejas

Al intentar elevar las cejas, la ceja del lado sano sí que se elevará de forma normal, pero la del lado enfermo se quedará casi estática, los *landmarks* del lado sano estarán más altos que los del lado enfermo. Se puede medir este desplazamiento de los *landmarks* mediante distancias entre varios de ellos.

En resumen, los *landmarks* faciales nos permiten traducir las manifestaciones clínicas en variables cuantificables de manera objetiva.

4. MATERIALES Y MÉTODOS

Durante este capítulo se explicarán de forma detallada todas las acciones llevadas a cabo y materiales necesarios para la elaboración de este trabajo de fin de grado. Esto incluye desde la creación de una interfaz gráfica de usuario, selección de una base de datos, extracción de puntos de referencia faciales, definición de métricas para el análisis automático hasta la clasificación del grado de severidad.

4.1 Interfaz de usuario: diseño y funcionalidades

Para facilitar el uso de la herramienta creada en este proyecto, se diseñó una interfaz de usuario en Python, es la principal fuente de interacción del usuario y el sistema de análisis automático. Entre las principales funcionalidades de la interfaz se encuentran las siguientes:

- Carga de vídeos: El usuario puede seleccionar vídeos de su dispositivo y cargarlos.
- Selección de acción facial: permite elegir entre distintos movimientos o expresiones faciales, “reposo”, “arrugar frente (subir cejas)”, “cerrar ojos suavemente”, “cerrar ojos fuerte” y “enseñar dientes”, lo que facilita el análisis específico para cada movimiento.
- Selección de *frames*: A través de una barra de desplazamiento podemos movernos por los fotogramas del vídeo y seleccionar el instante más representativo de cada movimiento facial, se da a aceptar y el *frame* queda guardado para ese movimiento. Se tendrán como norma general, cinco *frames* para cada paciente.
- Evaluación de sincinesias: selector con opciones Sí/No para indicar si el paciente padece o no sincinesias.
- Almacenamiento de los *frames*: se guardan los *frames* de cada acción en una carpeta cuyo nombre es el nombre del vídeo y la fecha, esto es útil para que puedan ser posteriormente comparados con nuevas imágenes del mismo paciente y así evaluar su evolución en el tiempo. También se guardan los datos de si el paciente padece sincinesias en un archivo .txt.
- Análisis automático: al pulsar “Confirmar análisis” el sistema activa automáticamente el modelo de MediaPipe y este se encarga de extraer los *landmarks* faciales. Posteriormente a la detección de *landmarks* se realiza el cálculo de las métricas para cada uno de los *frames* seleccionados y la estimación del grado de severidad de la enfermedad.
- Visualización de resultados: tras el análisis, el sistema muestra por pantalla una serie de porcentajes que reflejan el movimiento de cada lado de la cara, así como el grado de afectación según la escala de House-Brackmann. Esta información se guarda también en el archivo .txt junto con la presencia o no de sincinesias.

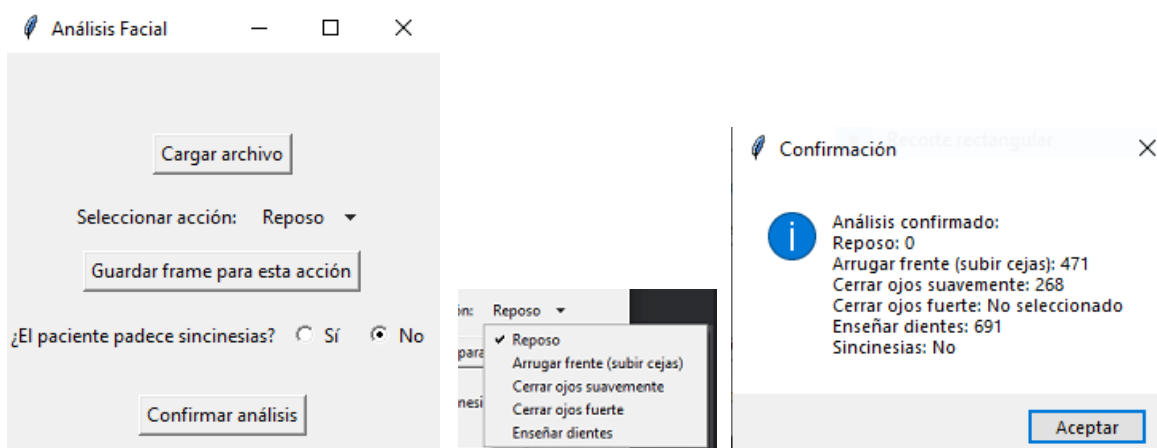


Figura 15. Interfaz de usuario, desplegable con acciones y botón de confirmación de análisis.

```

Todos los landmarks guardados en 'landmarks_todos_20250704_124058.json'
● Frente - La ceja izquierda se eleva respecto a la posición de reposo un 30.79%
● Frente - La ceja derecha se eleva respecto a la posición de reposo un 0.16%
● Comparación frente: lado izquierdo se levanta 30.63% más que lado derecho
● Ojos suavemente cerrados - El ojo derecho se cierra un 54.31% respecto a la posición de reposo
● Ojos suavemente cerrados - El ojo izquierdo se cierra un 61.23% respecto a la posición de reposo
● Comparación ojos suave: ojo izquierdo cierra un 6.92% más que ojo derecho
● Ojos fuertemente cerrados - El ojo derecho se cierra un 57.70% respecto a la posición de reposo
● Ojos fuertemente cerrados - El ojo izquierdo se cierra un 60.59% respecto a la posición de reposo
● Comparación ojos fuerte: ojo izquierdo cierra un 2.90% más que ojo derecho
● Boca - Elevación del lado derecho de la comisura respecto a posición de reposo 3.77%
● Boca - Elevación del lado izquierdo de la comisura respecto a posición de reposo 9.28%
● Comparación boca: lado izquierdo se levanta 5.51% más que lado derecho
Predicción (House-Brackmann): 2.18

```

Figura 16. Resultado del análisis que sale por pantalla.

```

resultado_20250704_124057.txt x
1 Sincinesias: No
2 Reposo: Seleccionado ●
3 Arrugar frente (subir cejas): Seleccionado
4 Cerrar ojos suavemente: Seleccionado
5 Cerrar ojos fuerte: Seleccionado
6 Enseñar dientes: Seleccionado
7 Frente - La ceja izquierda se eleva respecto a la posición de reposo un 30.79%
8 Frente - La ceja derecha se eleva respecto a la posición de reposo un 0.16%
9 Frente - Comparación: lado izquierdo se levanta 30.63% más que lado derecho
10 Ojos suavemente cerrados - El ojo derecho se cierra un 54.31% respecto a la posición de reposo
11 Ojos suavemente cerrados - El ojo izquierdo se cierra un 61.23% respecto a la posición de reposo
12 Ojos suavemente cerrados - Comparación: ojo izquierdo cierra un 6.92% más que ojo derecho
13 Ojos fuertemente cerrados - El ojo derecho se cierra un 57.70% respecto a la posición de reposo
14 Ojos fuertemente cerrados - El ojo izquierdo se cierra un 60.59% respecto a la posición de reposo
15 Ojos fuertemente cerrados - Comparación: ojo izquierdo cierra un 2.90% más que ojo derecho
16 Boca - Elevación del lado derecho de la comisura respecto a posición de reposo 3.77%
17 Boca - Elevación del lado izquierdo de la comisura respecto a posición de reposo 9.28%
18 Boca - Comparación: lado izquierdo se levanta 5.51% más que lado derecho
19 Predicción (House-Brackmann): 2.18

```

Figura 17. Contenido del archivo .txt generado.

4.2 Bases de datos utilizadas

A lo largo de este proyecto se han utilizado dos fuentes principales de datos para realizar las pruebas de los modelos de detección de puntos de referencia faciales y el modelo final.

Por un lado, se han recopilado vídeos de pacientes diagnosticados con parálisis facial del Hospital Universitario Río Hortega. Hasta el momento solo se han podido recopilar un total de 6 vídeos correspondientes a distintos pacientes con distinto grado de severidad de parálisis facial unilateral. Para poder realizar estas grabaciones fue necesario la elaboración de la documentación con la información sobre el estudio, la cual fue enviada y aprobada el día 12 de marzo de 2025 por el comité ético de Valladolid. La documentación enviada al comité constaba de una solicitud de evaluación de un TFG, un compromiso de confidencialidad, un documento de conformidad del jefe de servicio, una carta de presentación del estudio, el protocolo y el consentimiento informado para los pacientes. Este consentimiento informado debe ser firmado por todos los pacientes que quieran participar en el estudio, asegura el cumplimiento de los principios éticos y legales sobre el tratamiento de los vídeos y datos personales de los pacientes y sobre el derecho a retirarse del estudio.

Los vídeos fueron grabados mediante los dispositivos móviles del personal sanitario de la unidad de otorrinolaringología. Para la adquisición de los vídeos el dispositivo móvil debe estar lo más recto posible y frente a la cara del sujeto, posteriormente se le pide al paciente que realice un protocolo sencillo de acciones:

- Reposo: Se pide al paciente que relaje el rostro sin realizar ningún movimiento.
- Arrugar frente (subir cejas): el paciente debe levantar ambas cejas lo máximo posible.
- Cerrar ojos de forma suave: se pide que el paciente cierre los ojos con suavidad sin ejercer excesiva fuerza.
- Cerrar ojos de forma fuerte: el paciente debe cerrar los ojos apretándolos con fuerza.
- Sonreír: el paciente debe sonreír enseñando los dientes.

Estos vídeos han sido utilizados para probar el programa generado y comparar los resultados obtenidos con la valoración de un conjunto de especialistas del hospital.

Por otro lado, se solicitó acceso a una base de datos pública “YouTube-Facial-Palsy-Database” encontrada en el repositorio oficial [GitHub]. Esta base de datos está formada por secuencias de imágenes extraídas de vídeos de pacientes con parálisis facial en la plataforma YouTube. La base de datos incluye un total de 21 pacientes con distintos grados de afectación. Dependiendo del fotograma que se esté observando, la imagen puede mostrar parálisis de alta intensidad, parálisis de intensidad media o de intensidad baja (Vision, 2020/2025). Esta base de datos se utilizó especialmente para evaluar las herramientas de detección de *landmarks* faciales.

La combinación de ambos conjuntos de datos ha facilitado tanto la verificación del correcto funcionamiento del sistema en datos reales de pacientes como la comparación de diversas herramientas para la detección de puntos de referencia faciales.

4.3 Extracción de puntos de referencia faciales con MediaPipe

Para la extracción de puntos de referencia faciales se utilizó el modelo de MediaPipe desarrollado por Google, en específico el algoritmo Face Mesh que combina una red neuronal convolucional con un *pipeline* de postprocesado para extraer *landmarks* de forma rápida. La elección de este modelo se basa en las pruebas comparativas realizadas con los modelos de detección de *landmarks* mencionados previamente en este trabajo. Este modelo coloca los puntos de referencia con mayor precisión, y, además, dispone de muchos más puntos, lo que permite que se puedan realizar mediciones más detalladas y fiables en las distintas regiones del rostro.

La forma de implementar el modelo está explicado en su guía oficial de uso, en ella se puede encontrar paso a paso como debe implementarse para diferentes plataformas (ios, Android y Raspberry Pi) y lenguajes, también se encuentran ejemplos de aplicación. En nuestro caso, seguimos la guía de implementación en Python.

El modelo ofrece 478 puntos, de los cuales tenemos su posición tridimensional. Sin embargo, no todos los puntos son necesarios y útiles para nuestro objetivo. Por ello se realizó una selección de los más relevantes para nuestro estudio, serán aquellos relacionados con las comisuras labiales, párpados inferior y superior y cejas.

4.4 Métricas: definición, cálculo e interpretación

La evaluación cuantitativa de la PF mediante análisis automático se realiza a partir de la posición relativa de los puntos de referencia faciales, colocados previamente, durante las distintas acciones. Para cuantificar el desplazamiento de los *landmarks* durante los movimientos realizados se utiliza la información de la posición tridimensional de cada punto (coordenadas x, y, z) proporcionada por el modelo de MediaPipe.

La métrica utilizada para la valoración de la simetría de ambos lados del rostro es la distancia euclidiana (1):

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

Dependiendo de la acción que se analice, se seleccionan distintos pares de *landmarks*, cuya distancia se espera que varíe de una forma específica

Valoración de cierre de ojos (tanto fuerte como suave)

Para la acción de cerrar los ojos tanto de forma suave como intensa, se seleccionaron los *landmarks* 145 y 159 para el lado derecho y 374 y 386 para el lado izquierdo, como se observa en la Figura 18, según la numeración de MediaPipe. Estos puntos de referencia corresponden al borde del párpado superior y del párpado inferior en ambos ojos.

Midiendo la distancia euclidiana entre ambos párpados podemos cuantificar el grado de cierre ocular. Un ojo completamente cerrado tendrá valores cercanos a 0, mientras que el ojo que no se cierre del todo tendrá valores mayores, es decir, se espera que el lado paralizado tenga mayor distancia entre párpados (no se consigue cerrar el ojo del todo) que el lado sano, esto nos permitirá detectar limitaciones de movimiento.



Figura 18. Distancia entre landmarks seleccionados para valoración de cierre de ojos suave/fuerte.

Valoración de arrugar frente (subir cejas)

La evaluación del movimiento asociado a levantar las cejas se realiza mediante los *landmarks* 230 y 105 para el lado derecho y 450 y 334 para el lado izquierdo, según la numeración de MediaPipe. La selección de estos puntos se debe a que los puntos 230 (derecha) y 450 (izquierda) se encuentran en una zona fija del rostro, encima del hueso. Estos puntos, a pesar de la presencia de parálisis, no se van a ver significativamente afectados debido a que el hueso va a funcionar de “soporte”, evitando el descenso en situaciones de debilidad muscular. De esta forma podemos evaluar únicamente el ascenso de la ceja (*landmarks* 105 y 334) respecto a una referencia estática o punto fijo.

Este enfoque se ha encontrado en artículos que tratan de realizar valoraciones de PF objetivas. El Nottingham System, mostrado en la Figura 19 y desarrollado por Murty et al. (1994), surge como un refinamiento de los métodos objetivos de evaluación definidos por Burres y Fisch (Burres & Fisch, 1986), y emplea medidas lineales para valorar la simetría facial en distintos movimientos o expresiones, entre estas medidas se encuentra la mencionada previamente, distancia entre ceja (punto supraorbitario) y un punto infraorbitario (Quesada Marín et al., 2010).

Por otro lado, los puntos 105 (derecho) y 334 (izquierda) en la Figura 20 localizados en la parte superior de la ceja, se corresponden con la zona de máximo movimiento de las cejas.

Al medir la distancia euclidiana entre estos puntos podemos cuantificar de forma objetiva el desplazamiento de la ceja en la elevación, se espera que el lado que no tenga parálisis tenga una distancia mayor que el lado paralizado debido a que los músculos de lado enfermo están paralizados y no se consigue levantar la ceja con normalidad. De esta forma podemos detectar

asimetrías o limitaciones de movimientos en alguno de los lados.

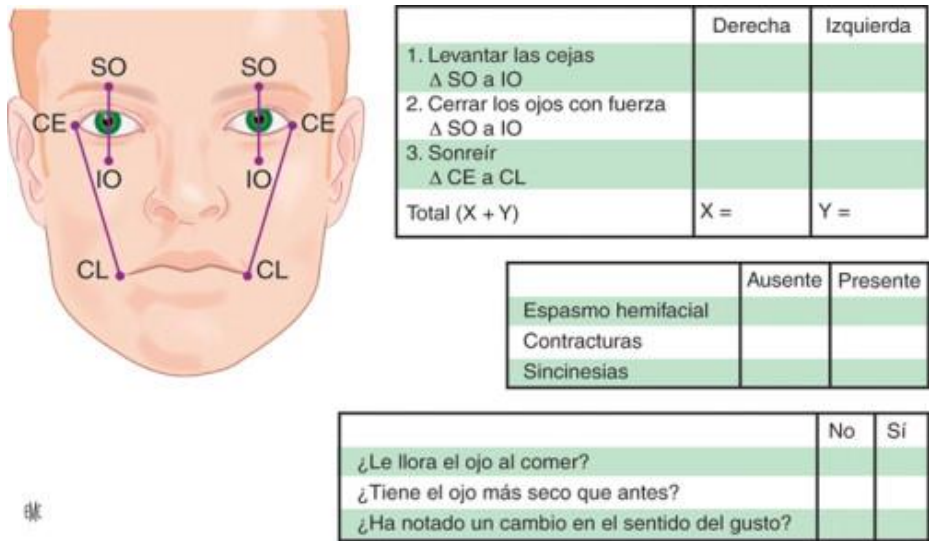


Figura 19. Distancias lineales que utiliza Nottingham System para la evaluación de parálisis facial (Quesada Marín et al., 2010).

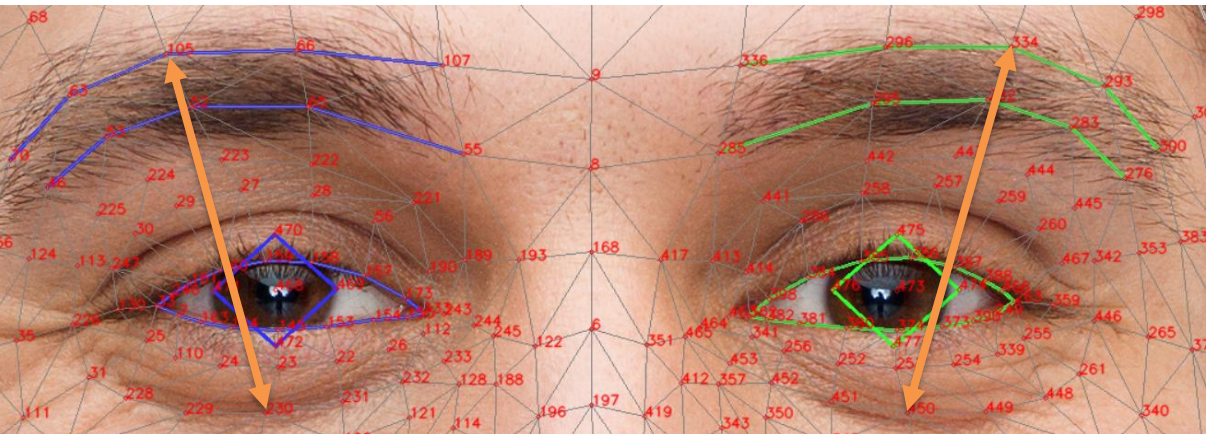


Figura 20. Distancia entre landmarks seleccionados para valoración de levantamiento de cejas.

Valoración de sonreír

La evaluación del movimiento asociado a sonreír se realiza mediante la distancia entre los *landmarks* 78 y 226 para el lado derecho, y 308 y 446 para el lado izquierdo, según la numeración de MediaPipe.

La elección de los *landmarks* está basada en el estudio mencionado anteriormente, (Quesada Marín et al., 2010), en específico siguiendo el Nottingham System. Este sistema toma como referencias las comisuras de los labios, el punto 78 (derecho) y el 308 (izquierdo) y las comisuras de los ojos. Sin embargo, en nuestro estudio, en lugar de coger las comisuras de los ojos se han elegido los puntos 226 (derecho) y 446 (izquierdo) que se localizan en zonas donde hay hueso, como se ve en la Figura 21. Esta decisión se debe a que hemos considerado que el hueso actúa como soporte, por lo que estas zonas no sufrirán desplazamientos ni variaciones significativas en el lado paralizado, y, por tanto, se pueden utilizar como una referencia fija. Por el contrario, las comisuras de los ojos sí que se verán afectadas en el lado que está paralizado, ya que se encontrarán ligeramente desplazadas hacia abajo respecto al sano.

Al medir la distancia euclidiana entre estos puntos podemos cuantificar de forma objetiva la elevación de las comisuras bucales, y por tanto evaluar el nivel de movimiento de la boca durante la sonrisa. Se espera que el lado sano tenga una distancia menor que el lado paralizado debido a que la comisura del lado sano estará más elevada, no hay limitación de movimiento, y, por tanto, más cerca del punto elegido del ojo. De esta forma podemos detectar asimetrías o limitaciones de movimientos en alguno de los lados.

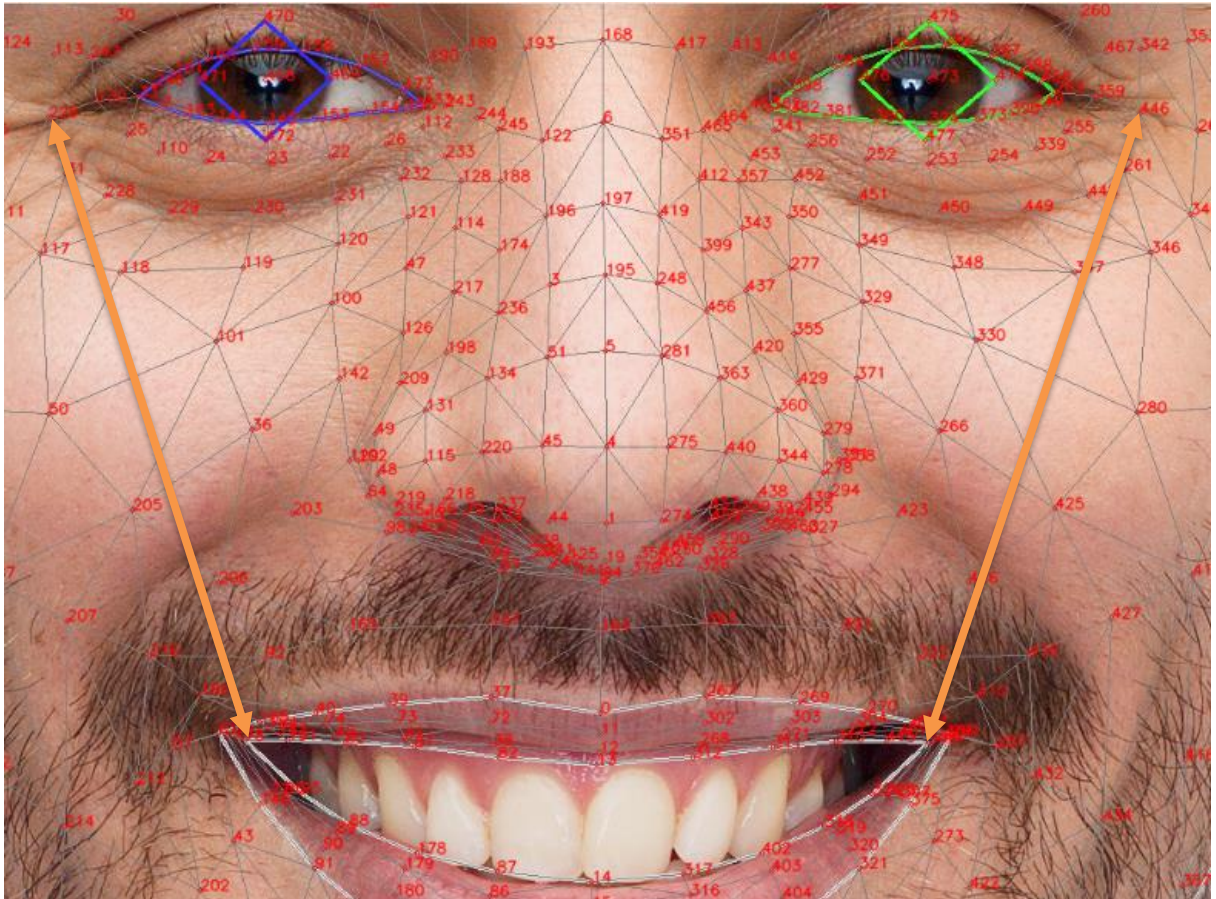


Figura 21. Distancia entre landmarks seleccionados para valoración de levantamiento de comisuras al sonreír.

Debido a que en la imagen no es posible ver cuales son los puntos que corresponden a las comisuras de la boca, buscamos en GitHub el código sobre las conexiones de los puntos de la malla facial. En la Figura 22 se detalla la conexión de puntos de referencia para todas las estructuras.

```
FACEMESH_LIPS = frozenset([(61, 146), (146, 91), (91, 181), (181, 84), (84, 17),
                             (17, 314), (314, 405), (405, 321), (321, 375),
                             (375, 291), (61, 185), (185, 40), (40, 39), (39, 37),
                             (37, 0), (0, 267),
                             (267, 269), (269, 270), (270, 409), (409, 291),
                             (78, 95), (95, 88), (88, 178), (178, 87), (87, 14),
                             (14, 317), (317, 402), (402, 318), (318, 324),
                             (324, 308), (78, 191), (191, 80), (80, 81), (81, 82),
                             (82, 13), (13, 312), (312, 311), (311, 310),
                             (310, 415), (415, 308)])
```

Figura 22. Puntos de referencia faciales que forman los labios en MediaPipe (*google-ai-edge, s.f.*).

Reposo

La posición de reposo se utiliza como punto de referencia para la evaluación del resto de acciones. Esta acción consiste en obtener la posición de los *landmarks* en posición neutral.

Para poder cuantificar el movimiento/desplazamiento producido por los músculos faciales en cada acción, es necesario disponer de todas las medidas base en reposo. Por ello, se calculan todas las distancias euclidianas mencionadas previamente. Estas medidas permiten comparar cuanto se ha desplazado cada región facial durante los movimientos voluntarios.

Cálculos comparativos

Una vez obtenidas las métricas de todas las acciones (levantar cejas, cerrar ojos fuerte y suave y sonreír), se realiza una comparación entre ellas y la posición de reposo para cuantificar el movimiento en forma de porcentaje. Posteriormente se realiza una comparación entre ambos lados del rostro.

Para la comparación con posición de reposo se utiliza la fórmula de cambio porcentual (2) respecto a esta posición:

$$\% \text{ de cambio} = \left(\frac{\text{distancia tras movimiento} - \text{distancia en reposo}}{\text{distancia en reposo}} \right) \times 100 \quad (2)$$

Cuando se trata de movimientos que impliquen una reducción de la distancia, como es el caso del cierre ocular y de sonreír, la fórmula (3) será la siguiente:

$$\% \text{ de cambio} = \left(\frac{\text{distancia en reposo} - \text{distancia tras movimiento}}{\text{distancia en reposo}} \right) \times 100 \quad (3)$$

El cálculo de estos porcentajes se realiza de forma independiente para el lado izquierdo y para el lado derecho del rostro, de esta forma se puede hacer una comparación entre ambos porcentajes (4).

$$\text{diferencia entre lados} = \% \text{mayor} - \% \text{menor} \quad (4)$$

Se calcula la diferencia entre al lado sano y el lado que tiene menos movimiento.

4.5 Criterios de clasificación del grado de severidad

Para la transformación de la diferencia porcentual entre ambos lados para todas las acciones en una estimación del grado de la escala de House-Brackmann fue necesario la colaboración de varios expertos clínicos.

El especialista que ha trabajado con nosotros a lo largo del proyecto utilizó el sistema creado para analizar cuatro, de los 6 vídeos recopilados. A partir de estos vídeos se seleccionaron cinco *frames* (en alguno de los vídeos no se realizaban todos los movimientos y por tanto tendrán menos *frames*), los cuales se guardaron en una carpeta y se enviaron a un grupo de expertos de la unidad de Otorrinolaringología. Cada experto realizó una valoración sobre el grado de afectación facial de manera independiente a partir de los cinco *frames* enviados de cada vídeo.

A continuación, se muestra el cuestionario que respondieron 15 expertos para cada vídeo enviado:

VIDEO 1

ASIMETRIA ESTATICA: ☐ LEVE ☐ MODERADA ☐ SEVERA ☐ COMPLETA

ASIMETRIA ESTATICA ESCALA SUNNYBROOK: TOTAL: ()

- CIERRE OCULAR NORMAL(0) CIERRE SUBTOTAL(1) CIERRE PARCIAL(2).. AMPLIA EXPOSICION (3)
 - SURCO NASOGENIANO NORMAL(0) PRONUNCIADO (1) POCO PRONUNCIADO(2).. AUSENTE (3)
 - ASIMETRIA LABIAL NORMAL(0) COMISURA CAIDA (1) COMISURA TRACCIONADA(2)..
 -

ASIMETRIA DINAMICA

-AL ARRUGAR LA FRENTE ☐ LEVE ☐ MODERADA ☐ SEVERA ☐ COMPLETA
 -AL CERRAR LOS OJOS SUAVE ☐ LEVE ☐ MODERADA ☐ SEVERA ☐ COMPLETA
 -AL CERRAR OJOS FUERTE ☐ LEVE ☐ MODERADA ☐ SEVERA ☐ COMPLETA
 -AL MOSTRAR LOS DIENTES ☐ LEVE ☐ MODERADA ☐ SEVERA ☐ COMPLETA

ASIMETRIA DINAMICA

ESCALA

SUNNYBROOK:

ARRUGAR LA FRENTE:
 GUIÑAR EL OJO:
 ENSEÑAR LOS DIENTES:
 SONREIR:
 FRUNCIR LOS LABIOS:

AUSENCIA COMPLETA DE MOVIMIENTO (1) ;INICIO SUAVE DE MOVIMIENTO (2) MOVIMIENTO DE EXCURSIÓN MEDIA (3) MOVIMIENTO CASI COMPLETO (4). MOVIMIENTO COMPLETO (5)

GRADACION DE HOUSE - BRACKMAN: ()

(1: FUNCION NORMAL EN TODAS LAS ÁREAS; 2.- DEBILIDAD MUSCULAR EN EXPLORACION DINAMICA; 3.- DEBILIDAD MUSCULAR NOTORIA PERO NO DESFIGURANTE EN REPOSO; 4.- ASIMETRIA EVIDENTE EN REPOSO; 5.ESCASO MOVIMIENTO PERCEPTIBLE; 6.-AUSENCIA COMPLETA DE MOVIMIENTO

Figura 23. Cuestionario realizado por profesionales médicos para la evaluación clínica tradicional de la parálisis facial de 4 vídeos.

A partir de ese cuestionario se generó un Excel con múltiples valores para cada vídeo. De este Excel se extrajeron las columnas más relevantes, mostradas en las Tabla 6, 7, 8 y 9.

VÍDEO 1					
Años ejercicio	ASIMETRIA AL ARRUGAR LA FRENTE	ASIMETRIA AL CERRAR FUERTE LOS OJOS	ASIMETRIA AL CERRAR SUAVE LOS OJOS	ASIMETRIA AL SONREIR	GRADO DE AFECTACIÓN GLOBAL
35	MODERADA	MODERADA	MODERADA	SEVERA	III
28	MODERADA	SEVERA	MODERADA	SEVERA	III
26	SEVERA	MODERADA	SEVERA	SEVERA	V
21	MODERADA	MODERADA	MODERADA	SEVERA	IV
19	MODERADA	MODERADA	SEVERA	MODERADA	IV
17	MODERADA	MODERADA	MODERADA	MODERADA	
15	SEVERA	MODERADA	SEVERA	MODERADA	V
14	MODERADA	SEVERA	SEVERA	MODERADA	IV
13	SEVERA	SEVERA	SEVERA	SEVERA	V
10	LEVE	MODERADA	LEVE	LEVE	II
4	LEVE	LEVE	MODERADA	MODERADA	IV
3	MODERADA	MODERADA	SEVERA	COMPLETA	IV
2	LEVE	LEVE	MODERADA	SEVERA	III
1	SEVERA	LEVE	MODERADA	SEVERA	IV
0	LEVE	MODERADA	MODERADA	MODERADA	II

Tabla 6. Resultados de la valoración del vídeo 1 por parte de los especialistas.

VÍDEO 2					
Años ejercicio	ASIMETRIA AL ARRUGAR LA FRENTE	ASIMETRIA AL CERRAR FUERTE LOS OJOS	ASIMETRIA AL CERRAR SUAVE LOS OJOS	ASIMETRIA AL SONREIR	GRADO DE AFECTACIÓN GLOBAL
35	MODERADA	MODERADA	MODERADA	MODERADA	V
28	MODERADA	SEVERA	SEVERA	SEVERA	IV
26	SEVERA	COMPLETA	COMPLETA	COMPLETA	VI
21	SEVERA	COMPLETA	COMPLETA	COMPLETA	V
19	SEVERA	SEVERA	MODERADA	SEVERA	IV
17	SEVERA	SEVERA	SEVERA	SEVERA	
15	COMPLETA	COMPLETA	COMPLETA	COMPLETA	VI
14	SEVERA	SEVERA	SEVERA	SEVERA	V
13	MODERADA	SEVERA	MODERADA	SEVERA	IV
10	LEVE	SEVERA	SEVERA	COMPLETA	IV
4	MODERADA	MODERADA	SEVERA	SEVERA	V
3	MODERADA	COMPLETA	COMPLETA	COMPLETA	V
2	MODERADA	SEVERA	SEVERA	SEVERA	IV
1	MODERADA	SEVERA	SEVERA	SEVERA	IV
0	MODERADA	SEVERA	SEVERA	MODERADA	II

Tabla 7. Resultados de la valoración del vídeo 2 por parte de los especialistas.

VÍDEO 3					
Años ejercicio	ASIMETRIA AL ARRUGAR LA FRENTE	ASIMETRIA AL CERRAR FUERTE LOS OJOS	ASIMETRIA AL CERRAR SUAVE LOS OJOS	ASIMETRIA AL SONREIR	GRADO DE AFECTACIÓN GLOBAL
35	MODERADA	MODERADA	LEVE	MODERADA	II
28	MODERADA	MODERADA	MODERADA	SEVERA	IV
26	SEVERA	LEVE	LEVE	MODERADA	III
21	SEVERA	MODERADA	MODERADA	SEVERA	III
19	MODERADA	MODERADA	LEVE	MODERADA	III
17	MODERADA	LEVE	MODERADA	MODERADA	
15	SEVERA	MODERADA	MODERADA	MODERADA	IV
14	SEVERA	MODERADA	MODERADA	SEVERA	IV
13	MODERADA	SEVERA	MODERADA	MODERADA	II
10	LEVE	LEVE	LEVE	MODERADA	II
4	SEVERA	LEVE	LEVE	LEVE	III
3	COMPLETA	COMPLETA	COMPLETA	COMPLETA	III
2	LEVE	LEVE	MODERADA	LEVE	III
1	COMPLETA	LEVE	MODERADA	MODERADA	III
0	MODERADA	LEVE	LEVE	LEVE	II

Tabla 8. Resultados de la valoración del vídeo 3 por parte de los especialistas.

VÍDEO 4					
Años ejercicio	ASIMETRIA AL ARRUGAR LA FRENTE	ASIMETRIA AL CERRAR FUERTE LOS OJOS	ASIMETRIA AL CERRAR SUAVE LOS OJOS	ASIMETRIA AL SONREIR	GRADO DE AFECTACIÓN GLOBAL
35	SEVERA	SEVERA	LEVE	SEVERA	V
28	MODERADA	LEVE	LEVE	MODERADA	III
26	MODERADA	MODERADA	LEVE	SEVERA	IV
21	SEVERA	LEVE	MODERADA		IV
19	SEVERA	SEVERA	SEVERA	SEVERA	V
17	MODERADA	MODERADA	MODERADA	SEVERA	
15	SEVERA	SEVERA	SEVERA	SEVERA	V
14	SEVERA	LEVE	LEVE	SEVERA	IV
13					V
10	LEVE	LEVE	LEVE	COMPLETA	III
4	SEVERA	MODERADA	MODERADA	SEVERA	V
3	COMPLETA	LEVE	SEVERA	COMPLETA	IV
2	SEVERA	MODERADA	MODERADA	SEVERA	IV
1	COMPLETA	LEVE	LEVE	SEVERA	IV
0	SEVERA	MODERADA	MODERADA	SEVERA	II

Tabla 9. Resultados de la valoración del vídeo 4 por parte de los especialistas.

Es importante destacar la gran disparidad que hay en el grado de afectación global para un mismo paciente entre los especialistas. Esta variabilidad se ha cuantificado mediante Fleiss' Kappa. Se ha elegido esta estadística porque permite medir como de acuerdo están múltiples evaluadores que clasifican ítems en categorías cualitativas, como es nuestro caso (Landis & Koch, 1977).

El cálculo se ha realizado en Python a través de la librería *statsmodels*. Como entrada para el cálculo se creó la siguiente matriz de frecuencias, cada fila es un paciente y cada columna se corresponde con el número de médicos que asignaron cada una de las seis categorías de la escala:

	Número de veces que aparece cada grado					
	I	II	III	IV	V	VI
Paciente 1	0	2	3	6	3	0
Paciente 2	0	1	0	6	5	2
Paciente 3	0	4	7	3	0	0
Paciente 4	0	1	2	6	5	0

Tabla 10. Matriz de frecuencias de valoraciones para cada paciente según la escala House-Brackmann.

Debido a que el especialista de 17 años de experiencia no ha rellenado el grado de afectación, se ha dejado fuera para los siguientes cálculos. Del mismo modo, se ha dejado fuera también al especialista de 0 años de experiencia, esta decisión se ha tomado en base a que su valoración difiere bastante de los demás, y debido a su corta experiencia se cree que puede ser un diagnóstico menos preciso que puede dar lugar a un aprendizaje erróneo de nuestro modelo.

Las evaluaciones de los 13 especialistas restantes se dividieron en 3 grupos según los años de experiencia:

- Grupo 1: menos de 5 años de experiencia (a este grupo pertenecerán los de 1, 2, 3 y 4 años de ejercicio).
- Grupo 2: entre 10 y 20 años de experiencia (los especialistas de 10, 13, 14, 15 y 19 años de ejercicio).

- Grupo 3: más de 20 años de experiencia (los que tienen 21, 26, 28 y 35 años de ejercicio).

Esta división en grupos se ha utilizado para crear 2 conjuntos de *train* y *test*. Para el conjunto de *train* se cogieron 2 especialistas de cada grupo (médicos de 28 y 26 años de experiencia del grupo 1, médicos de 15 y 14 años de experiencia del grupo 2 y médicos de 3 y 2 años de experiencia del grupo 3) y para el de *test* el resto.

Posteriormente se hizo una media ponderada (5) de la estimación del grado de afectación de los grupos de *train* y de *test*, en base a los años de experiencia de los profesionales, con el fin de obtener una clasificación consensuada para cada paciente.

$$\text{Grado de afectación promedio} = \frac{\sum \text{años ejercicio}_i * \text{grado de afectación}_i}{\sum \text{años ejercicio}_i} \quad (5)$$

A través de esta fórmula obtenemos para cada paciente un número de 1 a 6, que será el grado de afectación promedio de la escala House-Brackmann.

A continuación, se muestran los porcentajes de asimetría obtenidos mediante el sistema automático desarrollado:

VIDEO 1	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	-0.87	27.25	28.12
Cerrar ojos suave	54.32	61.23	6.91
Cerrar ojos fuerte	59.39	61.71	2.32
Sonreír	4.75	9.48	4.73

Tabla 11. Resultados de los porcentajes calculados para el vídeo 1 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

VIDEO 2	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	17.17	12.47	4.7
Cerrar ojos suave	79.07	66.06	13.01
Cerrar ojos fuerte	-	-	-
Sonreír	6.94	-6.50	13.44

Tabla 12. Resultados de los porcentajes calculados para el vídeo 2 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

VIDEO 3	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	-3.20	6.90	10.1
Cerrar ojos suave	52.75	72.70	19.95
Cerrar ojos fuerte	70.58	70.71	0.13
Sonreír	7.22	22.75	15.53

Tabla 13. Resultados de los porcentajes calculados para el vídeo 3 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

VIDEO 4	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	-	-	-
Cerrar ojos suave	51.83	72.11	20.28
Cerrar ojos fuerte	76.81	71.14	5.67
Sonreír	-1.68	3.01	4.69

Tabla 14. Resultados de los porcentajes calculados para el vídeo 4 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

Para estimar el grado de afectación se utilizó la última columna de las tablas anteriores. Debido a que de los cuatro vídeos que se han analizado hay dos de ellos que tienen datos incompletos, se estableció la relación entre los porcentajes y los grados de afectación calculados mediante un ajuste de mínimos cuadrados. El ajuste consistió en obtener el valor de A, B, C y D de la ecuación (6). Con estos valores calculados el modelo puede estimar los grados de afectación para nuevos datos.

$$\hat{y} = A * X_1 + B * X_2 + C * X_3 + D * X_4 \quad (6)$$

Siendo $\hat{y}$ las estimaciones promedio calculadas del conjunto de *train* para cada paciente mediante la ecuación 5, y las X_1, X_2, X_3, X_4 las diferencias entre los porcentajes de ambos lados para cada acción.

Por último, para evaluar la precisión del modelo se estimó el error mediante el cálculo del error cuadrático medio (MSE) (7). Para este cálculo se utilizó el conjunto de *test*, que está formado por las valoraciones $\hat{y}$ promedio calculadas a partir de las valoraciones de los especialistas no incluidas en el conjunto de *train*.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Donde y es el valor de las $\hat{y}$ promedio calculadas para el grupo de *test*, $\hat{y}$ es la predicción del modelo para los pacientes y n es el número de pacientes, en este caso 4.

Se probó este modelo con el quinto y sexto vídeo y se obtuvieron estimaciones del rango de severidad fuera de la escala House-Brackmann. Esto probablemente se debe a que el modelo no puede aprender las relaciones entre las estimaciones de los profesionales y las diferencias de los porcentajes con solo con 4 vídeos, de los cuales no todos contienen la información completa. Además, la gran variabilidad de nuestras etiquetas reales (las valoraciones realizadas por los médicos) también dificulta el aprendizaje del modelo. Por esta razón se ideó otra forma de estimar los parámetros de la ecuación 6, el modelo 2.

MODELO 2

Para reducir la subjetividad y variabilidad introducida en el modelo anterior a través de las valoraciones realizadas por los médicos, se solicitó a los profesionales que volvieran a reunirse y llegasen a un consenso único para cada paciente. De forma que solo hubiese un diagnóstico para cada vídeo. Además, en esta ocasión, en vez de separar los vídeos en *train* y *test*, se han utilizado todos para el entrenamiento con el fin de mejorar el aprendizaje del modelo.

A continuación, se muestran las nuevas valoraciones para cada vídeo:

Valoración conjunta					
Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Vídeo 6
II	V	III	IV	III	I

Tabla 15. Valoración conjunta de los especialistas para cada paciente según escala House-Brackmann.

Los porcentajes calculados a través del sistema automático para los vídeos 5 y 6 fueron los siguientes:

VIDEO 5	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	13.42	1.51	11.91
Cerrar ojos suave	48.95	11.34	37.61
Cerrar ojos fuerte	32.79	3.87	28.92
Sonreír	10.20	0.99	9.21

Tabla 16. Resultados de los porcentajes calculados para el vídeo 5 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

VIDEO 6	% de cambio derecho respecto a reposo	% de cambio izquierdo respecto a reposo	Diferencia entre ambos lados
Arrugar frente	10.61	-1.01	11.62
Cerrar ojos suave	32.63	32.75	0.12
Cerrar ojos fuerte	28.77	19.05	9.72
Sonreír	0.69	-1.71	2.4

Tabla 17. Resultados de los porcentajes calculados para el vídeo 6 a través del modelo automático desarrollado y utilizados para la comparación con la evaluación tradicional.

A partir de estos datos se vuelven a estimar los parámetros de la ecuación 6, pero ahora con los 6 vídeos y las valoraciones de la tabla 15.

Por último, se realizó el cálculo del MSE entre el valor real de cada vídeo (tabla 15) y los valores predichos.

4.6 Flujo del análisis automático y salida de resultados

El sistema desarrollado sigue una secuencia de pasos definida para estudiar la simetría del rostro y la estimación del grado de severidad de la parálisis facial. A continuación, se describe la secuencia de pasos:

1. Adquisición del vídeo: para la utilización de la herramienta desarrollada es necesario disponer de un vídeo del paciente que se quiere analizar. La adquisición se puede realizar mediante un dispositivo móvil, se debe pedir al paciente que realice los 4 movimientos (reposo, arrugar frente, cerrar ojos fuerte, cerrar ojos suave y sonreír).
2. Carga del vídeo: A través de la interfaz, el usuario selecciona el vídeo que se quiera analizar y, de manera manual, selecciona a través de una barra de progreso, un *frame* del vídeo para cada movimiento: reposo, arrugar frente, cerrar ojos suave, cerrar ojos fuerte y sonreír. Si en alguno de los vídeos no se realizan todas las expresiones, y por tanto no se selecciona ese *frame*, el modelo analizará las expresiones disponibles sin provocar errores.

3. Detección de *landmarks* faciales: El modelo Face Mesh de MediaPipe analiza los *frames* seleccionados en el paso anterior y nos devuelve un archivo .csv que contiene la información de la posición 3D de 478 puntos de referencia colocados en el rostro.
4. Cálculo de métricas: De los 478 puntos de referencia faciales, el sistema utiliza 12 puntos relevantes, correspondientes a las zonas de interés. Con estos 12 puntos realiza las medidas mencionadas para cada *frame* de reposo, arrugar frente, cerrar ojos suave, cerrar ojos fuerte y sonreír. Después, realiza una comparación de cada movimiento con la posición de reposo y nos da un porcentaje de diferencia para cada lado del rostro. Por último, el sistema realiza una comparación entre ambos porcentajes (lado derecho y lado izquierdo) y da como resultado un porcentaje diferencia.

Por tanto, para cada *frame* el sistema calcula de manera automática los siguientes datos:

Reposo

Se obtienen 3 medidas (distancia entre párpado superior e inferior, distancia entre comisura bucal y punto infraorbitario y distancia entre ceja y punto infraorbitario) para el lado izquierdo y para el lado derecho del rostro.

Arrugar frente

Tanto para el lado derecho como para el izquierdo se realiza el cálculo de la distancia entre un punto de la ceja y un punto infraorbitario. Se comparan los valores tras este movimiento y los valores en reposo y obtenemos un porcentaje para cada lado. Luego, se obtiene la diferencia entre los porcentajes del lado derecho e izquierdo.

Cerrar ojos fuerte y suave

Se calcula la distancia entre un punto del párpado superior y un punto del párpado inferior del lado derecho y del lado izquierdo. Después realiza un cálculo para comparar las distancias de ambos lados tras el movimiento y en reposo. Y por último la diferencia de porcentaje entre los porcentajes izquierdo y derecho.

Sonreír

El sistema calcula la distancia entre la comisura bucal y un punto de la zona infraorbitaria externa, para el lado derecho y para el lado izquierdo. Da unos porcentajes de comparación con la posición de reposo y por último un porcentaje de diferencia entre lado derecho e izquierdo.

5. Clasificación en grados de la escala House-Brackmann: A partir de los resultados de los porcentajes diferencia entre lado izquierdo y derecho para cada movimiento, excepto reposo, se realiza una clasificación automática según la severidad de la enfermedad. Esta clasificación se realiza en grupos del I al VI, como en la escala House-Brackmann, siguiendo criterios mencionados previamente.
6. Visualización y exportación de resultados: Una vez realizados todos los cálculos, el sistema mostrará por pantalla el valor de todos porcentajes y el resultado final correspondiente a la clasificación del I al VI. Por otro lado, se genera un archivo .txt con la información calculada y una carpeta con todos los *frames* seleccionados para cada paciente, esto es muy importante para que se puedan realizar comparaciones de evolución del paciente a lo largo del tiempo.

5. RESULTADOS Y DISCUSIÓN

A lo largo de este capítulo se mostrarán los resultados del sistema desarrollado. Se realiza un análisis para la interpretación del estadístico Fleiss' Kappa, se comentan los parámetros estimados para el cálculo del grado de afectación para el modelo inicial y para el segundo modelo planteado. Por último, se comenta la precisión de los modelos planteados.

5.1 Análisis de concordancia en las valoraciones

En primer lugar, con la cuantificación de la variabilidad de las evaluaciones de los profesionales sanitarios a través de Fleiss' Kappa obtuvimos un valor de:

Fleiss' Kappa: 0.032

Para su interpretación nos basamos en el artículo de Landis y Koch (1977) en el que se establecen una serie de rangos, como se ve en la Figura 24, para ver la aleatoriedad con la que se llega a un acuerdo.

<i>Kappa Statistic</i>	<i>Strength of Agreement</i>
<0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

Figura 24. Interpretación de los valores de Fleiss' Kappa (Landis & Koch, 1977).

El estadístico Fleiss' Kappa toma valores entre -1 y 1. Valores menores de 0 indican que el acuerdo es peor que el azar, y valores cercanos a 1 indican acuerdos muy buenos con alta concordancia entre los evaluadores. Valores cercanos a 0 indican que el acuerdo entre los jueces es prácticamente el que se esperaría por coincidencia.

Nuestro valor (0.032) es muy bajo, lo que indica un acuerdo leve y ligeramente mejor que el azar, quiere decir que la concordancia entre los especialistas médicos está muy limitada. Esta subjetividad en los diagnósticos tradicionales dificulta que nuestro modelo automático pueda aprender patrones a la hora de establecer una estimación del grado de afectación de la PF y, es, por tanto, una fuente de error en el diagnóstico automático.

Se crearon diagramas de barras para la visualización de las diferencias en la evaluación de cada paciente con el fin de identificar si hay algún paciente que genere mayor discordancia en las evaluaciones:

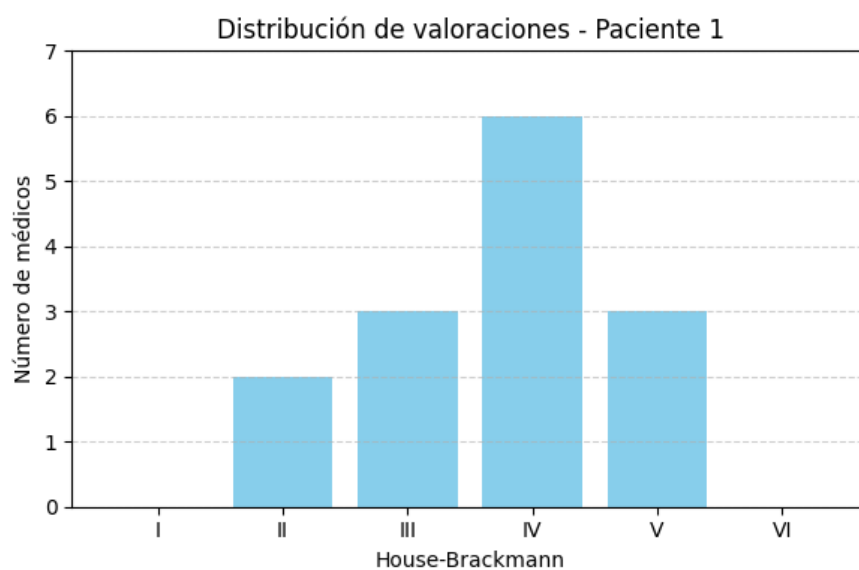


Figura 25. Diagrama de barras de variabilidad en el diagnóstico del vídeo 1.

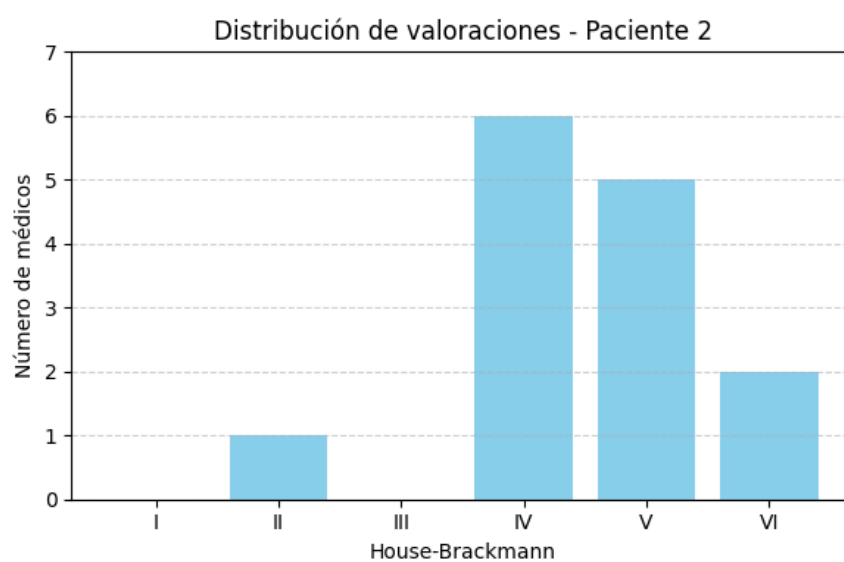


Figura 26. Diagrama de barras de variabilidad en el diagnóstico del vídeo 2.

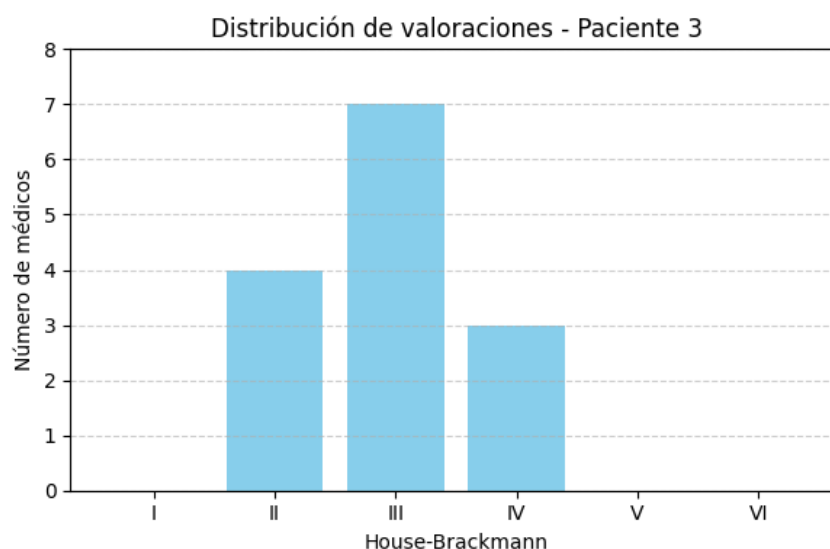


Figura 27. Diagrama de barras de variabilidad en el diagnóstico del vídeo 3.

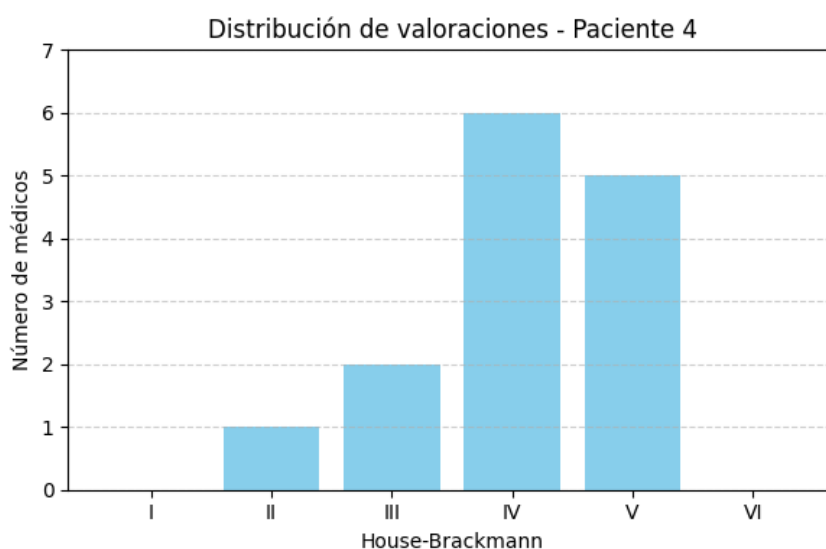


Figura 28. Diagrama de barras de variabilidad en el diagnóstico del vídeo 4.

Observamos que en todos los vídeos menos en el tercero, se han clasificado a los pacientes en cuatro grados distintos de severidad según la escala House-Brackmann, lo que demuestra una gran heterogeneidad en los diagnósticos.

Este análisis pone de manifiesto la necesidad existente de herramientas automáticas que reduzcan la subjetividad en los diagnósticos, de forma que se consigan evaluaciones más estandarizadas y consistentes.

5.2 Primer intento de estimación del grado de afectación

La relación entre los porcentajes obtenidos con el modelo automático y la valoración tradicional que se estableció través del ajuste de mínimo cuadrados (6) proporcionó los siguientes valores de los parámetros A, B, C, D:

A	B	C	D
-0.0197	-0.4893	1.7165	0.8618

Tabla 18. Resultados del primer ajuste de mínimos cuadrados.

Comparación real ($\hat{y}$ del conjunto de test) vs predicho:

Video 1: $y_{\text{real}} = 3.59223301$, $y_{\text{pred}} = 4.125$, error = -0.53276699

Video 2: $y_{\text{real}} = 4.58252427$, $y_{\text{pred}} = 5.125$, error = -0.54247573

Video 3: $y_{\text{real}} = 2.4368932$, $y_{\text{pred}} = 3.64772727$, error = -1.21083407

Video 4: $y_{\text{real}} = 4.59223301$, $y_{\text{pred}} = 3.85227273$, error = 0.73996028

El conjunto de test nos proporcionó un valor total de MSE de:

$$\text{MSE} = 0.6479$$

Sin embargo, a pesar de que el MSE no es muy elevado, al probar esta estimación de parámetros en los nuevos vídeos, se obtuvieron estimaciones del grado de afectación global muy alejadas de lo esperado y fuera del rango de la escala utilizada, House-Brackmann:

Prueba con vídeo 5:

Predicción para $x_{\text{nueva}} = [11.91, 37.61, 28.92, 9.21]$: $y = 38.9430$

Prueba con vídeo 6:

Predicción para $x_{\text{nueva}} = [11.62, 0.12, 9.72, 2.4]$: $y = 18.4654$

Del mismo modo ocurría si se probaba el modelo con otros *frames* de los vídeos con los que se había entrenado el modelo. Es decir, este modelo no era capaz de aprender relaciones consistentes entre los porcentajes y los grados de afectación por lo que se ideó otra forma de calcular los parámetros, el segundo modelo.

5.3 Estimación final del grado de afectación

El ajuste de la ecuación 6 a través del consenso de los profesionales para cada vídeo y el uso de los 6 vídeos disponibles para la estimación de los parámetros, nos dio unos resultados para los parámetros de:

A	B	C	D
0.0151	0.0834	-0.0462	0.1793

Tabla 19. Resultados del ajuste de mínimos cuadrados final.

Comparación real vs predicho:

Video 1: $y_{\text{real}} = 2.00$, $y_{\text{pred}} = 1.74$, error = -0.26

Video 2: $y_{\text{real}} = 5.00$, $y_{\text{pred}} = 3.57$, error = -1.43

Video 3: $y_{\text{real}} = 3.00$, $y_{\text{pred}} = 4.60$, error = 1.60

Video 4: $y_{\text{real}} = 4.00$, $y_{\text{pred}} = 2.27$, error = -1.73

Video 5: $y_{\text{real}} = 3.00$, $y_{\text{pred}} = 3.63$, error = 0.63

Video 6: $y_{\text{real}} = 1.00$, $y_{\text{pred}} = 0.17$, error = -0.83

Podemos destacar que los vídeos 3 y 4 son los que tienen mayor error, esto probablemente se deba a que en estos vídeos hay acciones incompletas, por los que no se han podido calcular todos los porcentajes requeridos. Estos resultados destacan la importancia de cumplir el protocolo de adquisición de vídeos para la obtención de diagnósticos fiables con el sistema automático. Debido a que el modelo automático basa su predicción del grado de afectación en las diferencias porcentuales entre lado derecho e izquierdo de varias acciones, si alguna de estas acciones no se ha realizado, disminuye la calidad de los datos y por tanto la fiabilidad de

los diagnósticos.

Con esta estimación de parámetros se obtiene un error cuadrático medio de:

MSE: 1.4587

A pesar de que el error cuadrático medio es algo superior, se comprobó que el modelo se comporta de forma más estable, proporcionando estimaciones similares para los mismos vídeos, aunque se seleccionen distintos *frames*.

Este modelo de ajuste que está basado en valoraciones consensuadas entre profesionales y un mayor número de datos, proporciona una solución más coherente y robusta, aunque probablemente pueda ser mejorado si se tuviera una mayor cantidad de datos de entrenamiento. La forma previa que se utilizó para la estimación del grado de severidad era muy sensible a la variabilidad de los evaluadores y a la falta de datos, lo que proporcionó estimaciones fuera de rango y predicciones inestables.

6. LIMITACIONES

En este capítulo se analiza y se establece cuál será la utilidad del modelo desarrollado en el entorno clínico como herramienta de diagnóstico y seguimiento de la parálisis facial. Por otro lado, se realiza un análisis de las limitaciones encontradas durante el desarrollo de este trabajo, lo que permite identificar mejoras y líneas futuras de investigación.

6.1 Utilidad clínica y aportación del modelo

El modelo desarrollado es capaz de proporcionar un diagnóstico objetivo y cuantitativo, lo que tiene gran utilidad en el ámbito clínico como herramienta de apoyo al diagnóstico. Este sistema se contrapone a las limitaciones de la evaluación tradicional, disminuyendo de manera considerable la subjetividad y la posibilidad de producir diagnósticos erróneos. Hemos podido comprobar a través del análisis Fleiss Kappa la importancia del desarrollo de herramientas objetivas, ya que la variabilidad existente entre los profesionales es muy elevada, hasta el punto de que el grado de acuerdo entre ellos es ligeramente superior al azar.

El sistema desarrollado toma las decisiones en base a una serie de distancias para la cuantificación del grado de asimetría en ambos lados del rostro, esto permite realizar una estimación automática del grado de severidad (del I al VI) según la escala de House-Brackmann. El hecho de ser una herramienta automática, rápida y fácil de usar permite que sea especialmente útil en entornos donde hay gran carga asistencial. Es una solución para agilizar los diagnósticos en aquellas consultas con muchos pacientes.

El modelo no solo permite un diagnóstico objetivo, también es una muy buena herramienta para el estudio de la evolución de los pacientes con el paso del tiempo. Esto es debido a que las imágenes elegidas por el especialista se guardan en una carpeta que puede ser consultada posteriormente, del mismo modo los resultados de los porcentajes y la estimación final también queda guardada en un archivo. La comparación de varios resultados de un paciente obtenidos en distintos momentos temporales permitiría valorar si un tratamiento está funcionando y evaluar la evolución durante las distintas fases de recuperación.

Podría ser de especial interés integrarlo en unidades como otorrinolaringología, neurología o rehabilitación. No como una herramienta sustitutiva del especialista, si no como una herramienta de ayuda para la valoración. Gracias al diseño de la interfaz gráfica, es muy fácil de utilizar por lo que no se necesitan conocimientos técnicos para su manejo.

En resumen, el sistema que hemos desarrollado ofrece un gran valor clínico al unir la facilidad de uso, la rapidez en el análisis y la objetividad en los resultados. Esto lo convierte en una base sólida para crear aplicaciones médicas que evalúen y hagan seguimiento de la parálisis facial de una manera más estandarizada y accesible.

6.2 Limitaciones técnicas y clínicas del sistema

Durante la realización de este proyecto se han ido encontrando una serie de limitaciones o de aspectos que podrían mejorarse.

La primera limitación encontrada ha sido la escasez de vídeos de pacientes con parálisis facial. Debido a que el número de pacientes que son atendidos en la unidad de otorrinolaringología del Hospital Universitario Río Hortega no es muy elevado, y a que no todos los pacientes están dispuestos a participar en el estudio, solo ha sido posible la adquisición de seis vídeos durante el período de prácticas y de realización del TFG. Esta escasez en el número de vídeos disponibles ha limitado bastante el número de pruebas que se han podido realizar con el

modelo desarrollado, por lo que no se puede asegurar que el modelo sea preciso para todos los grados de parálisis facial. Además, el disponer de una base de datos tan pequeña ha imposibilitado la capacidad de dividir la base de datos en validación y test. La idea principal era la utilización de una parte de los vídeos para el desarrollo y la mejora del sistema de evaluación automático, y la parte restante de los vídeos para evaluar la precisión real y objetiva de diagnóstico del sistema. Finalmente, para mejorar el aprendizaje, se utilizaron los 6 vídeos para el entrenamiento y desarrollo del sistema automático.

Otra limitación encontrada ha sido el escaso número de modelos de detección de puntos de referencia faciales encontrados. Tras una exhaustiva búsqueda en distintas plataformas y repositorios solo se encontraron 3 modelos que realizaran esta detección. Además, alguno de ellos mostraba una baja precisión en la identificación de estructuras faciales, lo que redujo significativamente las opciones para el desarrollo del sistema. Todos los modelos encontrados han sido entrenados en personas sanas, por lo que al usar el modelo en pacientes con PF, donde la simetría y los patrones anatómicos están alterados, la capacidad de detección de *landmarks* puede verse comprometida. Para solventar este problema habría que reentrenar los modelos con imágenes de pacientes de PF con todos los grados de afectación. Sin embargo, no fue posible debido a que no se disponía ni de base de datos ni de tiempo para ello.

Por último, destacamos que el proceso de elección del *frame* correspondiente a cada acción se realiza de forma manual a través de la interfaz gráfica. Esto introduce cierto grado de subjetividad ya que es posible que distintos especialistas seleccionen diferentes momentos del vídeo para una misma acción. Puesto que el objetivo del desarrollo de este sistema es crear una herramienta de diagnóstico objetiva y lo más automatizada posible, de cara al futuro podría realizarse la elección del *frame* de manera automática lo que reduciría la variabilidad interobservador.

7. CONCLUSIONES Y LÍNEAS FUTURAS

En este último capítulo se realizará un análisis sobre las conclusiones extraídas del desarrollo del proyecto, se valorará si se han cumplido los objetivos planteados al inicio del trabajo y la eficacia del sistema. Por otro lado, se proponen una serie de mejoras que podrían ayudar a optimizar el modelo, de manera que permita la continuación del desarrollo de herramientas diagnósticas automáticas.

7.1 Cumplimiento de los objetivos

Para la elaboración de este proyecto se ha desarrollado un sistema que permite la colocación de puntos de referencia faciales, la medida de ciertas distancias de interés, la comparación de estas medidas y la estimación de un grado de severidad de la PF. De este modo se ha conseguido crear una herramienta que permite la valoración clínica de la parálisis facial de forma automática y objetiva que servirá como ayuda a los profesionales a la hora de realizar los diagnósticos.

En el capítulo de introducción se enumeraron un conjunto de objetivos específicos que se esperaban cumplir a lo largo del trabajo, por consiguiente, se va a valorar si se ha producido el cumplimiento de los mismos:

- I. Se han revisado numerosos artículos, tesis y libros acerca de la parálisis facial, el análisis automático para el diagnóstico de enfermedades y sobre modelos de detección de puntos faciales. A través de esta información obtenida se ha estudiado el estado actual y las limitaciones de las herramientas disponibles para el diagnóstico automático de la PF, lo cual ha justificado la necesidad del sistema desarrollado.
- II. Se ha desarrollado una interfaz gráfica en Python que permite la carga de vídeos, la selección de un *frame* para cada acción pedida, el análisis de todos los *frames* elegidos, la especificación de la presencia o no de sincinesias y el guardado tanto de los resultados obtenidos como de los *frames* seleccionados. La interfaz facilita el uso del sistema, de forma que no se necesiten conocimientos técnicos para su correcta utilización.
- III. Se ha conseguido implementar de forma correcta el modelo de MediaPipe y se ha desarrollado una herramienta en Python que realiza una serie de cálculos de distancias a partir de los *landmarks* faciales colocados por MediaPipe. A partir de estos cálculos la herramienta calcula una serie de porcentajes que posteriormente utiliza para conseguir una estimación del grado de severidad según la escala de House-Brackmann.
- IV. Aunque el número de pacientes ha sido reducido, se ha podido comparar los porcentajes calculados mediante el sistema desarrollado con la evaluación de un conjunto de expertos para poder establecer una correlación entre los porcentajes calculados y los grados de severidad.
- V. Tras los hallazgos realizados en este trabajo, se exponen una serie de reflexiones, sobre las limitaciones del proyecto y las líneas futuras para la mejora del sistema. Estos aspectos abren un camino a futuras investigaciones.

7.2 Conclusiones

Una vez discutidos los resultados y analizado las limitaciones, se han extraído una serie de conclusiones que se enumeran a continuación:

1. El sistema desarrollado permite la realización de un diagnóstico automático de manera sencilla y rápida. Por lo que es posible obtener una estimación objetiva del grado de severidad de la parálisis facial unilateral según la escala de House-Brackmann a partir de las distancias sobre puntos de referencia faciales.
2. El modelo preentrenado de MediaPipe utilizado en este proyecto ha demostrado tener gran precisión a la hora de detectar y colocar los puntos de referencia en el rostro. Sin embargo, la precisión del modelo se podría mejorar si este fuese reentrenado con sujetos que padezcan PF.
3. La interfaz gráfica desarrollada permite que el uso de la herramienta sea mucho más sencilla e intuitiva para el personal sanitario. Este punto es muy importante para que se pueda integrar el sistema en la práctica clínica real.
4. A pesar del limitado número de vídeos, se ha demostrado que el sistema puede, en cierta medida, imitar el criterio clínico de los expertos, mostrando un comportamiento relativamente coherente al estimar el grado de afectación. Sin embargo, la precisión de las estimaciones se puede considerar limitada. Se espera que a medida que se aumente la base de datos y esta sea más completa y representativa de todos los grados de parálisis se produzca una mejora en la precisión del modelo.
5. El sistema que hemos desarrollado no solo es útil para el diagnóstico inicial, sino que también funciona como una herramienta de seguimiento a lo largo del tiempo, para valorar la evolución de los pacientes. Esto es posible gracias al almacenamiento automático de resultados y fotogramas seleccionados de cada paciente.
6. En el futuro, esta herramienta podría servir de gran ayuda para la evaluación clínica tradicional, facilitando diagnósticos más objetivos y reproducibles. Además, gracias a que se trata de una herramienta muy rápida, su uso podría resultar especialmente útil en áreas con gran cantidad de pacientes, como otorrinolaringología o neurología.

7.3 Líneas futuras

Teniendo en cuenta las limitaciones identificadas en el apartado anterior, una de las líneas más prometedoras para el futuro sería el aumento del tamaño de la base de datos. Ampliar el número de vídeos disponibles podría permitirnos dividir la base de datos en un conjunto de validación y un conjunto de test. Esto no solo permitiría el desarrollo de un sistema con mayor capacidad de generalización, sino que también nos permitiría evaluar la precisión real del modelo en pacientes que no han sido utilizados para la creación del sistema, lo que evitaría obtener resultados sesgados o demasiado optimistas.

En el caso de disponer de una base de datos muy extensa, también se podría valorar la opción de realizar un reentrenamiento del modelo de MediaPipe utilizando imágenes de sujetos enfermos de parálisis facial. Esto aumentaría de forma significativa la precisión del modelo en la detección de estructuras faciales: cejas, ojos, boca, nariz...

En segundo lugar, otra línea de mejora sería evitar la introducción de cierto grado de subjetividad mediante la elección manual de los *frames* del vídeo. Para ello se podría estudiar la opción de que en vez de realizar la selección de un solo *frame*, que el sujeto elija una ventana de múltiples *frames*, esto reduciría las diferencias en la elección del momento del vídeo para el

análisis. También se podría hacer un análisis totalmente automático en el que el propio programa es el encargado de la elección del *frame* para cada acción. El programa deberá elegir el momento más representativo para cada acción, que será aquel en el que la distancia entre la ceja y el punto infraorbitario sea máxima para el movimiento de levantar las cejas, el punto del vídeo en el que la distancia entre el párpado inferior y el párpado superior sea la mínima para la acción de cerrar ojos (suave y fuerte), para la acción de sonreír deberá elegir el momento del vídeo en el que la distancia entre la comisura bucal y el punto infraorbitario externo sea la mínima.

Otro aspecto que podría verse mejorado de cara al futuro sería la oferta de modelos públicos o de código abierto que permitan la colocación de *landmarks* en el rostro. Aunque en el momento de realización del estudio no se han encontrado muchas opciones precisas y funcionales, es posible que en un futuro se desarrollen nuevos modelos con mejores prestaciones. Esto permitirá probar un mayor número de modelos y permitirá sustituir el modelo actual en el caso de que se encuentre alguno que ofrezca mejores resultados.

Por último, se podría mejorar la forma de presentación y visualización de los resultados. Resultaría interesante que se mostrasen de una manera más visual e intuitiva, como diagramas de barras que muestren la probabilidad de pertenencia a cada grado de severidad de la escala House-Brackmann. Y del mismo modo que los porcentajes de movimiento de ambos lados del rostro se muestren como un diagrama de barras en el que puedas ver de forma rápida cuanta diferencia de movimiento hay entre ambos lados. Esto sería importante ya que el personal sanitario busca herramientas que sean sencillas de usar y de interpretar.

Todas estas implementaciones podrían mejorar la calidad de los resultados y la precisión a la hora de diagnosticar la PF a partir de esta herramienta desarrollada.

8. BIBLIOGRAFÍA

Andrews, J., Alwafi, A., Bichu, Y. M., Pliska, B. T., Mostafa, N., & Zou, B. (2023). Validation of three-dimensional facial imaging captured with smartphone-based photogrammetry application in comparison to stereophotogrammetry system. *Heliyon*, 9(5), e15834. <https://doi.org/10.1016/j.heliyon.2023.e15834>

Benítez S., S., Danilla E., S., Troncoso O., E., F. Kinesióloga, A. M., & A. Kinesióloga, J. M. (2016). MANEJO INTEGRAL DE LA PARÁLISIS FACIAL. *Revista Médica Clínica Las Condes*, 27(1), 22-28. <https://doi.org/10.1016/j.rmcl.2016.01.004>

Burres, S., & Fisch, U. (1986). The comparison of facial grading systems. *Archives of Otolaryngology--Head & Neck Surgery*, 112(7), 755-758. <https://doi.org/10.1001/archotol.1986.03780070067015>

Características de los grados de afectación del nervio facial y la discapacidad facial en pacientes con parálisis facial periférica de un año de evolución. (s. f.). Recuperado 4 de junio de 2025, de <https://core.ac.uk/download/pdf/323352811.pdf>

Descripción general: Documentación de MMPose 1.3.2. (s. f.). Recuperado 13 de junio de 2025, de <https://mmpose.readthedocs.io/en/latest/overview.html>

Garrido Calvo, A. M., González Espallargas, E., Pinós Laborda, P. J., & Gil Romea, I. (2000). Una parálisis facial periférica. *Medicina Integral*, 36(8), 285-293.

GitHub—1adrianb/face-alignment: :Fire: Biblioteca de alineación de rostros 2D y 3D creada con PyTorch. (s. f.). Recuperado 13 de junio de 2025, de <https://github.com/1adrianb/face-alignment/tree/master>

Gómez Cuarán, F. N., & Higuera Castro, G. A. (2019). State of the art at mobile applications for facial paralysis: Information, diagnostic and rehabilitation. *Visión Electrónica*, 2(Extra 1), 169-182.

google-ai-edge. (s. f.). *Mediapipe/mediapipe/python/solutions/face_mesh_connections.py at master · google-ai-edge/mediapipe.* GitHub. Recuperado 17 de junio de 2025, de https://github.com/google-ai-edge/mediapipe/blob/master/mediapipe/python/solutions/face_mesh_connections.py

Guía de detección de puntos de referencia faciales | Google AI Edge. (s. f.). Google AI for Developers. Recuperado 14 de junio de 2025, de https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker?hl=es-419

Guía de detección de puntos de referencia faciales | Google AI Edge | Google AI for Developers. (s. f.). Recuperado 13 de junio de 2025, de https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker?hl=es-419

Hasebe, K., Kojima, T., Okanou, Y., Yuki, R., Yamamoto, H., Otsuki, S., Fujimura, S., & Hori, R. (2024). Novel evaluation method for facial nerve palsy using 3D facial recognition system in iPhone. *Auris Nasus Larynx*, 51(3), 460-464. <https://doi.org/10.1016/j.anl.2024.02.003>

Katsumi, S., Esaki, S., Hattori, K., Yamano, K., Umezaki, T., & Murakami, S. (2015). Quantitative analysis of facial palsy using a three-dimensional facial motion measurement system. *Auris Nasus Larynx*, 42(4), 275-283. <https://doi.org/10.1016/j.anl.2015.01.002>

Lagarde, J., & Améri, A. (2012). Parálisis facial. *EMC - Tratado de Medicina*, 16(3), 1-7. [https://doi.org/10.1016/S1636-5410\(12\)62726-3](https://doi.org/10.1016/S1636-5410(12)62726-3)

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>

Lassaletta, L., Morales-Puebla, J. M., Altuna, X., Arbizu, Á., Arístegui, M., Batuecas, Á., Cenjor, C., Espinosa-Sánchez, J. M., García-Iza, L., García-Raya, P., González-Otero, T., Mañós, M., Martín, C., Moraleda, S., Roda, J. M., Santiago, S., Benítez, J., Cavallé, L., Correia, V., ... Gavilán, J. (2020). Parálisis facial: Guía de práctica clínica de la Sociedad Española de ORL. *Acta Otorrinolaringológica Española*, 71(2), 99-118. <https://doi.org/10.1016/j.otorri.2018.12.004>

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). *MediaPipe: A Framework for Building Perception Pipelines* (No. arXiv:1906.08172). arXiv. <https://doi.org/10.48550/arXiv.1906.08172>

Mmpose/docs/es/dataset_zoo/2d_face_keypoint.md en main · open-mmlab/mmpose · GitHub. (s. f.). Recuperado 13 de junio de 2025, de https://github.com/open-mmlab/mmpose/blob/main/docs/en/dataset_zoo/2d_face_keypoint.md

Murty, G. E., Diver, J. P., Kelly, P. J., O'Donoghue, G. M., & Bradley, P. J. (1994). The Nottingham System: Objective assessment of facial nerve function in the clinic. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, 110(2), 156-161. <https://doi.org/10.1177/019459989411000203>

National Library of Medicine. (s. f.). *Parálisis de Bell: MedlinePlus enciclopedia médica.* Parálisis de Bell. Recuperado 4 de junio de 2025, de <https://medlineplus.gov/spanish/ency/article/000773.htm>

Pillai, K. (2024, enero 20). *Answer to «Mediapipe facemesh vertices mapping»* [Post]. Stack Overflow. <https://stackoverflow.com/a/77852970>

Quesada Marín, P., López Aguado, D., & Quesada Martínez, J. L. (2010). *Parálisis facial periférica complicaciones y secuelas.* E.U.R.O.M.E.D.I.C.E.

Rosebrock, A. (2017). *Facial landmarks with dlib, OpenCV, and Python—PyImageSearch.* <https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>

Rubin, M. (2023). *Parálisis del nervio facial—Trastornos neurológicos—Manual Merck versión para profesionales.* <https://www.merckmanuals.com/es-us/professional/trastornos-neurol%C3%B3gicos/trastornos-neurooftalmol%C3%B3gicos-y-de-los-pares-cranianos/par%C3%A1lisis-del-nervio-facial>

Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. (2022, marzo 24). *La parálisis facial: Tipos, diagnóstico y tratamiento | SEORL-CCC.* <https://seorl.net/la-paralisis-facial-tipos-diagnostico-y-tratamiento/>

Su, Z., Liang, B., Shi, F., Gelfond, J., Šegalo, S., Wang, J., Jia, P., & Hao, X. (2021). Deep learning-based facial image analysis in medical research: A systematic review protocol. *BMJ Open*, 11(11), e047549. <https://doi.org/10.1136/bmjopen-2020-047549>

Taeger, J., Bischoff, S., Hagen, R., & Rak, K. (2021). Utilization of Smartphone Depth Mapping Cameras for App-Based Grading of Facial Movement Disorders: Development and Feasibility Study. *JMIR mHealth and uHealth*, 9(1), e19346. <https://doi.org/10.2196/19346>

Ten Harkel, T. C., de Jong, G., Marres, H. A. M., Ingels, K. J. A. O., Speksnijder, C. M., & Maal, T. J. J. (2023). Automatic grading of patients with a unilateral facial paralysis based on the Sunnybrook Facial Grading System—A deep learning study based on a convolutional neural

network. *American Journal of Otolaryngology*, 44(3), 103810.
<https://doi.org/10.1016/j.amjoto.2023.103810>

Vision, A.-A. (2025). *AvLab-CV/YouTube-Facial-Palsy-Database* [Software].
<https://github.com/AvLab-CV/YouTube-Facial-Palsy-Database> (Obra original publicada en 2020)

Vrochidou, E., Papić, V., Kalampokas, T., & Papakostas, G. A. (2023). Automatic Facial Palsy Detection—From Mathematical Modeling to Deep Learning. *Axioms*, 12(12), 1091.
<https://doi.org/10.3390/axioms12121091>