

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



**“Validación de Algoritmo para la Cuantificación del Grado
de Parálisis Facial”**

ROXANA ALEJANDRA SAGÜÉS LAGOS

TESIS PARA OPTAR AL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA.

**Director de Tesis: Prof. Dr. Mauricio Cerda
Co-Director: Prof. Daniel Rappoport**

2022

AGRADECIMIENTOS

Gracias a todos los que han estado presente de alguna u otra forma en este mágico proceso.

A Mauricio Cerda su permanente apoyo, dedicación, tiempo y paciencia.

A los Dres. Daniel Rappoport y Camilo Bustos y el equipo médico por el tiempo y dedicación al etiquetado de videos.

A Sandra de la Fuente por sus consejos y apoyo.

A Rodrigo Gálvez por todas sus aclaraciones y toda su ayuda para echar a andar el inicialmente misterioso software.

A mis queridos amigos Vale, Vivi, Su, Ricardo y Nancy por ayudarme a no claudicar.

INDICE

RESUMEN	5
ABSTRACT	7
1. INTRODUCCIÓN	9
1.1. ANTECEDENTES.....	9
1.2. EVALUACIÓN CLÍNICA DE LA PARÁLISIS FACIAL	11
1.3. EVALUACIÓN INFORMÁTICA DE LA PARÁLISIS FACIAL (NO CLÍNICA)	12
1.4. MOTIVACIÓN	14
2. HIPÓTESIS Y OBJETIVOS	15
2.1. HIPÓTESIS.....	15
2.2. OBJETIVO GENERAL	15
2.3. OBJETIVOS ESPECÍFICOS.....	15
3. MATERIALES Y MÉTODOS	16
3.1. CONSTRUIR UNA BASE DE DATOS DE PERSONAS PARA CUANTIFICAR LA PARÁLISIS FACIAL A PARTIR DE VIDEOS DISPONIBLES EN INTERNET.	16
3.2. ENTRENAR ALGORITMOS AUTOMÁTICOS DE CLASIFICACIÓN (ML) CON EJEMPLOS DE MEDICIONES DE PARÁLISIS FACIAL.	19
3.3. DEFINIR PROTOCOLOS PARA LA EVALUACIÓN DEL GRADO DE PARÁLISIS FACIAL PARA SER APLICADOS EN UNA ENTIDAD HOSPITALARIA Y EN EL HOGAR DEL PACIENTE.....	27
4. RESULTADOS OBTENIDOS.....	28
4.1. ANÁLISIS DE LA BASE DE DATOS CONSTRUIDA	28
4.2. RESULTADOS DE LOS ENTRENAMIENTOS CON ALGORITMOS DE CLASIFICACIÓN (ML)	32
4.3. PROTOCOLOS PARA LA EVALUACIÓN DEL GRADO DE PARÁLISIS FACIAL PARA SER APLICADOS EN UNA ENTIDAD HOSPITALARIA Y EN EL HOGAR DEL PACIENTE.	39
5. DISCUSIÓN.....	45
6. CONCLUSIONES	49
7. REFERENCIAS.....	51
ANEXO A: VIDEOS DE INTERNET.....	53
ANEXO B: GRÁFICOS DE DISTRIBUCIÓN DE DATOS DE ENTRENAMIENTO .	53
ANEXO C: RESULTADOS DE ENTRENAMIENTOS CON ALGORITMOS DE CLASIFICACIÓN (ML) POR MÉDICO.....	60
ANEXO D: FORMULARIO DE CONSENTIMIENTO.....	62

ANEXO E: DATOS ESTADÍSTICOS COMPLEMENTARIOS	65
--	-----------

Resumen

La parálisis facial consiste en la pérdida parcial o total del movimiento del rostro, involucrando a uno o ambos lados de la cara. Dentro de las posibles causas de parálisis facial están: infecciones, tumores, traumas, toxinas, causas congénitas, neurológicas o idiopáticas. En el caso de la parálisis unilateral una escala ampliamente usada para diagnosticar el grado de parálisis es la de House-Brackmann. La escala House-Brackmann (HB) establece y describe seis categorías respecto de las cuales el médico evalúa el grado de disfunción del paciente. Este proceso corresponde a un método clínico de evaluación. Por otra parte, con el advenimiento de la informática y computación han surgido nuevos métodos denominados no-clínicos que buscan entregar mediciones más objetivas de parálisis facial. En un trabajo anterior se desarrolló un método no-clínico para la evaluación del grado de parálisis facial basado en la detección automática de expresiones faciales asimétricas (tracking facial), a partir de las cuales se asigna un grado de parálisis según la escala HB. Las expresiones faciales asimétricas se calculan mediante niveles de simetría (indicadores de simetría) que reflejan las diferencias en la movilidad del lado izquierdo y derecho del rostro para un conjunto de expresiones.

En este trabajo de tesis se buscó validar la cuantificación del grado de parálisis facial unilateral mediante algoritmos automáticos de clasificación (*Machine Learning*) que hacen uso de indicadores de simetría del rostro desarrollados anteriormente. De esta manera se busca comprobar la factibilidad de aplicar métodos no-clínicos automatizados al problema de evaluación de la parálisis facial unilateral.

Los resultados obtenidos no son concluyentes respecto de la superioridad de las evaluaciones clínicas por sobre las predicciones con algoritmos automatizados (ML). Así, por el mismo motivo, no es posible rechazar el uso de una herramienta basada en tracking facial y algoritmos de clasificación ML. Para afinar los resultados, se requiere contar con una base de datos mayor, y adecuar los algoritmos de tracking facial para la zona ocular.

Los niveles de exactitud o concordancia, al disminuir el número de categorías HB aumentan, esto es: agrupando categorías adyacentes de a dos y de a tres, simplificando el problema. Así, en un análisis más grueso (menor número de categorías) se obtiene una clasificación con mayor exactitud.

Respecto del protocolo desarrollado para realizar la grabación en el hogar del paciente, su uso podría significar un aporte -más allá del proyecto- como alternativa a la evaluación presencial; ya que ofrece beneficios a pacientes; por ejemplo, frente a situaciones en que no desean salir y ser vistos, y también a médicos, al tener la posibilidad de evaluar en otro momento o por videoconferencia. También es una respuesta frente a restricciones de movilidad o riesgo en la salud.

Finalmente, con relación a la base de videos, aunque se trata de casos obtenidos en internet, se estima de utilidad para cualquier desarrollo futuro, que requiera evaluar o probar herramientas de tracking sobre gestos asimétricos ya que a la fecha no existen bases públicas de videos con estas características.

Abstract

Facial paralysis is the partial or total loss of movement of the face, involving one or both sides of the face. Among the possible causes of facial paralysis are infections, tumors, traumas, toxins, congenital, neurological, or idiopathic causes. In the case of unilateral paralysis, a widely used scale to diagnose the degree of paralysis is the House-Brackmann scale. The House-Brackmann (HB) scale defines and describes six categories against which the doctor evaluates the degree of dysfunction of the patient. This process corresponds to a clinical evaluation method. On the other hand, with the advent of informatics and computing, new methods called non-clinical have emerged seeking to provide more objective measurements. In a previous work a non-clinical method, based on the automatic detection of asymmetric facial expressions (facial tracking), was developed for the evaluation of the degree of facial paralysis according to the HB scale. Asymmetrical facial expressions are calculated by means of symmetry levels (symmetry indicators) that reflect the differences in the mobility of the left and right sides of the face

This thesis work sought to validate the quantification of the degree of unilateral facial paralysis carried out by automatic classification algorithms (*Machine Learning*) that make use of facial symmetry indicators previously developed, to check the feasibility of applying automated non-clinical methods.

The results indicate that the degree of accuracy achieved by the evaluations carried out with automated algorithms (ML) is similar to that of health professionals. Thus, for the same reason, it is not possible to reject the use of a tool based on facial tracking and ML classification algorithms. To fine-tune the results, it is necessary to have a larger database, and to adapt the facial tracking algorithms for the ocular zone.

The levels of accuracy or agreement are low in both cases, however, by reducing the number of HB categories, that is: grouping adjacent categories of two and three, and

thus, simplifying the problem, the accuracy increases. In a coarser analysis (less number of categories) a more accurate classification is obtained.

Regarding the protocol developed to make the recording in the patient's home, it could be included as an alternative to face-to-face evaluation, since it offers benefits and ease of use both to patients (cases in which the patient does not want to be seen in public) and to medical doctors (deferred evaluation or by video conference); it could also it could be used when facing mobility restrictions or health risk issues.

Finally, regarding the video database, although these videos were rescued from the Internet, it is viewed as will be useful in any future development, that require to test tracking software or tools for asymmetric gestures. To the date there are no public video databases or repositories with these characteristics.

.

1. Introducción

1.1. Antecedentes

La parálisis facial consiste en la pérdida del movimiento del rostro debido a una afectación del nervio facial que puede involucrar uno o ambos lados de la cara. Los gestos, en el caso de parálisis unilateral, se observan sólo en el lado no afectado.

La parálisis facial puede ser de origen central o periférico. Las lesiones periféricas del nervio facial producen parálisis de los músculos faciales de un lado de la cara mayoritariamente; sólo en el 1% de los casos se produce una parálisis bilateral. En la parálisis facial de origen central se ven afectados los músculos de la parte inferior de la cara, se conserva la musculatura frontal y el paciente no suele tener dificultades para levantar la ceja o cerrar el ojo.

En la categoría correspondiente a parálisis de origen periférico, entre el 30 y 40% de los casos corresponden a parálisis de Bell o idiopática (causa desconocida), con una incidencia anual que oscila entre 11 y 40 casos por cada 100.000 habitantes [6]. Un 25% es consecuencia de algún traumatismo, fractura, compresión nerviosa o cirugía, denominándose parálisis facial traumática. Otro 13 % es de origen tumoral [3] entre los que encuentran los tumores parotídeos y que dependiendo del diagnóstico pueden requerir cirugía. El resto de los casos se deben a virus, infecciones auditivas o enfermedades en el organismo que afectan al nervio facial [3].

La Tabla 1 resume los distintos tipos de parálisis por causa, indicando sus porcentajes. Sin embargo, cabe mencionar que la literatura presenta variabilidad en los porcentajes observados [4].

Clasificación de la Parálisis Facial por Causa (Etiología)		Distribución %
Causas Idiopáticas	Parálisis de Bell, sarcoidosis, parálisis de Bell hereditaria, myasthenia gravis, esclerosis múltiple, arteritis temporal	40%
Infección	Otitis externa, varicella Zoster, sífilis, otitis media, poliomyelitis, botulismo, mastoiditis, coxsackievirus, tétano, difteria, VIH, colesteatoma, enfermedad de Lyme	13%
Causas Neoplásicas	Schwannoma, teratoma, meningioma, enfermedad de Von Recklinghausen, hemangioma, tumor parotídeo, sarcoma, carcinoma (metastásico).	13%
Trauma	Al nacer, por forceps, cirugía parotídea, fractura de la mandíbula, lesiones sobre el nervio.	25%
Causas Congénitas	Síndrome de Moebius	9%
Causas Neurológicas	Síndrome opercular, síndrome de Wernicke-korsakoff, pseudotumor cerebral, infarto lacunar	
Toxinas	Etanol, talidomida, etilenglicol	

Tabla 1. Clasificación de Tipos de Parálisis Facial por Causa. A la tabla original de [5] se agregó la columna de proporción de casos por origen en base a información en [3].

Independiente de su origen, el diagnóstico de una parálisis facial requiere cuantificar su severidad. Las evaluaciones, dependiendo del uso de tecnología se clasifican en 1) clínicas o 2) No-clínicas (evaluación informática).

1.2. Evaluación clínica de la parálisis facial

En la evaluación clínica no se usan de elementos tecnológicos, y la evaluación se realiza en base a escalas o encuestas para asignar el grado de parálisis unilateral. La escala de House-Brackmann (HB) [1] es una de las más usadas, consta de seis categorías, en que el grado 1 corresponde a una persona sin parálisis y 6 a aquella con un grado severo. El médico solicita al paciente realizar una serie de gestos los cuales son comparados con las descripciones de la escala y así asignar el grado de parálisis.

La escala SunnyBrook [9] se ha ido posicionando como alternativa, ya que, entre otras diferencias con la escala HB, incorpora la evaluación por regiones del rostro, siendo este punto una de las observaciones frecuentes a la escala HB [9][11][12]. Cabe mencionar que para este estudio las evaluaciones HB se realizaron por zonas, ya que es la forma en que el equipo médico usa la escala. Una graduación general no permite la diferenciación; en [11] se observó que, al comparar evaluaciones por regiones versus evaluación general, se correlacionan mejor en la zona ocular, y la evaluación general tiende a ser más baja que la mayor por zonas (se evalúa al paciente con un grado de parálisis menor usando la evaluación general).

Independiente de la escala o encuesta utilizada, las valoraciones son realizadas por un médico especialista, lo que conlleva un grado de subjetividad.

Respecto a los niveles de concordancia (*Agreement*) y confiabilidad (medición con el indicador Kappa de Cohen o Fleiss) en [13] y [14] se reportan altos valores de kappa inter-evaluadores ($\text{kappa} > 0.8$), sin indicar los valores para la concordancia. En [15] se reporta una concordancia de un 93% para el estudio realizado, no se entregan los valores de kappa asociados. Este estudio arroja valores de Kappa variables en el rango 0.1-0.35. La concordancia también presenta variabilidad (entre un 23% y 47%). Los resultados se encuentran en la Tabla 4, casos 6 categorías HB. La experiencia

de los evaluadores se estima uno de los motivos para las diferencias en la evaluación [11].

1.3. Evaluación Informática de la parálisis facial (no clínica)

Los métodos que se apoyan en un soporte computacional se denominan métodos no-clínicos. Estos métodos buscan entregar una medición más objetiva y precisa de la parálisis facial [7] [8], sin embargo, su uso no se ha masificado. Los métodos no clínicos pueden requerir de elementos externos tales como marcadores visuales, mucho tiempo de preparación y así su uso resultar engorroso limitando su aplicación en entornos clínicos [2].

En [2] se propuso un método de graduación no clínico que detecta expresiones faciales asimétricas como se ilustra en la **Figura 1**. Se basa en un algoritmo de tracking facial (seguimiento de expresiones faciales) y no requiere de elementos externos como marcadores. A partir de los datos que entrega el tracking facial se obtienen **indicadores de los niveles de simetría (AUC)** de expresiones (ejemplo de una expresión es “estirar los labios hacia los lados”). El indicador del nivel de simetría mide a variabilidad del movimiento de un lado del rostro respecto del otro. Así los niveles de simetría son los datos de entrada a un árbol de decisión (algoritmo de graduación) para la determinación del grado de parálisis, que se ilustra en las **Figuras 2 y 3**.

El método propuesto en [2] entregó una graduación coherente con el grado de parálisis observado para el conjunto de pacientes utilizado, sin embargo, los resultados no se consideraron concluyentes en términos estadísticos debido a que el universo para la evaluación contó sólo con tres pacientes (N=3), sin alcanzar a cubrir las 6 categorías HB. Por otra parte, los umbrales de corte del árbol de decisión también se calcularon a partir de un universo de pacientes reducido (9 pacientes). Es así que no se descartó un fenómeno de sobreajuste (**overfitting**) del modelo.

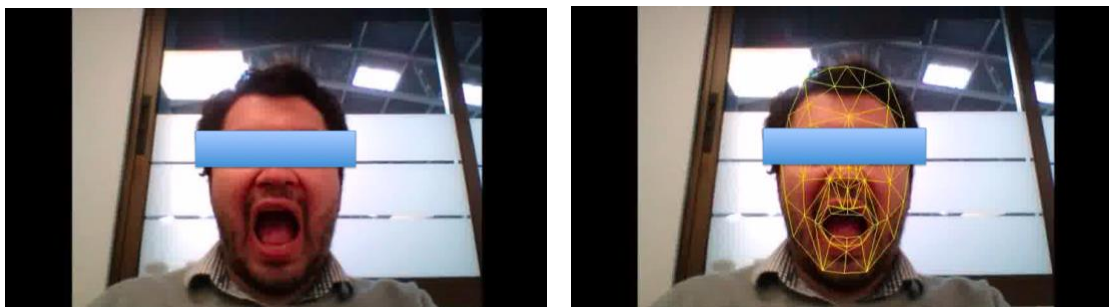


Figura 1. Video original y modelo de cara usado para seguimiento. En la imagen de la izquierda (frame de la grabación) se muestra a la persona efectuando un gesto a evaluar. En la segunda imagen se observa la máscara superpuesta usada en el proceso de evaluación del grado de parálisis, el modelo (máscara) se construye a posteriori y no forma parte de las actividades a realizar durante la grabación y evaluación clínica. Imágenes pertenecientes a material de apoyo de [2].

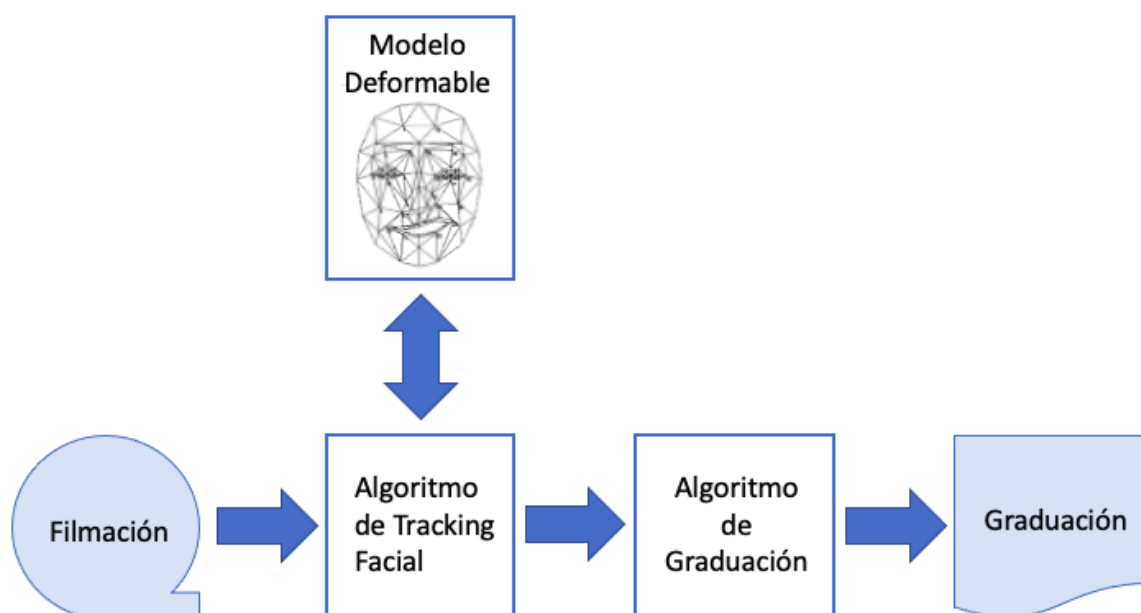


Figura 2. Método de Graduación con Tracking Facial.[2]

El método de graduación se compone de tres partes: un modelo deformable con soporte de asimetría, un algoritmo de tracking facial y un algoritmo de decisión que se apoya en los datos reportados por el tracking facial y la deformación del modelo para entregar la graduación. La entrada es una filmación y la salida es la graduación en escala HB. Diagrama desarrollado en [2]

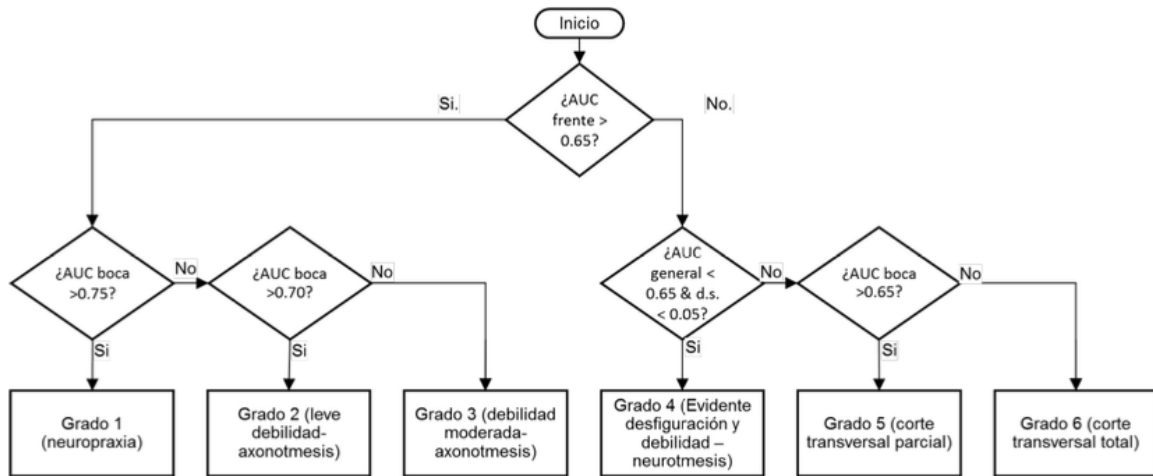


Figura 3. Algoritmo de Graduación. En [2] se propuso el algoritmo de valorización de la figura, que comparan valores de entrada (denominados AUC), con valores de corte previamente definidos.

1.4. Motivación

La subjetividad inherente a la graduación basada en métodos clínicos de la parálisis facial unilateral es motivación para el desarrollo de herramientas tecnológicas/computacionales de apoyo. Es así, que resulta de particular interés validar los resultados del trabajo desarrollado en [2] ya que el método descrito no requiere de elementos adicionales excepción de una cámara o celular para la realizar las grabaciones de los videos.

El árbol de decisión utilizado para la graduación en [2], es reemplazado por algoritmos de clasificación (*Machine Learning*) utilizando los indicadores de niveles de simetría calculados a partir de información del tracking facial.

2. Hipótesis y Objetivos

2.1. Hipótesis

La cuantificación/graduación de la parálisis facial obtenida usando indicadores de niveles de simetría desarrollados en [2] en algoritmos automáticos (ML), tiene al menos el mismo grado de exactitud que la medición actual de profesionales de la salud con una menor dispersión.

2.2. Objetivo General

Validar la cuantificación del grado de parálisis facial calculada con algoritmos automáticos (ML) que usan como variables explicativas indicadores de niveles de simetría.

2.3. Objetivos Específicos

1. Construir una base de datos de personas con parálisis facial a partir de videos disponibles en internet.
2. Entrenar algoritmos automáticos (ML) con ejemplos de mediciones de parálisis facial.
3. Definir protocolos para la evaluación del grado de parálisis facial para ser aplicados en una entidad hospitalaria y en el hogar del paciente.

3. Materiales y Métodos

3.1. Construir una base de datos de personas para cuantificar la parálisis facial a partir de videos disponibles en internet.

Inicialmente se definió que la muestra estaría conformada por pacientes sometidos a cirugía de parótida y con tratamiento finalizado o en período de seguimiento, del Hospital Clínico de la Universidad de Chile.

Sin embargo, con la llegada de la **pandemia** y su impacto en la movilidad, no fue posible contar con pacientes en la consulta hospitalaria y se analizaron otras alternativas; siendo la obtención de videos de internet la opción tomada. Esta alternativa permitió obtener un número levemente superior al mínimo tamaño de la muestra.

El tamaño mínimo se estimó en 24 casos, de forma de cubrir las categorías de la escala HB con datos suficientes para la ejecución de algoritmos de clasificación ML.

Los videos cumplen dos funciones:

1. Entrada (*input*) para el software basado en tracking facial y
2. Reemplazo a la evaluación en la consulta del médico

También se definió construir un cuestionario electrónico de forma que los médicos pudiesen ver los videos y registrar sus evaluaciones en una misma aplicación.

Así, la base de datos (repositorio de información) queda conformada por el conjunto de videos y las respuestas asociadas a los videos de cada médico evaluador.

3.1.1. Obtención de Videos

Desde internet (YouTube, Instagram, TikTok) se rescató un conjunto de videos de personas con parálisis facial, siendo seleccionados 30 para este proyecto de tesis. La mayoría de los videos, corresponden a casos de personas con parálisis facial idiopática (parálisis de Bell).

El uso de videos en las actividades indicadas en 3.1 requiere que las grabaciones cumplan con un conjunto de requisitos, entre los cuales cabe mencionar: nitidez, estabilidad de la grabación, iluminación adecuada (luz frontal), contener los gestos requeridos, velocidad de la gestualidad adecuada para el tracking, posición de la cara adecuada, entre otros. En el punto 4.3.4. se encuentra el detalle de los aspectos a considerar en la grabación del video.

Así, la selección de los 30 videos se basó en aquellos con mayor cumplimiento de los requisitos definidos y adicionalmente se buscó variabilidad en las severidades de parálisis observadas. En el **Anexo A** se entrega la lista de los casos utilizados en las evaluaciones.

3.1.2. Evaluación Médica y Registro

La metodología utilizada por el equipo médico (especialidad Cirugía Cabeza y Cuello) que evaluó los videos, analiza el rostro por zonas: frontal (frente), ocular, labio superior y labio inferior. Cada zona es evaluada por separado asignando valores entre 1 y 6 como en la escala HB tradicional.

The screenshot shows a Google Forms interface for a medical evaluation. It is titled '1. Paciente 1'. Below the title is a video player labeled 'video 1' showing a man's face with the text 'Caso 8 - día 14'. Below the video is a question '1.a ¿Cuál es el grado de parálisis? (por zona, escala HB) *'. This question has a table with 6 columns (1 to 6) and 4 rows (Zona frontal, Zona ocular, Zona labio superior, Zona labio inferior). Each cell contains a radio button. Below this is another question '1.b ¿Cuál es el grado de certeza en su evaluación? *'. This question has a row with 5 columns (1 to 5) and two labels: 'Incertidumbre total' and 'Gran certeza'. Each cell contains a radio button. At the bottom is a section '1.c Comentarios' with a text input field labeled 'Tu respuesta'.

	1	2	3	4	5	6
Zona frontal	☐	☐	☐	☐	☐	☐
Zona ocular	☐	☐	☐	☐	☐	☐
Zona labio superior	☐	☐	☐	☐	☐	☐
Zona labio inferior	☐	☐	☐	☐	☐	☐

	1	2	3	4	5	
Incertidumbre total	☐	☐	☐	☐	☐	Gran certeza

Figura 4. Ejemplo de pregunta del cuestionario para evaluación HB por zonas del rostro.

Para registrar la evaluación de los médicos se construyó un cuestionario con la herramienta Google Forms. En la Figura 4 se tiene una vista del cuestionario en que se observa que contiene el video a analizar. Además de la evaluación por

zonas, se consultó por el grado de certeza en las respuestas y se definió una sección para observaciones, esto con el objetivo de detectar posibles problemas asociados a la calidad de los videos.

3.2. Entrenar algoritmos automáticos de clasificación (ML) con ejemplos de mediciones de parálisis facial.

3.2.1 Datos para el entrenamiento de los algoritmos automáticos (ML)

Tamaño de la muestra

El tamaño de la muestra consideró 30 videos obtenidos de internet, de forma de cubrir las 6 categorías HB con suficientes datos para efectuar el entrenamiento y validación. Por cada categoría HB se debe contar con al menos 3 pacientes, por lo que 30 videos es un número levemente superior al mínimo requerido para el uso con algoritmos de clasificación.

El tamaño de las muestras observado en otros trabajos es variable; así, debido a que la incidencia es baja, aún en estudios extensos en el tiempo el número de sujetos es conservador. En [4] la muestra es de 41 pacientes, para un período de un año, en [7] el número de pacientes es de 50, y en (Ravikumar, Singh and Batish. 1999) de 25 pacientes.

A partir de los videos de pacientes se genera la siguiente información para ser utilizada por los algoritmos de clasificación (ML)

Indicadores de niveles de simetría

Para cada video de paciente se obtienen 7 indicadores de niveles de simetría (valores entre 0 y 1, en que 1 es la mayor simetría posible y 0 la menor), asociados a diferentes movimientos de zonas del rostro producto de gestos realizados. Ejemplo: al efectuar el gesto levantar las cejas se mueve tanto la parte interior de ceja como la

parte exterior, a estas acciones las denominamos “expresiones” y es a cada expresión que se calcula el nivel de simetría (lo denominamos indicador ya que toma valores entre 0 y 1). En el ejemplo anterior, la expresión asociada a que se levante la parte interior de la ceja se denomina **ibr** y a la exterior **obr**.

Las definiciones anteriores y su implementación fueron desarrolladas en [2],

El modulo de tracking facial y rutinas asociadas procesan el video de entrada entregando como resultado los niveles de simetría correspondientes. En la Figura 5 en el punto 3.2.3 se detalla el proceso.

Las expresiones sobre las que se calculan los niveles de simetría son las siguientes:

- nw:** Nose wrinkle (indicador de nariz arriscada)
- bl:** Brow lowerer (indicador de mover cejas hacia abajo)
- obr:** Outer Brow raiser (indicador de elevación de ceja zona exterior)
- ibr** Inner Brow raiser (indicador de elevación de ceja zona interior)
- ulr:** Upper Lip raiser (indicador de elevación de labio superior)
- ls:** Lip stretcher (indicador de estiramiento de labios hacia los lados)
- cd:** Lip corner depressor (indicador de mover labios hacia abajo)

En [2] se concluyó que este es el conjunto de indicadores relevante para la evaluación del grado de parálisis. También cabe mencionar que no hay una expresión (y por lo tanto indicador) asociado a la zona ocular, esto es debido a que el software de tracking facial no realiza un seguimiento adecuado sobre esa zona, siendo este punto una posible limitante para las predicciones de algoritmos ML.

La selección de los indicadores de niveles de simetría está estrechamente asociada a los movimientos que los pacientes deben realizar para su evaluación.

Evaluación HB

Cada video fue evaluado por médicos por separado utilizando la escala HB por zonas. Así, las grabaciones tienen asociadas evaluaciones HB para las siguientes zonas: zona Frontal (frente), zona Ocular, zona Labio Superior y zona Labio Inferior.

Debido a que la exactitud en las predicciones entregadas por los algoritmos de clasificación para 6 categorías HB arrojó bajos resultados en entrenamientos iniciales, se decidió fusionar las categorías HB para el análisis, bajo el supuesto que al agrupar la exactitud en las predicciones debía aumentar, lo que efectivamente se observó en los resultados finales. Se unificó de a dos y de a tres categorías adyacentes.

Cabe mencionar que, en una primera fase de la tesis, se realizó el entrenamiento tomando como valor HB el máximo entre las cuatro zonas evaluadas, sin embargo, con un mejor entendimiento del uso del grupo médico evaluador de esta escala, se optó por reprocesar por zonas. Realizando también un nuevo análisis exploratorio de los datos.

3.2.2 Herramientas

Se hace uso de los desarrollos de [2] que procesan los videos de entrada y entregan los indicadores de los niveles de simetría. (tracking facial)

Se incorpora el proceso de ML (algoritmos de clasificación) en reemplazo del algoritmo de graduación propuesto en [2] para las predicciones.

Se utilizó RStudio y las librerías correspondientes para la ejecución de algoritmos de clasificación y estadísticas.

3.2.3 Metodología

Cambios al Método de Graduación basado en Tracking Facial

La **Figura 5** muestra el flujo modificado para la estimación (predicción) del grado de parálisis por zona con escala HB. Los algoritmos de clasificación entregan las predicciones para el grado de parálisis en la escala HB por zona.

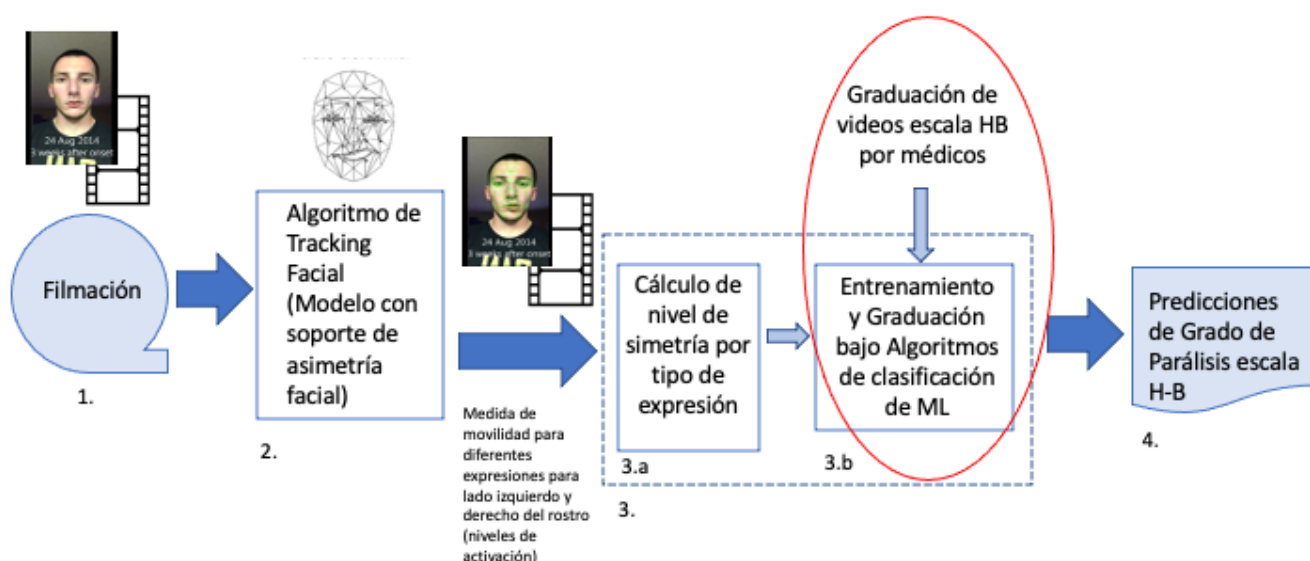


Figura 5. Flujo del Método de Graduación con Tracking Facial [2] modificado. El input (1) al tracking facial es un video en frames. Para cada video se ajusta manualmente una máscara (malla de puntos) sobre el rostro del paciente. El software de tracking facial a partir de esta máscara inicial efectúa el tracking. La salida de (2) entrega el video con las máscaras superpuestas para cada frame y un archivo con valores denominados **niveles de activación** asociados al grado de movimiento para expresiones definidas para cada lado del rostro. (3a) Los **niveles de activación** son los datos de entrada para calcular los indicadores de niveles de simetría por video para el conjunto de expresiones definido. (3b) el conjunto de los niveles de simetría para todos los videos y las evaluaciones de los médicos son los datos de entrada para el entrenamiento con los algoritmos de clasificación y 4.

30 videos fueron evaluados de manera independiente por 3 médicos utilizando la escala HB aplicada por zonas (frontal, ocular, labio superior, labio inferior).

Se entrenó con los siguientes 6 algoritmos de clasificación (lineales, no lineales y avanzados):

LDA: Linear Discriminant Analysis
CART: Classification & Regression Tree
KNN: K- nearest neighbors
SVM: Support Vector Machine
RF: Random Forest
NB: Naïve Bayes

El entrenamiento con algoritmos de clasificación (ML) considera:

- Evaluación HB por 4 zonas del rostro (Zona Frontal, Ocular, Labio Superior, Labio Inferior) por 3 médicos
- 7 indicadores de nivel de simetría (por video)
- 3 escenarios asociados a Categorías HB
 - 6 categorías HB (1 al 6)
 - 3 grupos, conformados por categorías (1,2), (3,4), y (5,6)
 - 2 grupos conformados por categorías (1,2,3) y (4,5,6)
- % de datos para entrenamiento y testing (la selección de datos es aleatoria)
- Parámetro de entrenamiento K-fold Cross Validation
- 6 algoritmos de clasificación ML

Los datos en entrenamiento y testing son divididos según el % indicado.

La ejecución con los datos de testing entrega entre otros datos, la exactitud (*Accuracy*) alcanzada.

Los 7 indicadores de nivel de simetría son predictores para ML.

Criterios para la Validación de Resultados

La validación de la hipótesis requiere comparar los resultados de la graduación de parálisis facial de la evaluación clínica versus la evaluación automática con algoritmos de clasificación que usan los indicadores de simetría desarrollados en [2].

Para alcanzar este propósito se construye una evaluación denominada ***Evaluación Promedio*** para ser usada como medida de comparación con las evaluaciones médicas y en las predicciones (evaluaciones) resultantes de los algoritmos de clasificación (ML).

En la ***Evaluación Promedio*** se asignó a cada video una graduación HB igual al promedio de las evaluaciones de los médicos (Dr1, Dr2, Dr3) para ese video, valor redondeado al entero más cercano.

La ***Evaluación Promedio*** es usada en las siguientes operaciones:

- Entrenamiento con los algoritmos automatizados (ML) cuyo resultado es la Exactitud (*Accuracy*) de las predicciones.
- Cálculo de la Concordancia y Confiabilidad para Evaluaciones Médicas.
- Comparación de la Exactitud (*Accuracy*) de los algoritmos de clasificación y la Concordancia (*Agreement*) de una cuarta evaluación clínica, ambas respecto de la ***Evaluación Promedio***.
- Cálculo de la Dispersión.

Con los resultados obtenidos se valida o rechaza la hipótesis respecto del grado de acuerdo (*Accuracy, Agreement*) y dispersión.

Métricas

- Para medir el grado de concordancia de los algoritmos de clasificación se utilizó la métrica **Accuracy** (Exactitud), que corresponde al % de casos predichos correctamente sobre el total de casos.
- Para medir el grado de concordancia entre médicos se utilizó la métrica **Agreement** (Concordancia), que es equivalente a **Accuracy** de ML.

Agreement y **Accuracy** se calculan de la misma manera y la diferencia en los nombres se debe a que la primera se usa en general, para comparar evaluaciones entre personas evaluadoras y la segunda respecto de herramientas automáticas, en este caso los algoritmos de clasificación (ML), en que se comparan las predicciones con los datos esperados.

- Para evaluar el grado de confiabilidad de la concordancia obtenida, se utilizan los indicadores Kappa de Cohen (entre dos evaluadores) y Kappa de Fleiss (para 3 o más evaluadores). El grado de confiabilidad también se denomina “fuerza de la Concordancia”.

Los valores de Kappa de Cohen y Fleiss varían entre -1 y 1. En la tabla siguiente se entrega su interpretación aplicable para ambos Kappas. Existen otras escalas más estrictas, sin embargo, esta es una de las más usadas.

Coeficiente Kappa	Fuerza de la Concordancia
<= 0,00	Pobre (<i>Poor Agreement</i>)
0,01 – 0,22	Leve (<i>Slight Agreement</i>)
0,21 – 0,40	Aceptable (<i>Fair Agreement</i>)
0,41 – 0,60	Moderada (<i>Moderate Agreement</i>)
0,61 – 0,80	Considerable (<i>Substantial Agreement</i>)
0,81 – 1,00	Casi Perfecta (<i>Almost Perfect Agreement</i>)

Tabla 2. Escala de valoración para Kappa de Cohen. (Landis & Coch 1977)

- Se calcula la **Dispersión** como la desviación media (promedio aritmético de las de la suma de las diferencias, en valor absoluto, respecto de la **Evaluación Promedio**)

$$DM = \frac{\sum_{i=1}^N |x_i - \bar{x}_i|}{N}$$

DM: Desviación Media

N: tamaño de la muestra, en este caso N videos

x_i : i-ésima graduación del video i-ésimo

$\bar{x}_i$: graduación promedio para video i-ésimo (i-ésimo dato de la **Evaluación Promedio**)

La fórmula para la dispersión indica que los valores posibles se mueven en el rango: 0 a 5. Siendo 5 el mayor grado de dispersión (diferencia máxima posible entre 2 evaluaciones), y su valor refleja en valor absoluto la distancia respecto del valor promedio esperado.

3.3. Definir protocolos para la evaluación del grado de parálisis facial para ser aplicados en una entidad hospitalaria y en el hogar del paciente.

Para la autorización de la grabación de videos de pacientes se desarrollaron dos protocolos de consentimiento informado:

3.3.1. Protocolo presencial. Para ser aplicado en la consulta médica con el objetivo de obtener la evaluación del médico del grado de parálisis facial del paciente y un video del paciente para una posterior evaluación automatizada (método no clínico).

3.3.2. Protocolo remoto. Para ser aplicado en el hogar del paciente y grabarlo. En este caso el video no sólo es usado para una posterior evaluación automatizada; también es usado para la evaluación por parte del médico. Se desarrolló este segundo protocolo producto de las restricciones de movilidad y riesgos asociados a la pandemia. Esta alternativa se considera mejor que la obtención de videos desde internet, por varios motivos entre ellos porque los gestos grabados son los requeridos.

Cabe recalcar que, en los controles post operación de pacientes sometidos a cirugía de parótida, la evaluación bajo escala HB por zonas es parte del protocolo habitual.

4. Resultados Obtenidos

4.1. Análisis de la base de datos construida

Se efectúa un análisis exploratorio de los datos para verificar la diversidad del universo de casos requerida para el entrenamiento de los algoritmos (ML).

4.1.1 Niveles de Simetría para el Conjunto de 30 Videos Utilizados

Se graficaron los niveles de simetría para los 30 videos de este estudio, que incluye 6 ya utilizados en [2] de pacientes representativos de cada uno de los grados de la escala HB.

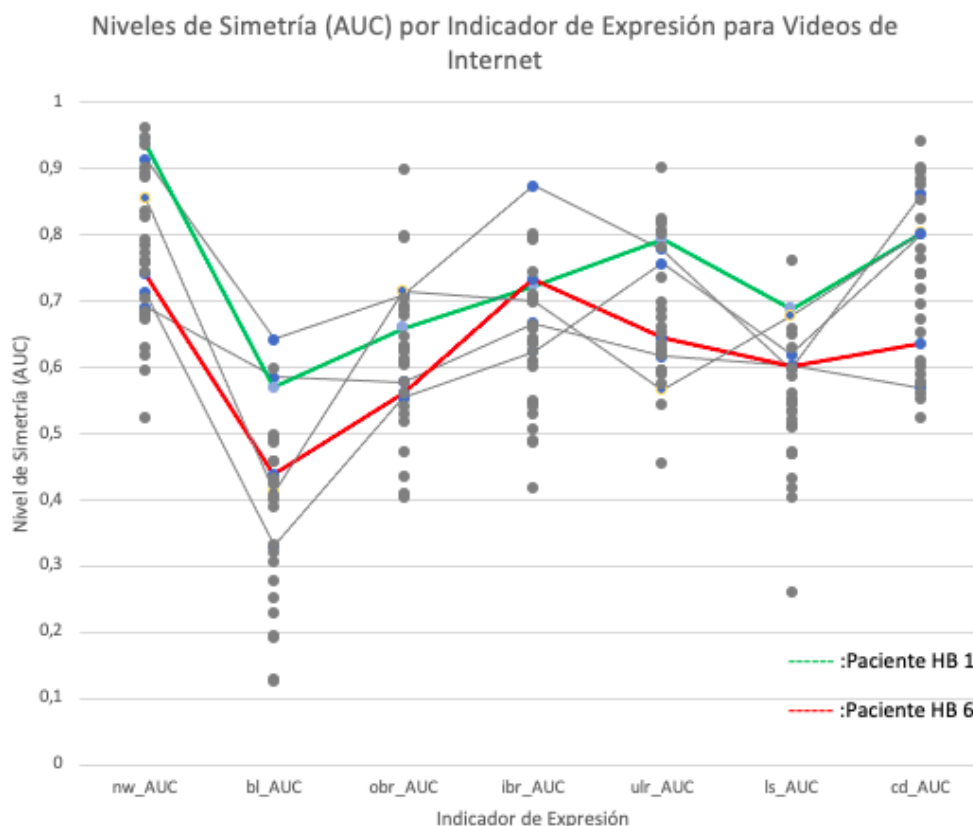


Figura 6. Por cada uno de los 30 videos se grafican con puntos grises los niveles de simetría por Indicador de Expresión. Los 7 indicadores de expresión utilizados se asocian a los siguientes sectores del rostro: nariz (nw_AUC), frente (bl_AUC, obr_AUC, ibr_AUC) y boca (ulr_AUC, ls_AUC, cd_AUC)). Los puntos unidos por líneas corresponden a los 6 videos tipo utilizados originalmente en [2], cuyas evaluaciones HB cubren las 6 categorías de la escala. Las líneas verde y roja corresponden a los videos tipo con pacientes etiquetados HB=1 y HB=6.

Se observa que los niveles de simetría se mueven en distintos rangos dependiendo del Indicador de Expresión, y que el conjunto total de videos presenta un comportamiento similar al de los 6 videos aunque con una mayor dispersión y con cotas inferiores por debajo de las de los videos tipo (mayor asimetría). La mayor asimetría puede tener orígenes diversos, tales como: un tracking defectuoso, o que los gestos en los videos no son exactamente los mismos gestos en los 6 videos tipo.

Ya que los rangos registrados para los videos de prueba incluyen a los de los videos tipo se estima de manera preliminar que la muestra es representativa y cubriría los diferentes grados de parálisis.

4.1.2 Distribuciones respecto de la clasificación HB

Para cada evaluación médica se graficó la distribución de casos por zona (zona frontal, zona ocular, labio superior, labio inferior) para conocer la cobertura de las categorías HB y su balance.

Se construyó un segundo gráfico con la distribución de los niveles de simetría por categoría HB abierto por expresión (nw-nose wrinkler, bl-brow lowerer, obr-outer brow raiser, ibr-inner brow raiser, ulr-upper lip raiser, ls-lip stretcher, cd-lip corner depresor y promedio) con el objetivo de visualizar posibles relaciones de los datos. En este gráfico está plasmada la misma información que utilizan los algoritmos de clasificación (ML).

A continuación, se muestra un caso de ejemplo.

La totalidad de los gráficos se encuentran en el **Anexo B**.

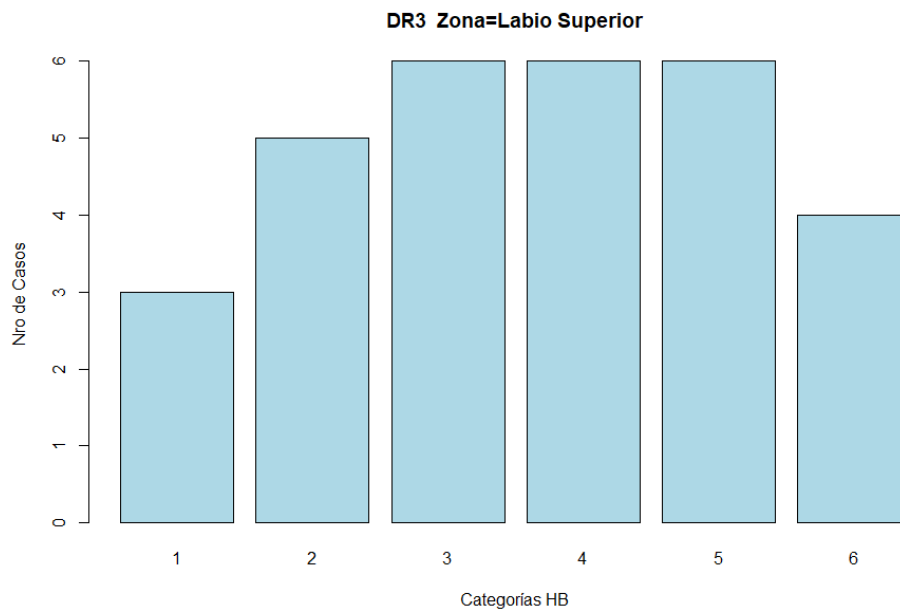


Figura 7. Distribución de casos por categoría HB

En la Figura 7 se observa una balanceada distribución de casos por categoría HB. La distribución de casos depende de la evaluación médica y de la zona evaluada como se observa en las gráficas del Anexo B.

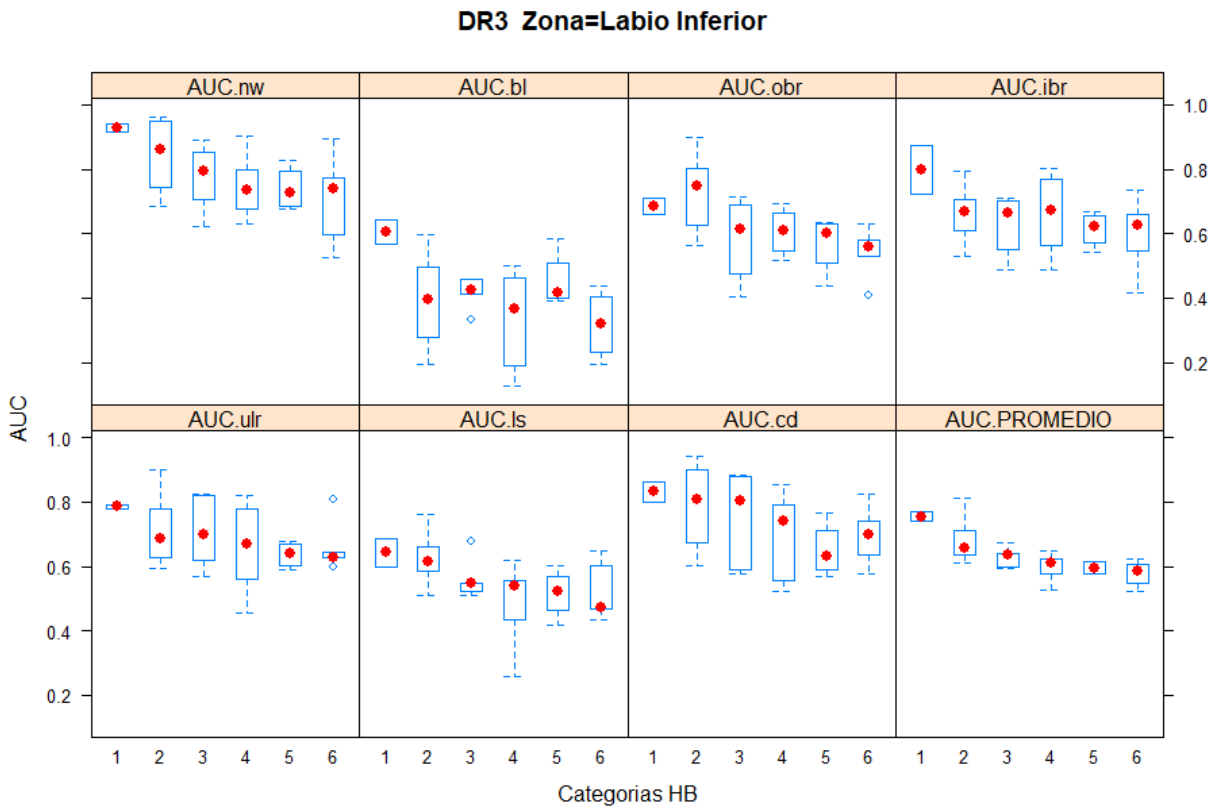


Figura 8. Distribución de niveles de simetría (AUC) por categoría HB por expresión

En la gráfica de la Figura 8 se observa:

- La relación inversa entre la categoría HB y el nivel de simetría, lo que era lo esperado, ya que mayor parálisis facial unilateral (mayor valor en la escala HB) menor nivel de simetría.
- Que hay traslape en las dispersiones entre categorías.
- Que los rangos de los niveles de simetría varían respecto de la expresión (ya observado en 4.1.1.).

4.2. Resultados de los Entrenamientos con Algoritmos de Clasificación (ML)

4.2.1 Análisis de *Accuracy* (Exactitud) para la Evaluación Promedio

En la Tabla 3 se entregan los niveles de Exactitud de entrenar con la ***Evaluación Promedio***. Los resultados de los entrenamientos individuales con las evaluaciones de cada médico (Dr1, Dr2, y Dr3) se encuentran en el **Anexo C**.

Médico	Nro. Categorías HB	Zona	Exactitud (Accuracy) por Algoritmo de Clasificación						MAX Accuracy	AVERAGE Accuracy	Promedio AVE Acc. por Zona			
			Ida	Cart	KNN	SVM	RF	Nbayes						
Evaluación Promedio (Dr123)	6	Frontal	42,9%	27,3%	42,9%	42,9%	36,4%	42,9%	42,9%	39,2%	51,7%			
	3		42,9%	42,9%	42,9%	40,0%	80,0%	80,0%	80,0%	54,8%				
	2		50,0%	66,7%	66,7%	66,7%	66,7%	50,0%	66,7%	61,1%				
	6	Ocular	44,4%	45,5%	44,4%	45,5%	33,3%	33,3%	45,5%	41,1%	60,5%			
	3		60,0%	60,0%	60,0%	60,0%	60,0%	60,0%	60,0%	60,0%				
	2		100,0%	83,3%	83,3%	83,3%	83,3%	50,0%	100,0%	80,5%				
	6	Labio Lsuperior	22,2%	30,0%	33,3%	30,0%	11,1%	44,4%	44,4%	28,5%	54,3%			
	3		40,0%	60,0%	42,9%	60,0%	60,0%	60,0%	60,0%	53,8%				
	2		83,3%	66,7%	83,3%	83,3%	83,3%	83,3%	83,3%	80,6%				
	6	Labio Linferior	27,3%	27,3%	44,4%	44,4%	33,3%	11,1%	44,4%	31,3%	56,5%			
	3		40,0%	60,0%	80,0%	60,0%	60,0%	80,0%	80,0%	63,3%				
	2		83,3%	66,7%	66,7%	66,7%	83,3%	83,3%	83,3%	75,0%				
Promedios	Clasificadores		53,0%	53,0%	57,6%	56,9%	57,6%	56,5%						
	Categorías													
	6 categorías		34,2%	32,5%	41,3%	40,7%	28,5%	32,9%						
	3 categorías		45,7%	55,7%	56,4%	55,0%	65,0%	70,0%						
	2 categorías		79,2%	70,8%	75,0%	75,0%	79,2%	66,7%						

Tabla 3. Niveles de Exactitud (*Accuracy*) por algoritmo, por zona, y número de categorías HB (agrupaciones) para la ***Evaluación Promedio***.

Se observa que:

- Al agrupar las categorías HB los algoritmos de clasificación aumentan en promedio la exactitud de las predicciones, como se puede observar en la columna Average Accuracy. A menor número de agrupaciones mejor Accuracy.
- Para 6 categorías HB el Máximo Accuracy por zona varía entre 42,9% y 45,5%.

- Los valores de Accuracy Promedio de la Zona Frontal y Ocular registran valores aproximadamente un 10% superiores respecto de las zonas Labio Superior e Inferior.
- En promedio los algoritmos LDA y Cart presentan los valores más bajos de Accuracy del grupo; KNN y RF los más altos, sin embargo, las diferencias son menores al 5%.
- El Accuracy Promedio por zona varía entre 51,7 y 60,5% (Tabla 3 columna gris del lado derecho).

4.2.2 Análisis de Concordancia y Confiabilidad para las Evaluaciones Clínicas

A continuación, se encuentran registrados en la Tabla 4 los niveles de Concordancia (*Agreement*) entre pares de médicos evaluadores y sus los niveles de confiabilidad (*Reliability*, medido usando el indicador kappa de Cohen). En la Tabla 5 se registran los datos para los mismos médicos evaluadores pero esta vez respecto de la ***Evaluación Promedio***.

Se observa que,

- Los registros de las Tablas 4 y 5 muestran que los niveles de Concordancia (*Agreement*) aumentan al comparar con la ***Evaluación Promedio***, lo que refleja el sesgo al usar un promedio obtenido a partir de los mismos datos.
- La Confiabilidad (kappa de Cohen) que mide la fuerza de la concordancia entre dos evaluadores, también aumenta al usar la ***Evaluación Promedio***.
- Respecto del número de categorías usadas para clasificar, en las Tablas 4 y 5 se observa que la Concordancia y Confiabilidad aumentan al disminuir el número de categorías a 3 y 2 comparado con 6 categorías. Sin embargo, la confiabilidad registra, en varios casos, valores más altos para 3 categorías que para 2 categorías, reflejando esto último, predicciones correctas probablemente producto del azar (al disminuir el número de categorías aumentaría la probabilidad de acertar por azar).

- Los niveles de Confiabilidad para la Concordancia inter-evaluadores (Tabla 4) toma valores entre 0,10 (concordancia leve) y 0,69 (concordancia considerable).
- Los niveles de Confiabilidad para la Concordancia respecto de la Evaluación promedio toman valores entre 0,22 (leve) y 0,87 (casi perfecta).
- Los niveles de Concordancia son mayores para la zona frontal respecto de las otras zonas (mayor consenso entre evaluadores).

Zona	Nro Categorías HB	Concordancia (Agreement) %			Confiabilidad (kappa de Cohen)			AVE Agree DRs %	AVE Kappa DRs
		DR1-DR2	DR1-DR3	DR2-DR3	DR1-DR2	DR1-DR3	DR2-DR3		
Frontal	6	46,70	46,70	43,30	0,35	0,35	0,32	45,57	0,34
	3	70,00	73,30	80,00	0,53	0,59	0,69	74,43	0,60
	2	80,00	76,70	76,70	0,58	0,52	0,53	77,80	0,54
Ocular	6	40,00	36,70	40,00	0,28	0,19	0,29	38,90	0,25
	3	66,70	66,70	66,70	0,49	0,48	0,52	66,70	0,50
	2	70,00	70,00	73,30	0,40	0,32	0,47	71,10	0,39
Labio Superior	6	23,30	46,70	30,00	0,10	0,35	0,15	33,33	0,20
	3	56,70	73,30	66,70	0,36	0,59	0,51	65,57	0,49
	2	66,70	83,30	76,70	0,37	0,66	0,54	75,57	0,52
Labio Inferior	6	26,70	26,70	30,00	0,12	0,12	0,15	27,80	0,13
	3	40,00	63,30	70,00	0,11	0,44	0,56	57,77	0,37
	2	66,70	76,70	76,70	0,34	0,52	0,54	73,37	0,47
Promedios:		54,46	61,68	60,84	0,33	0,43	0,44	58,99	0,40
6 categorías		34,18	39,20	35,83	0,21	0,25	0,23		
3 categorías		58,35	69,15	70,85	0,37	0,52	0,57		
2 categorías		70,85	76,68	75,85	0,42	0,50	0,52		

Tabla 4. Concordancia y Confiabilidad de las evaluaciones clínicas entre médicos.

Zona	Nro Categorías HB	Concordancia (Agreement) %			Confiabilidad (kappa de Cohen)			AVE Agree DRs %	AVE Kappa DRs
		DR1 y EvProm	DR2 y EvProm	DR3 y EvProm	DR1 y EvProm	DR2 y EvProm	DR3 y EvProm		
Frontal	6	63,30	60,00	56,70	0,55	0,51	0,48	60,00	0,51
	3	83,30	80,00	83,30	0,74	0,70	0,75	82,20	0,73
	2	86,70	86,70	90,00	0,72	0,73	0,80	87,80	0,75
Ocular	6	63,30	56,70	53,30	0,54	0,47	0,41	57,77	0,47
	3	90,00	76,70	76,70	0,84	0,65	0,61	81,13	0,70
	2	80,00	90,00	83,30	0,56	0,80	0,65	84,43	0,67
Labio Superior	6	63,30	36,70	66,70	0,55	0,22	0,59	55,57	0,46
	3	86,70	66,70	86,70	0,78	0,52	0,79	80,03	0,70
	2	83,30	83,30	93,30	0,66	0,67	0,87	86,63	0,73
Labio Inferior	6	43,30	50,00	60,00	0,31	0,40	0,50	51,10	0,40
	3	63,30	70,00	86,70	0,42	0,57	0,79	73,33	0,59
	2	83,30	83,30	93,30	0,66	0,67	0,86	86,63	0,73
Promedios:		74,15	70,01	77,50	0,61	0,57	0,68	73,89	0,62
6 categorías		58,30	50,85	59,18	0,49	0,40	0,50		
3 categorías		80,83	73,35	83,35	0,69	0,61	0,73		
2 categorías		83,33	85,83	89,98	0,65	0,72	0,79		

Tabla 5. Concordancia y Confiabilidad de las evaluaciones clínicas respecto de la Evaluación Promedio.

En el Anexo E punto II, se incluye el cálculo de la concordancia de un cuarto evaluador (Dr4) respecto de la Evaluación Promedio. Se observa que los niveles de concordancia de la cuarta evaluación son menores a las de los médicos Dr1, Dr2, y Dr3, cuyos sus datos se usaron en la construcción de la Evaluación Promedio.

En el Anexo E punto II, se incluyen el cálculo de la concordancia para 3 evaluadores, donde se observa que la zona frontal presenta mayor concordancia para todas las agrupaciones de categorías, al igual que en la comparación de pares.

4.2.3 Comparación de los Niveles Exactitud y Concordancia de los algoritmos de clasificación y las evaluaciones clínicas

La Tabla 6 Contiene los resultados de los niveles de exactitud (*Accuracy*) para el entrenamiento de los algoritmos de clasificación con la **Evaluación Promedio** y los niveles de Concordancia (*Agreement*) de los médicos Dr1, Dr2, Dr3, y Promedio Drs respecto de la misma **Evaluación Promedio**.

Zona	Nro Categorías HB	Concordancia %						Diferencia DRs - MLMAX Acc	Diferencia DRs - MLAVE Acc
		DR1	DR2	DR3	Average DRs	ML MAX Accuracy	ML Average Accuracy		
Frontal	6	63,30	60,00	56,70	60,00	42,86	39,18	17,14	20,82
	3	83,30	80,00	83,30	82,20	80,00	54,76	2,20	27,44
	2	86,70	86,70	90,00	87,80	66,67	61,11	21,13	26,69
Ocular	6	40,00	43,30	53,30	45,53	45,45	41,07	0,08	4,46
	3	66,70	66,70	76,70	70,03	60,00	60,00	10,03	10,03
	2	70,00	73,30	83,30	75,53	100,00	80,55	-24,47	-5,02
Labio Superior	6	63,30	36,70	66,70	55,57	44,44	28,52	11,13	27,05
	3	86,70	66,70	86,70	80,03	60,00	53,81	20,03	26,22
	2	83,30	83,30	93,30	86,63	83,33	80,55	3,30	6,08
Labio Inferior	6	43,30	50,00	60,00	51,10	44,44	31,31	6,66	19,79
	3	63,30	70,00	86,70	73,33	80,00	63,33	-6,67	10,00
	2	83,30	83,30	93,30	86,63	83,33	75,00	3,30	11,63
Promedios:		69,43	66,67	77,50	71,20	65,88	55,77	5,32	15,43
6 categorías		52,48	47,50	59,18	53,05	44,30	35,02		
3 categorías		75,00	70,85	83,35	76,40	70,00	57,98		
2 categorías		80,83	81,65	89,98	84,15	83,33	74,30		

Tabla 6. Comparación de los valores concordancia (*Agreement*) para médicos Dr1, Dr2, Dr3, promedio Drs. y valores de exactitud (*Accuracy*) de los Algoritmos de clasificación respecto de la **Evaluación Promedio**.

Los resultados muestran que la concordancia de las evaluaciones clínicas es más alta para todos los casos con escala HB de 6 categorías y en los promedios, que el *Accuracy* alcanzado por los algoritmos de clasificación (ML), sin embargo, si sólo se realiza la comparación con los médicos Dr1 y Dr2 por una parte y por otra se toman los resultados de los algoritmos con mejor desempeño la diferencia porcentual para HB 6 categorías es <5% en las 2 comparaciones (Dr1 vs ML MAX = 69,43 – 65,88 = 3,55%; Dr2 vs ML MAX= 66,67 – 65,88 = 0,79%). El impacto de la alta concordancia de Dr3 y la Evaluación Estándar justifica este análisis.

Los niveles de Concordancia de las evaluaciones comparados con los de *Agreement* de lo algoritmos (ML) disminuyen su brecha al disminuir el número de categorías.

Los resultados no son concluyentes respecto a la superioridad de la Concordancia de la evaluación clínica por sobre la evaluación automática, más aún teniendo en cuenta que la **Evaluación promedio** fue construida con los datos de los mismos médicos.

En el entrenamiento de los algoritmos (ML) también se utilizaron los datos de la **Evaluación promedio**, pero cabe recordar que la predicción de los algoritmos ML se realiza sobre una partición de los datos definida para este fin y que partición definida para entrenamiento no usa los datos de *testing*.

4.2.4 Comparación de los Niveles de Dispersión

Los niveles de dispersión se calcularon respecto de los datos de la **Evaluación Promedio**, de acuerdo con la fórmula descrita en el punto 3.2.3. El cálculo se efectuó sobre las evaluaciones con apertura en 6 categorías HB.

Para los algoritmos de ML se tomaron las predicciones de un conjunto de datos de pruebas (*testing ML*) y se calculó la dispersión respecto de los valores esperados (los de la **Evaluación Promedio**).

Nota: La dispersión refleja la diferencia promedio (o distancia) respecto de la evaluación esperada tomando posibles valores entre 1 y 5, no es un porcentaje.

	Dr1	Dr2	Dr3	Promedio (Dr1, Dr2, Dr3)	ML	diff % disp Ave Drs respecto ML
Zona Frontal	0,37	0,47	0,47	0,43	0,30	13,33
Zona Ocular	0,37	0,43	0,50	0,43	0,91	-47,57
Zona Labio Superior	0,37	0,70	0,33	0,47	0,80	-33,33
Zona Labio Inferior	0,63	0,53	0,43	0,53	0,50	3,33
Promedios	0,43	0,53	0,43	0,47	0,63	-16,06

Tabla 7. Dispersiones Evaluaciones médicos Dr1, Dr2, Promedio y Algoritmos ML respecto de la **Evaluación Promedio**.

Los resultados obtenidos en este caso tampoco son concluyentes respecto a la superioridad de la evaluación clínica por sobre la automática. En el caso de los algoritmos ML se observa una dispersión mucho más alta para la Zona ocular y Labio Superior lo que sube el promedio usado para comparar. En el punto 5 Discusión se analizan limitantes que pudieron afectar el mejor desempeño de los algoritmos de clasificación (ML). Al analizar por zonas se observan 2 zonas en que las dispersión es mayor para ML (diferencias de un 47% y 33%), y otras 2 zonas en que es mayor para la evaluación clínica (13% y 3%).

4.3. Protocolos para la evaluación del grado de parálisis facial para ser aplicados en una entidad hospitalaria y en el hogar del paciente.

4.3.1. Protocolo para consulta médica sin restricciones de pandemia

A continuación, se ilustran los cambios introducidos al procedimiento actual de evaluación del grado de parálisis facial para control en consulta médica.

1. Flujo sin pandemia en consulta

Flujo Actual



- Paciente acude a consulta para evaluación post cirugía
- Se evalúa grado de parálisis usando la escala House-Brackmann por zonas del rostro

Flujo Modificado

- Paciente acude a consulta para evaluación post cirugía
- Se solicita formalmente al paciente dar su consentimiento para ser grabado durante la evaluación



- Se evalúa grado de parálisis usando la escala House-Brackmann por zonas del rostro
- y
- la secuencia de gestos que debe realizar el paciente es grabada



Figura 9. Protocolo Caso paciente en la consulta médica. En la Figura se ilustra el flujo actual del profesional de la salud (izquierda) y el propuesto, donde se realiza una grabación de video.

El protocolo modificado incorpora dos nuevas etapas: Autorización formal del paciente (consentimiento), y la realización y grabación de gestos. El impacto de las modificaciones se estima bajo tanto en aspectos logísticos como en el tiempo adicional requerido. El tiempo adicional se estima inferior a 10 minutos. La obtención de la autorización requiere que el paciente sea claramente informado en que consiste la grabación y esta etapa tomar más tiempo. La grabación toma menos de un minuto.

4.3.2. Protocolo para grabación de Video bajo Pandemia.

A continuación, se ilustran los cambios introducidos al procedimiento actual para la grabación de un video en el hogar del paciente.

FLUJO BAJO PANDEMIA

Flujo Actual



- Paciente acude a consulta para evaluación post cirugía
- Se evalúa grado de parálisis usando la escala House-Brackmann por zonas del rostro

Flujo Modificado

I. Obtención Consentimiento Informado

Se contacta al paciente para invitarlo a participar en evaluaciones remotas del grado de parálisis facial. Si el paciente acepta, deberá indicar como prefiere dar su consentimiento.

a) Firmando documento físico. En este caso se acude al hogar del paciente para obtener firma. (Presencial).
o alternativamente (opción deseable),

b) Dando consentimiento por medio electrónico (No Presencial)

- En este caso se envía por medio electrónico documento de consentimiento informado para su revisión
- Se realiza 2º contacto telefónico para aclaraciones y formalización del consentimiento.
- El consentimiento se registra mediante: grabación (audio/video) de aceptación o correo con visto bueno.



II. Instrucciones

- Se envía video vía celular con instrucciones para grabación, y se indica que telefónicamente se pueden aclarar dudas.

III. Grabación

- Se realiza grabación con celular en el hogar del paciente.
- La puede realizar el mismo paciente o con ayuda de otra persona.



- Paciente envía grabación vía celular al médico tratante.

IV. Evaluación

- Una vez recibida la grabación Se evalúa el grado de parálisis usando la escala House-Brackmann por zonas
- La evaluación se realiza tanto por la herramienta automatizada como por los médicos.

Figura 10. Protocolo Caso paciente en su hogar. En la Figura se ilustra el flujo actual del profesional de la salud (izquierda) y el flujo para realizar la grabación del video en el hogar del paciente.

Para verificar la factibilidad de realizar grabaciones en el hogar del paciente y definir un protocolo, se grabaron videos a través de sesiones de Zoom y WhatsApp entre personas del equipo. Estos videos fueron inyectados al software de tracking facial sin inconvenientes.

4.3.3. Criterios de Selección de Pacientes

Los criterios siguientes aplican a la selección de posibles pacientes a grabar en general ampliando los casos más allá de los pacientes de operados de parótida. Y no aplica al caso de los videos obtenidos de internet.

Criterios de inclusión

- Pacientes capaces de recibir instrucciones.
- Pacientes con las siguientes patologías o diagnósticos:
 - Pacientes que ya han sido sometidos a cirugía de extirpación de parótidas y presentan parálisis facial unilateral.
 - Pacientes con parálisis de Bell (no requieren cirugía)
 - Pacientes con parálisis por otitis
 - Pacientes con parálisis por Herpes Zoster ótico

Criterios de exclusión

- Pacientes con parálisis facial bilateral
- Pacientes incapacitados para recibir instrucciones.
- Parálisis Facial de origen traumático no quirúrgico, yatrogénica, Síndrome Guillain Barré, otros.

Privacidad y manejo de datos: anonimizar nombres, y los videos.

Se considera y asegura el anonimato de todos los pacientes involucrados, en cuanto a nombres e imágenes en videos o fotografías. Se utiliza un identificador (ID) por cada paciente.

Consentimiento

Se dispone de un formulario de “Consentimiento Informado” para que el paciente autorice formalmente las grabaciones. El documento especifica las dos alternativas para realizar la grabación: aquella en que el paciente acude a la consulta médica y el caso en que la grabación se realiza en el hogar del paciente.

Ver **Anexo D**.

4.3.4. Instrucciones para la Grabación de un Video

Los siguientes puntos deben considerarse tanto para la grabación en la consulta médica como en la grabación en el hogar del paciente.

1. Lente de la cámara o celular a la altura de los ojos del paciente
2. Rostro del paciente despejado. Por ejemplo: eliminar chasquilla que pudiese cubrir cejas.
3. Iluminación pareja, de manera que no genere sombras, iluminando de frente al rostro, no desde atrás como en el caso de una ventana.
4. Fondo despejado y claro.
5. Distancia del lente al rostro: 40 y 45 cm aproximadamente
6. Fijar cámara o celular de manera de obtener un video estable
7. La grabación debe efectuarse en lo posible entre 60 y 120 frames/segundo (esta indicación está pensada para la grabación en la consulta).

Respecto de los gestos a realizar, se debe tener en cuenta:

8. Que el paciente efectúe los movimientos a una velocidad moderada, es preferible que los realice lentamente.
9. El paciente trate de mantener la cabeza fija al efectuar los gestos solicitados (no mover la cabeza, sólo efectuar los gestos).

Los gestos que el paciente debe realizar son los siguientes:

1. Levantar las cejas
2. Cerrar ambos ojos
3. Arriscar la nariz (oler algo desagradable)
4. Mostrar todos los dientes sonriendo
5. Tirar un beso hacia adelante

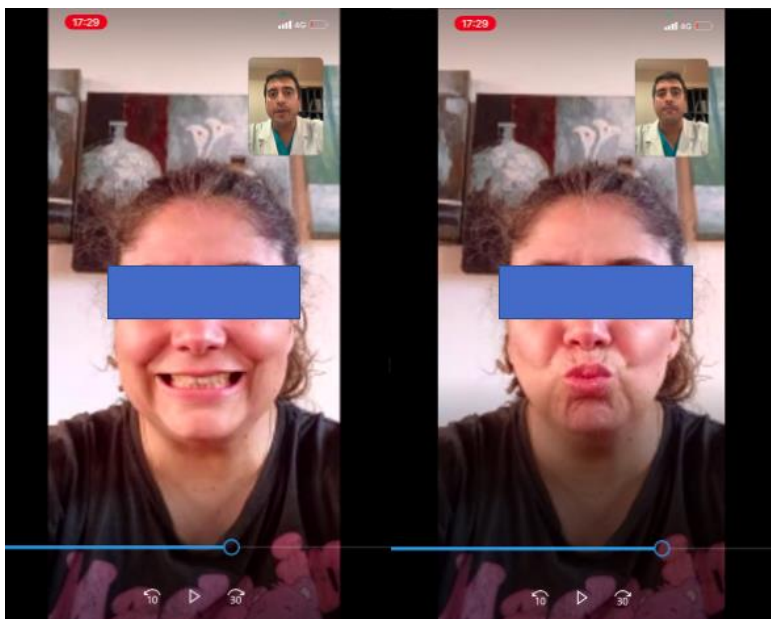


Figura 11. Caso real de paciente realizando la secuencia de gestos desde su celular. Etiquetado por Equipo Médico de Cirugía Cabeza y Cuello Hospital Clínico de la Universidad de Chile.

5. Discusión

La validación de la hipótesis requirió el desarrollo de tres objetivos específicos, cuyos resultados son discutidos a continuación.

Base de Datos

La base de pacientes se formó a partir de videos rescatados de internet. El conjunto de videos resultó heterogéneo respecto a calidad y contenido, no cumpliendo con la totalidad de los requisitos definidos. Esta falencia puede conducir a un tracking menos preciso y, por lo tanto, la información que alimenta a los algoritmos de clasificación (ML) también contener imprecisiones que distorsionen las predicciones. Sin embargo, a pesar de no contar con videos óptimos en calidad (requisitos en punto 4.3.4) el análisis de los datos no descartó su uso (punto 4.1).

Por otra parte, el no contar con videos con el contenido adecuado (gestos requeridos) se traduce en una evaluación médica con un mayor grado de incertidumbre y falta de datos en el tracking para esos gestos y así un cálculo de los niveles de simetría incompleto/inexacto.

Respecto del tamaño de la muestra (número de videos en la base de datos), es levemente superior al mínimo requerido para entrenar. Sin embargo, para testear es deseable contar con mayor cantidad casos. Si se toma un 80% de los datos para entrenamiento y 20% para pruebas, entonces en 30 casos, sólo hay sólo 6 para probar y obtener estadísticas. Por lo tanto, se estima deseable una base de datos del doble del tamaño actual (60 casos), lo cual puede ser lento y difícil de obtener; en el caso de la parálisis de Bell la incidencia es del orden 40 casos por cada 100.000 habitantes [6].

El uso de videos de internet llevó a la creación de un cuestionario en línea para que los médicos pudieran efectuar las evaluaciones. El cuestionario resultó sencillo de usar con un tiempo estimado de 1 1/2 horas para responder la encuesta de 30

pacientes, esto es 3 minutos por paciente. La información video y evaluación se puede conformar a futuro en una base para seguimiento en el tiempo de pacientes, con un registro histórico visual.

A raíz de la necesidad de obtención de videos producto de la pandemia, se investigó la existencia de bases de videos de parálisis facial públicas, encontrándose sólo una, que no pudo utilizarse por tratarse de secuencias de cuadros de videos no todos consecutivos. Por este motivo, el disponer de una base de datos contribuye al menos a facilitar casos de prueba en el desarrollo de herramientas de tracking de expresiones asimétricas.

Entrenamiento de algoritmos automáticos (ML) de Clasificación

En promedio los algoritmos LDA y Cart presentan los valores más bajos de Accuracy del grupo; KNN y RF los más altos, sin embargo, las diferencias son menores al 5%. Los resultados del entrenamiento con la Evaluación Promedio muestran que al disminuir el número de categorías (agrupando clases) aumenta -en general- la exactitud (**Accuracy**). Para el escenario con 6 categorías HB con algoritmo KNN, el Accuracy aumenta de 0,41 (41% de acierto) a valores entre 56% y 75% de acierto para escenarios con 3 y 2 categorías.

La clasificación en el escenario de 3 grupos equivale a clasificar el grado de parálisis en bajo, medio, y alto. Si se consideran 2 grupos equivale a dividir en: bajo y alto. Queda validar si es de alguna utilidad este nivel de agregación.

La Concordancia entre médicos es más baja que la Concordancia relativa a la Evaluación Promedio. Las diferencias porcentuales observadas para datos promediados varían entre un 8% y un 23%. Con el indicador Kappa ocurre lo mismo variando entre un 12% y un 28%. Las diferencias son atribuibles a que la **Evaluación Promedio** se construye a partir de los mismos datos que se comparan.

Los niveles de Concordancia y Kappa respecto del la ***Evaluación promedio*** son inferiores a los reportados en [13], [14]. Y [15]. Atribuible a los diferentes niveles de experiencia de los evaluadores [11] y también a que el conjunto de videos evaluado no cumple con la totalidad de los gestos requeridos.

La concordancia de las evaluaciones clínicas es más alta para todos los casos con escala HB de 6 categorías y en los promedios, que el *Accuracy* (Máximo y promedio) alcanzado por los algoritmos de clasificación (ML). Sin embargo, al comparar *Accuracy* Máximo con los resultados de concordancia de dos médicos los resultados muestran una diferencia porcentual promedio $< 5\%$. La tercera evaluación médica presenta un nivel de concordancia muy superior a los otros dos. Es por esta razón que no es concluyente la superioridad de la evaluación clínica por sobre la de algoritmos ML.

La concordancia de las evaluaciones clínicas arroja un valor más alto para la zona Frontal (60% respecto de 45%, 55%, y 51% de las otras zonas). El *Accuracy* para los algoritmos ML presenta diferencias porcentuales $< 3\%$.

Los resultados obtenidos indican una dispersión promedio para algoritmos de clasificación (ML) mayor que el promedio de las evaluaciones clínicas. En el análisis de dispersión por zonas, la dispersión es menor en ML para las Zonas Frontal y Labio Superior y mucho mayor para las otras dos zonas.

Protocolos para la evaluación del grado de parálisis facial

Los cambios al protocolo en la consulta médica agregan un tiempo adicional menor a 10 minutos por paciente, con modificaciones mínimas respecto del original.

Más allá del beneficio para el proyecto, la alternativa de una grabación en el hogar resulta atractiva para el paciente ya que la parálisis facial dependiendo de su severidad puede tener un impacto no menor en una serie de áreas, incluyendo aspectos sociales. Así la grabación de un video en el hogar frente a tener que exponerse a salir de la casa es una alternativa para la que se podría evaluar su uso.

Sin embargo, la evaluación en base a un video o a través de una video conferencia requiere la revisión de otros aspectos fuera del alcance de este proyecto, en particular, en lo que se refiere a las posibles diferencias de una evaluación médica presencial (paciente 3D) versus un video del paciente (2D); en [10] se presenta un estudio inicial al respecto que compara diferentes escalas de graduación.

Limitantes asociadas al Tracking Facial y los Indicadores de Niveles de Simetría

El conjunto de indicadores de niveles de simetría utilizados como dato de entrada para el entrenamiento de los algoritmos (ML) no incluye un indicador asociado a la zona ocular, debido a que el tracking para esa zona es muy impreciso con el software de tracking utilizado, como se indica en [2]. La zona ocular es una de las 4 zonas graduadas al usar la escala HB por sectores, y ya que el modelo actual no incluye la medición de nivel de simetría asociado a la zona ocular, hay información que se está perdiendo y pudiese mejorar la exactitud del tracking. Se estima necesario contar en futuros trabajos, con un tracking adecuado para el sector.

En este trabajo no se realizó un análisis particular para sincinesia por estar fuera del alcance, sin embargo, se requiere analizar como se ven afectadas las mediciones de niveles de simetría; en [2] se describe la situación. La sincinesia se refiere a un movimiento involuntario en un sector provocado por un movimiento voluntario en otro.

6. Conclusiones

No es posible rechazar la hipótesis planteada en esta tesis. Tampoco es posible afirmar la superioridad de las evaluaciones clínicas por sobre los algoritmos de clasificación ML. La Concordancia de las evaluaciones clínicas (respecto de la evaluación promedio) es superior en promedio a la de los algoritmos ML en un 5,3%. Sin embargo, al calcular la diferencia en base a 2 de los 3 médicos evaluadores la diferencia promedio es de 2,2%. Para la dispersión se observa un desempeño de los algoritmos de clasificación (ML) superior en dos zonas e inferior en las otras dos zonas.

Los niveles de exactitud y concordancia tanto para las evaluaciones realizadas con algoritmos automáticos (ML) como para las evaluaciones médicas, aumentan al agrupar categorías adyacentes de a dos y de a tres. Así, en un análisis más grueso se obtiene una clasificación con mayor exactitud.

Entre las posibles razones limitantes al estudio se cuentan: 1) el no contar grabaciones ad-hoc que cumplan con los requisitos de calidad y contenido. 2) No disponer de indicadores para la zona ocular que alimenten los algoritmos (ML). 3) tamaño de la muestra y 4) número de evaluaciones. Los puntos 3) y 4) apuntan a la definición de un Gold Standard independiente de los datos usados para entrenamiento y testing.

Los resultados anteriores, aunque no muestran la superioridad de las evaluaciones basadas en tracking facial y algoritmos de clasificación ML, no indican que deban ser descartadas; así es que se estima de interés continuar investigando en este tipo de herramientas de evaluación.

El protocolo de grabación remota es factible de realizar; prueba de ello son los videos recibidos recientemente, y que no alcanzaron a ser incorporados a este estudio, en la figura 11 se muestran dos imágenes tomadas de uno de los videos.

La incorporación de un protocolo de grabación remota como alternativa a la evaluación presencial se estima debiese ser analizada por los beneficios que conlleva para el pacientes y médicos evaluadores tales como respuesta frente a restricciones de movilidad. Sin embargo, su aplicación debe considerar la validación de videos (evaluación 2D) respecto de evaluaciones presenciales (Evaluación 3D) [10].

La conformación de un repositorio de grabaciones y evaluaciones además de constituir la pieza clave para realizar este estudio, se visualiza de utilidad para otras investigaciones o desarrollos de software que requieran de casos de rostros que presenten asimetría al movimiento, ya que no se han encontrado bases de datos o repositorios con videos.

El modelo utilizado para la graduación del grado de parálisis facial permite realizar cambios de manera controlada. En esta tesis se reemplazó el algoritmo graduación basado en puntos de corte para los niveles de simetría de un video, por entrenamiento con algoritmos de clasificación (ML). Otro ejemplo de cambio sería incorporar a los indicadores de niveles de simetría uno para la zona ocular (previo ajuste al componente de tracking para el seguimiento correcto en esa zona).

Futuros trabajos basados en tracking facial deben considerar un seguimiento adecuado en la zona ocular y como resolver las limitantes anteriormente, y también dar respuesta a la sincinesia y su impacto en el tracking.

7. Referencias

- [1] House JW, Brackmann DE. Facial nerve grading system. *Otolaryngol Head Neck Surg*. 1985;93(2):146-147. doi:10.1177/019459988509300202
- [2] Gálvez RE. Cuantificación del grado de parálisis facial mediante algoritmos basados en tracking facial. 2017.
- [3] Luque-López MC, Guevara S, Buforn A. Parálisis Facial. Manual de urgencias y emergencias. Bloque ORL. medynet.
<http://www.medynet.com/usuarios/jraguilar/Manual%20de%20urgencias%20y%20Emergencias/parafaci.pdf> Published 2002.
- [4] Alanazi WL, El-Fetoh NMA, Alanazi SL, et al. Profile of facial palsy in Arar, northern Saudi Arabia. *Electron Physician*. 2017;9(10):5596-5602. Published 2017 Oct 25. doi:10.19082/5596
- [5] Ravikumar A, Singh P, Batish VK. FACIAL PALSY - TREATMENT OPTIONS. *Medical Journal Armed Forces India*. 1999;55(1):41-44. doi:10.1016/s0377-1237(17)30312-x
- [6] De Diego-Sastre JI, Prim-Espada MP, Fernández-García F. Epidemiología de la parálisis facial de Bell [The epidemiology of Bell's palsy]. *Rev Neurol*. 2005;41(5):287-290.
- [7] Barbosa J, Lee K, Lee S, et al. Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier. *BMC Med Imaging*. 2016;16:23. Published 2016 Mar 12. doi:10.1186/s12880-016-0117-0
- [8] Banks CA, Jowett N, Azizzadeh B, et al. Worldwide Testing of the eFACE Facial Nerve Clinician-Graded Scale. *Plast Reconstr Surg*. 2017;139(2):491e-498e. doi:10.1097/PRS.0000000000002954
- [9] Galindo L, Sandoval S, Cerda J, Ulloa M, Vargas A. Homologación lingüística al idioma español, en población chilena, de la herramienta Sunnybrook Facial Grading System para evaluar parálisis facial. Estudio piloto. *Rev. Otorrinolaringol. Cir. Cabeza Cuello* 2021; 81: 33-39
- [10] Tan JR, Coulson S, Keep M. Face-to-Face Versus Video Assessment of Facial Paralysis: Implications for Telemedicine. *J Med Internet Res*. 2019;21(4):e11109. Published 2019 Apr 12. doi:10.2196/11109

- [11] Reitzen SD, Babb JS, Lalwani AK. Significance and reliability of the House-Brackmann grading system for regional facial nerve function. *Otolaryngol Head Neck Surg.* 2009;140(2):154-158. doi:10.1016/j.otohns.2008.11.021
- [12] Yen TL, Driscoll CL, Lalwani AK. Significance of House-Brackmann facial nerve grading global score in the setting of differential facial nerve function. *Otol Neurotol.* 2003;24(1):118-122. doi:10.1097/00129492-200301000-00023
- [13] Fonseca KM, Mourão AM, Motta AR, Vicente LC. Scales of degree of facial paralysis: analysis of agreement. *Braz J Otorhinolaryngol.* 2015;81(3):288-293. doi:10.1016/j.bjorl.2014.04.005
- [14] Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. *Otolaryngol Head Neck Surg.* 2005;132(4):543-549. doi:10.1016/j.otohns.2005.01.027
- [15] Evans RA, Harries ML, Baguley DM, Moffat DA. Reliability of the House and Brackmann grading system for facial palsy. *J Laryngol Otol.* 1989;103(11):1045-1046. doi:10.1017/s002221510011093x

Anexo A: Videos de Internet

Los videos seleccionados se ubicaron en una cuenta YouTube utilizada para este propósito, quedando disponibles desde el cuestionario.

Nro Video	Enlace Youtube	Nro Video	Enlace Youtube
1	 https://youtu.be/WJvhwijipWM	16	 https://youtu.be/RJrg35k7qMY
2	 https://youtu.be/AtoweCckG8s	17	 https://youtu.be/6642tAGHwgQ
3	 https://youtu.be/l8o4SBlohLU	18	 https://youtu.be/agDZwgcxJFQ
4	 https://youtu.be/2hEwGJvIFno	19	 https://youtu.be/YGSje8tkp8c
5	 https://youtu.be/2i6DgrLbU50	20	 https://youtu.be/_Rb3cCdx_A4
6	 https://youtu.be/6ECK_8Wjlp4	21	 https://youtu.be/4ku33tZaUIY
7	 https://youtu.be/SOtyzEs4yx0	22	 https://youtu.be/pvQIBCBUoHM
8	 https://youtu.be/NZhiTfur7Ps	23	 https://youtu.be/PghpPtfbj5U
9	 https://youtu.be/D7YXB0DEfnQ	24	 https://youtu.be/uwPLstWbEq0
10	 https://youtu.be/FKbnR1FOuQY	25	 https://youtu.be/EjT-sHcEbn0
11	 https://youtu.be/uxkh2256hGw	26	 https://youtu.be/rFgAYYf111s
12	 https://youtu.be/FRsT9EITsyU	27	 https://youtu.be/FF2sGlcPecY
13	 https://youtu.be/RblyA7T_3Uw	28	 https://youtu.be/5_7xvkwS4Nk
14	 https://youtu.be/zrUaoFLnDso	29	 https://youtu.be/klqsBdoBj8U
15	 https://youtu.be/nAhM7gVIHml	30	 https://youtu.be/g2yN644UcBE

Tabla 8. Lista de videos y enlaces para visualización.

Anexo B: Gráficos de Distribución de Datos de Entrenamiento

Médico Dr1 Zona Frontal

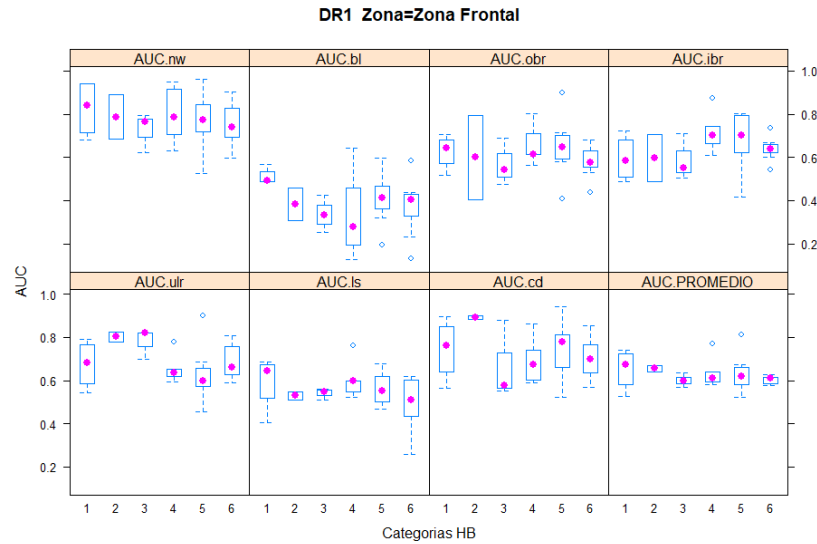
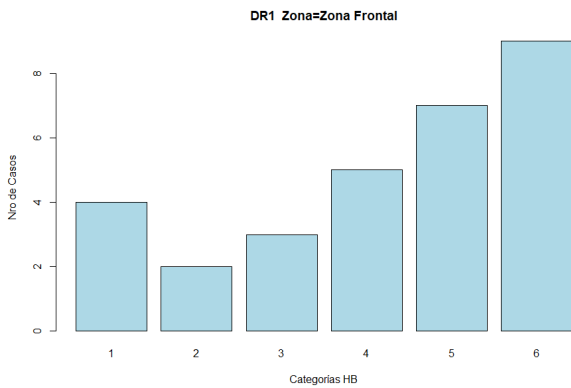


Figura 12. Distribución de casos por categoría HB

Figura 13. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Ocular

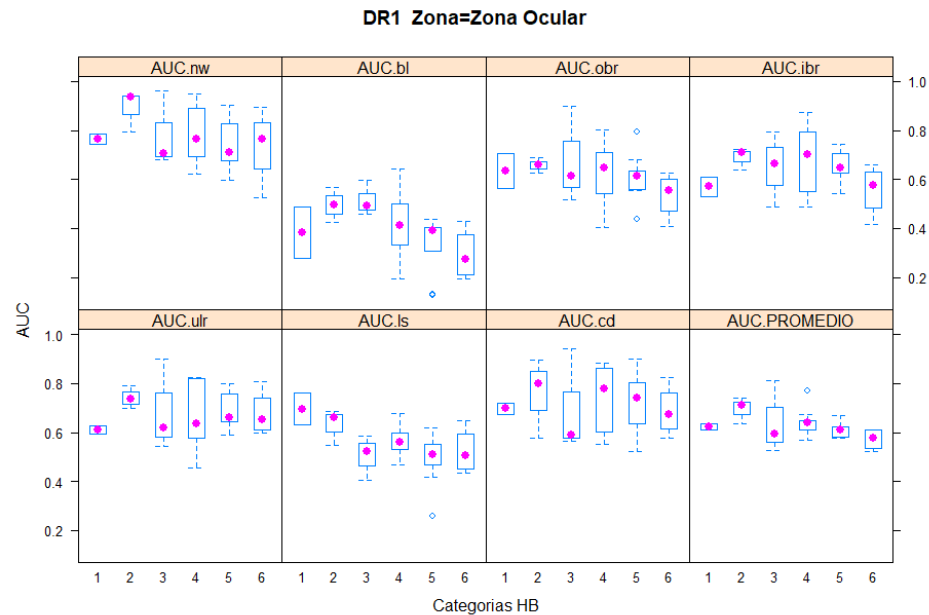
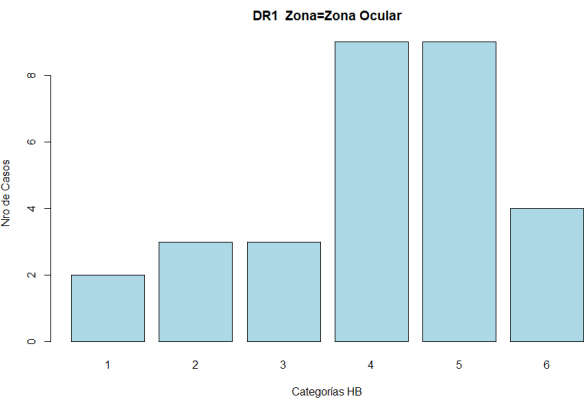


Figura 14. Distribución de casos por categoría HB

Figura 15. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Superior

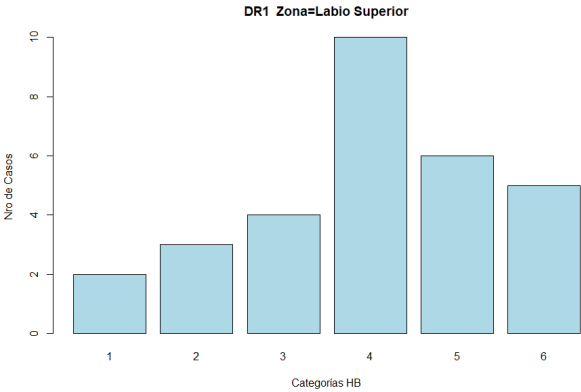


Figura 16. Distribución de casos por categoría HB

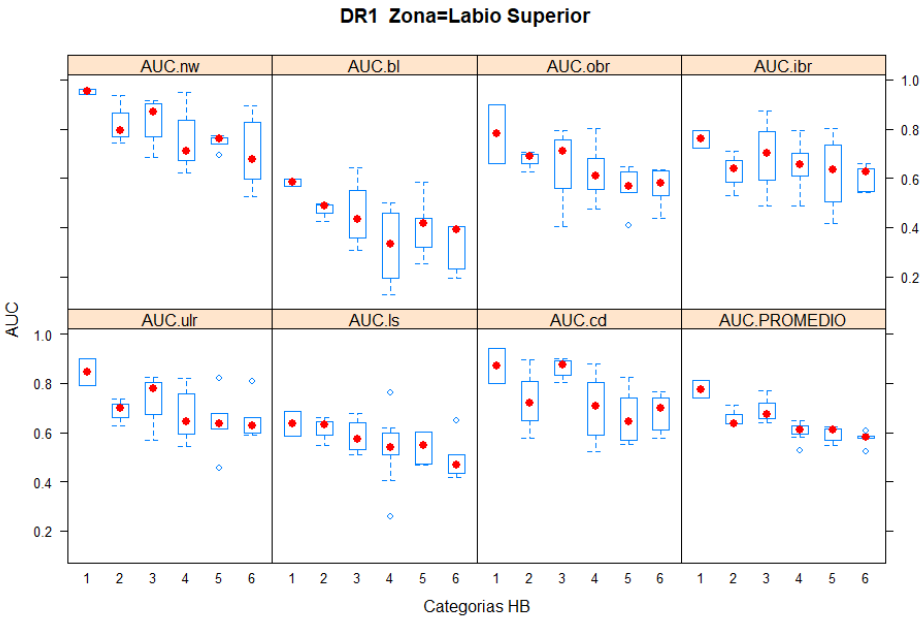


Figura 17. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Inferior

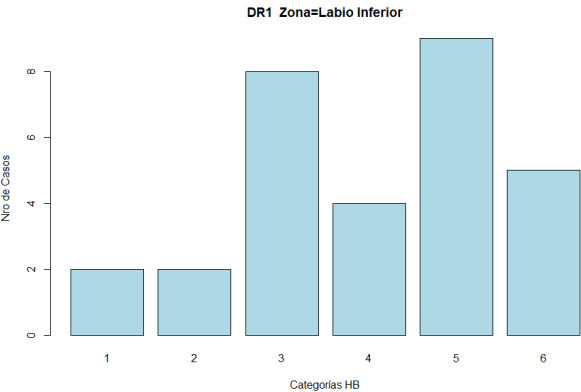


Figura 18. Distribución de casos por categoría HB

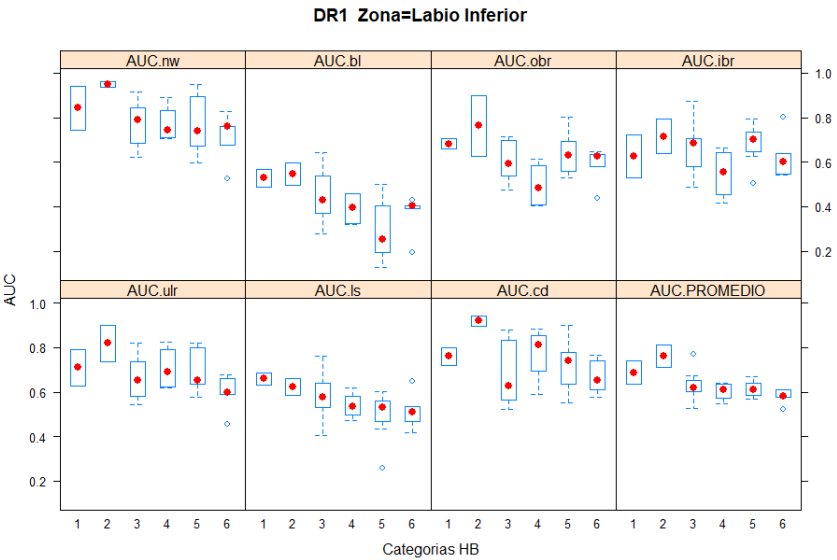


Figura 19. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Médico Dr2.
Zona Frontal

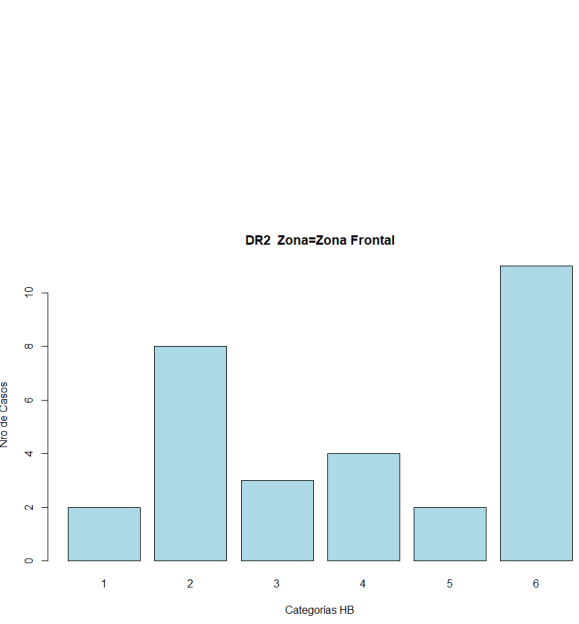


Figura 20. Distribución de casos por categoría HB

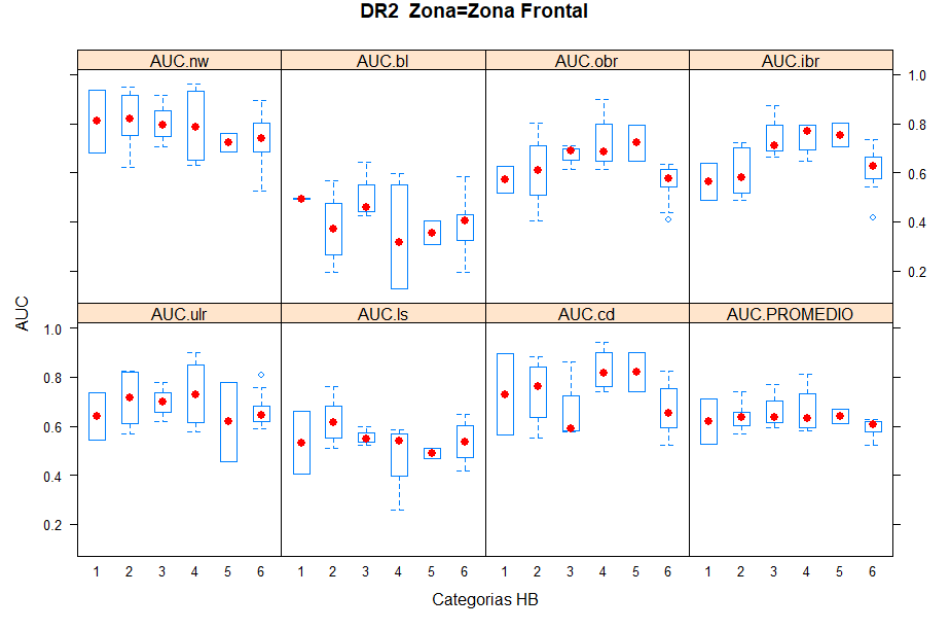


Figura 21. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Ocular

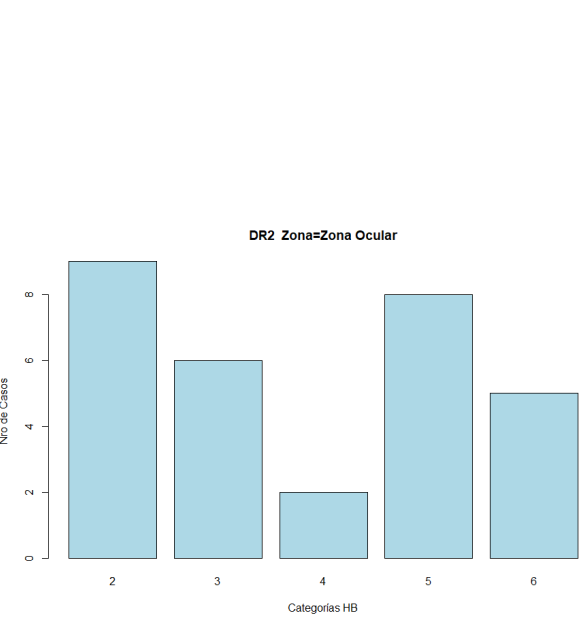


Figura 22. Distribución de casos por categoría HB

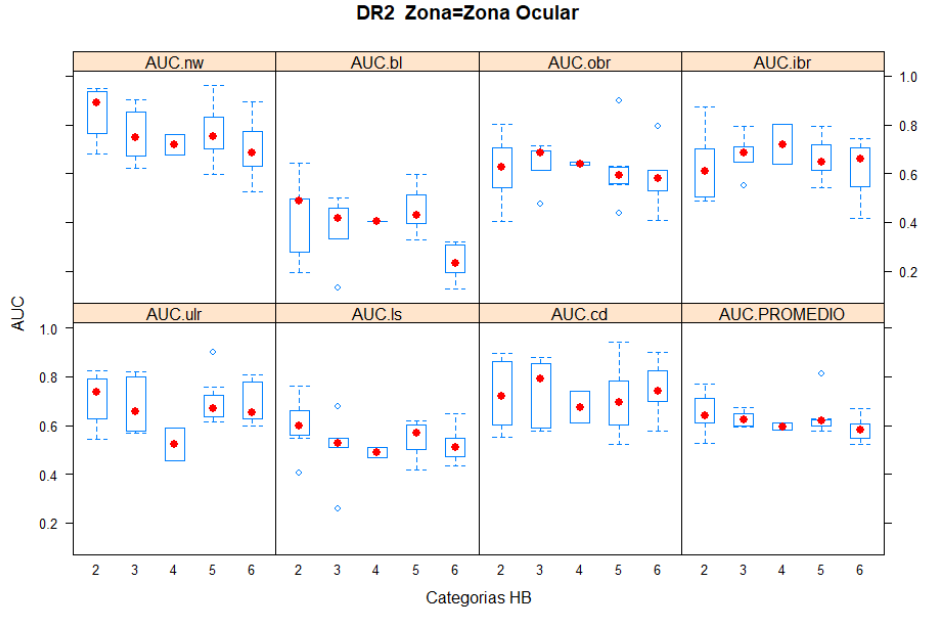


Figura 23. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Superior

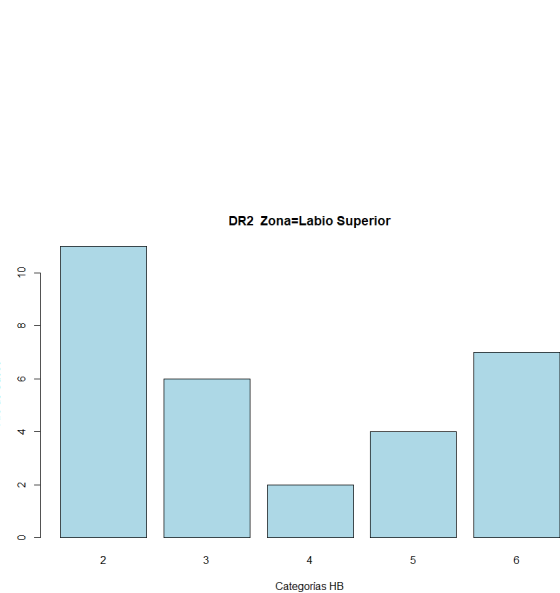


Figura 24. Distribución de casos por categoría HB

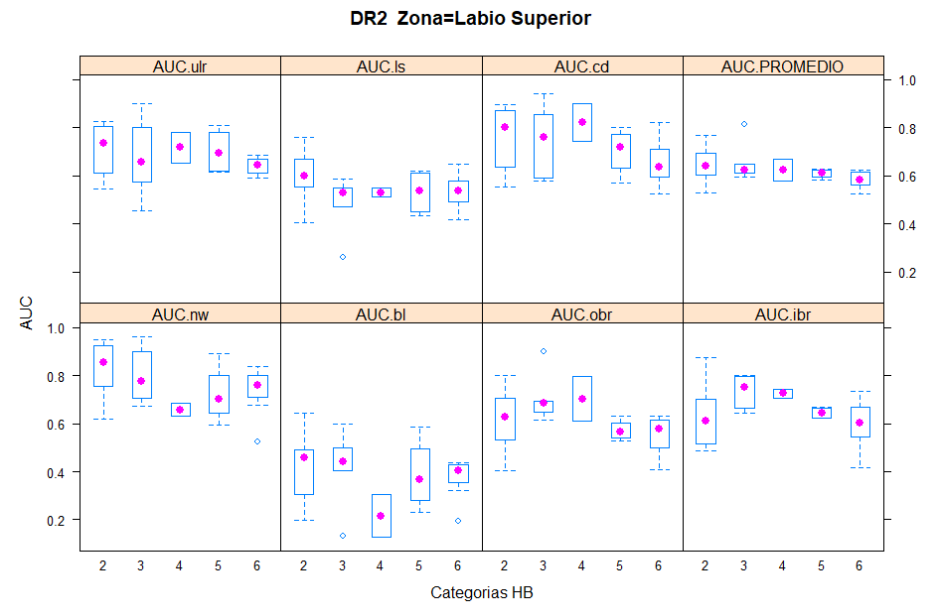


Figura 25. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Inferior

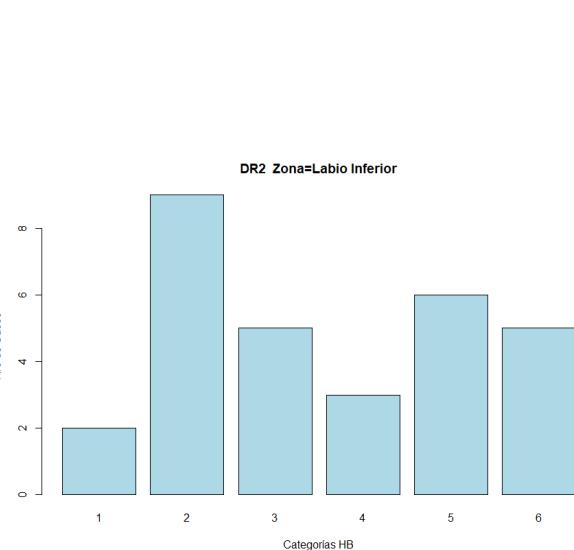


Figura 26. Distribución de casos por categoría HB

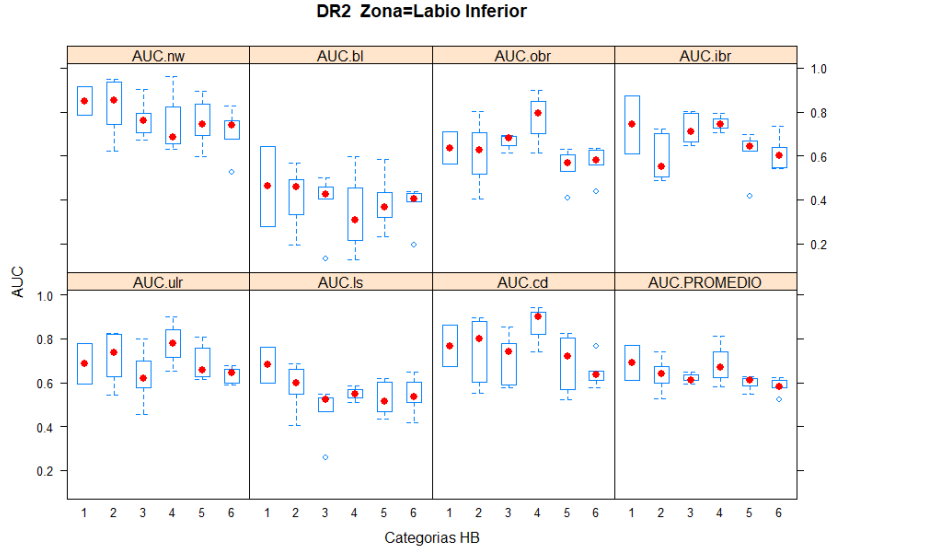


Figura 27. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Médico Dr3.
Zona Frontal

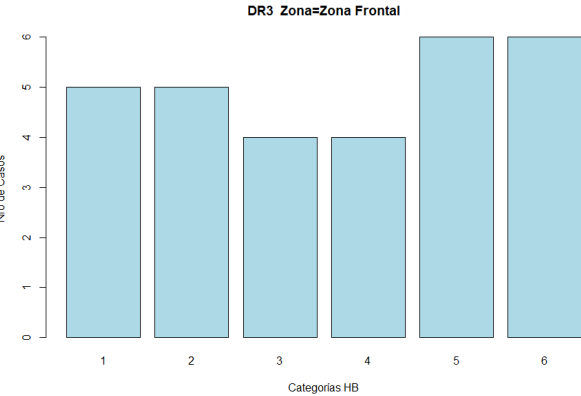


Figura 28. Distribución de casos por categoría HB

ZonaOcular

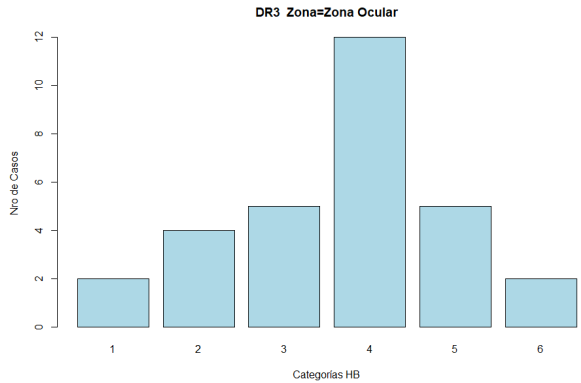


Figura 30. Distribución de casos por categoría HB

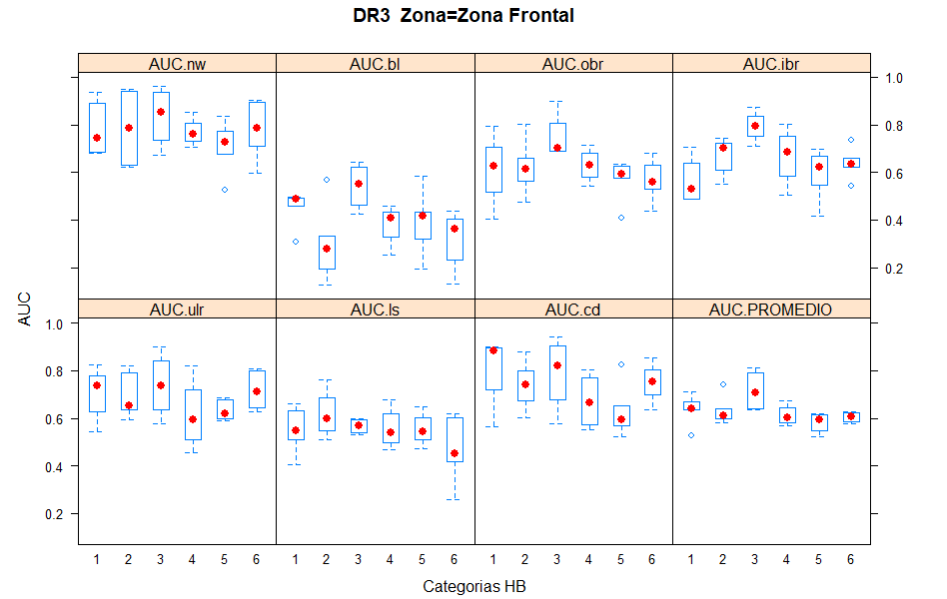


Figura 29. Distribución de niveles de simetría (AUC) por categoría HB por expresión

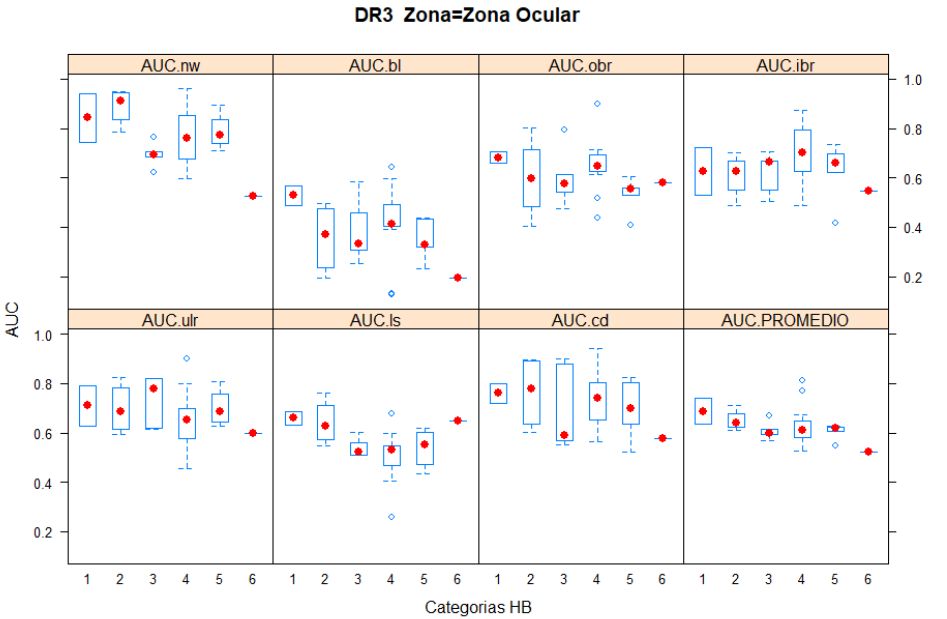


Figura 31. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Superior

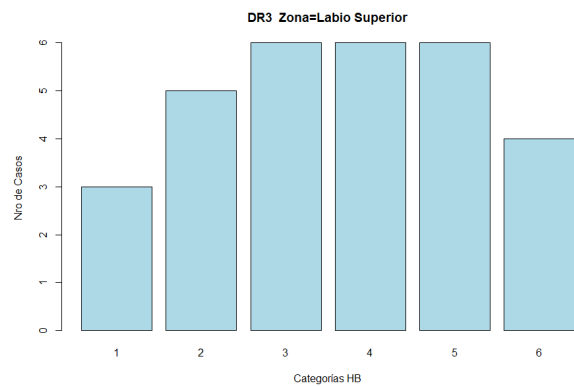


Figura 32. Distribución de casos por categoría HB

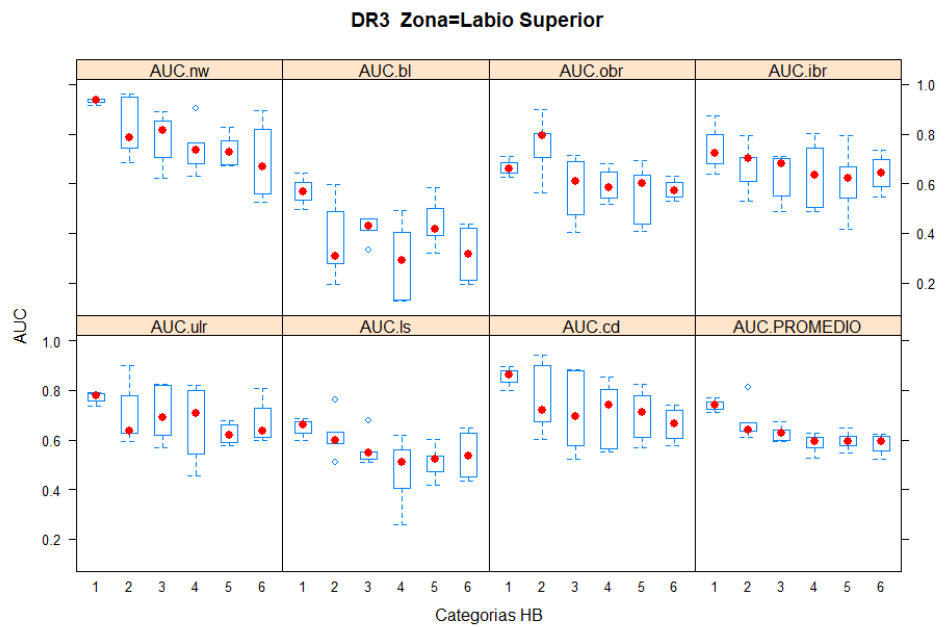


Figura 33. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Zona Labio Inferior

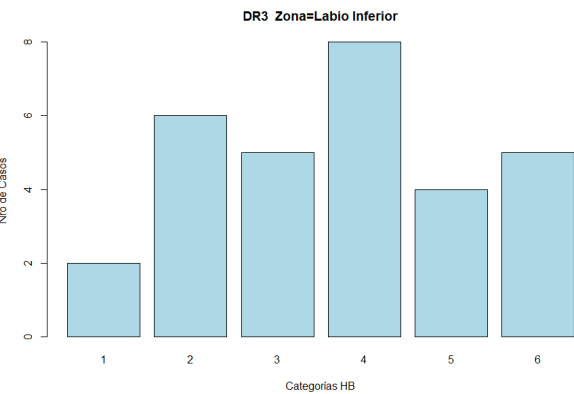


Figura 34. Distribución de casos por categoría HB

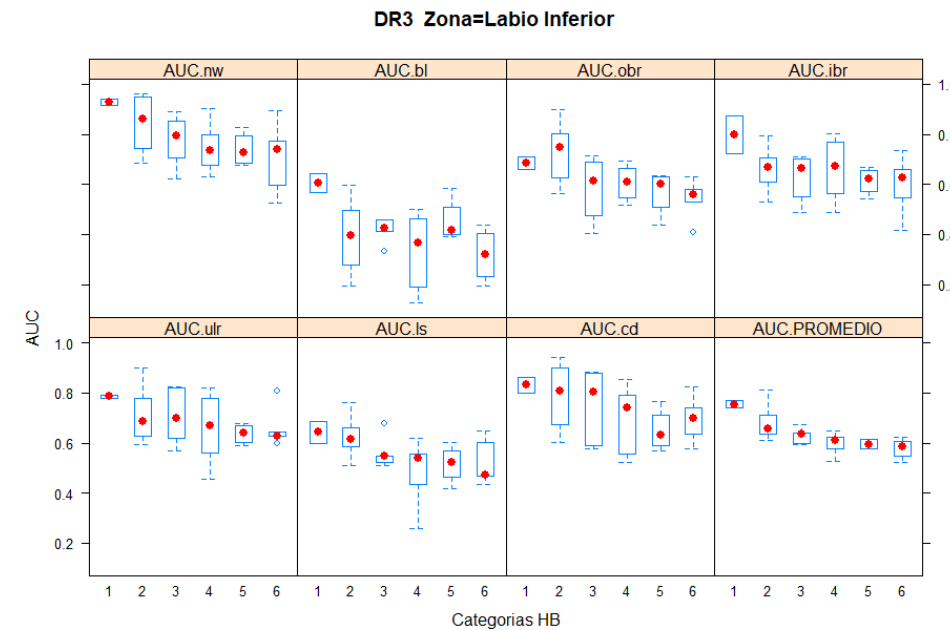


Figura 35. Distribución de niveles de simetría (AUC) por categoría HB por expresión

Anexo C: Resultados de Entrenamientos con Algoritmos de Clasificación (ML) por Médico.

En el entrenamiento con las evaluaciones de los médicos Dr1, Dr2, Dr3, y promedio de los tres, se utilizó una proporción de los datos para probar (testing) y la diferencia para entrenar. En la columna pTrain se indica el porcentaje utilizado para entrenamiento.

Entrenamiento ML Médico Dr1

Médico	Nro Categorías HB	Zona	pTrain	k-fold Resamples	Ida	Cart	KNN	SVM	RF	Nbayes	MAX Accuracy
DR1	6	Frontal	0,65	9	0,30	0,30	0,30	0,30	0,40	-	0,40
	3	Frontal	0,7	9	0,14	0,43	0,43	0,57	0,29	0,57	0,57
	2	Frontal	0,75	9	0,67	0,67	0,67	0,83	0,83	0,17	0,83
	6	Ocular	0,65	9	0,20	0,30	0,40	0,30	0,30	-	0,40
	3	Ocular	0,7	9	0,43	0,43	0,29	0,43	0,29	0,29	0,43
	2	Ocular	0,75	9	0,60	0,60	0,80	0,80	0,60	0,40	0,80
	6	Lsuperior	0,65	9	0,27	0,36	0,09	0,36	0,18	-	0,36
	3	Lsuperior	0,7	9	0,33	0,00	0,17	0,50	0,17	0,33	0,50
	2	Lsuperior	0,75	9	0,83	0,67	0,83	0,83	0,83	0,83	0,83
	6	Linferior	0,65	9	0,20	0,20	0,50	0,30	0,30	-	0,50
	3	Linferior	0,7	9	0,71	0,29	0,43	0,43	0,57	0,57	0,71
	2	Linferior	0,75	9	0,50	0,67	0,83	0,67	0,67	0,50	0,83

Tabla 8. Valores de Exactitud (Accuracy) para los diferentes algoritmos de clasificación.

Entrenamiento ML Médico Dr2

Médico	Nro Categorías HB	Zona	pTrain	k-fold Resamples	Ida	Cart	KNN	SVM	RF	Nbayes	MAX Accuracy
DR2	6	Frontal	0,65	9	0,27	0,27	0,27	0,27	0,36	0,27	0,36
	3	Frontal	0,7	9	0,57	0,43	0,43	0,43	0,71	0,43	0,71
	2	Frontal	0,75	9	0,67	0,33	0,67	0,50	0,67	0,33	0,67
	6	Ocular	0,65	9	0,33	0,25	0,17	0,17	0,25	0,17	0,33
	3	Ocular	0,7	9	0,71	0,43	0,43	0,43	0,71	0,14	0,71
	2	Ocular	0,75	9	0,67	0,50	0,67	0,67	0,50	0,67	0,67
	6	Lsuperior	0,65	9	0,40	0,40	0,50	0,30	0,40	-	0,50
	3	Lsuperior	0,7	9	0,83	0,33	0,50	0,67	0,83	0,83	0,83
	2	Lsuperior	0,75	9	0,50	0,67	0,67	0,67	0,50	0,83	0,83
	6	Linferior	0,65	9	0,55	0,27	0,18	0,27	0,36	-	0,55
	3	Linferior	0,7	9	0,50	0,33	0,33	0,33	0,50	0,33	0,50
	2	Linferior	0,75	9	0,50	0,67	0,67	0,67	0,67	0,83	0,83

Tabla 9. Valores de Exactitud (Accuracy) para los diferentes algoritmos de clasificación

Entrenamiento ML Médico Dr3

Médico	Nro Categorías HB	Zona	pTrain	k-fold Resamples	Ida	Cart	KNN	SVM	RF	Nbayes	MAX Accuracy
DR3	6	Frontal	0,65	9	0,09	0,09	0,27	0,18	0,27	0,18	0,27
	3	Frontal	0,7	9	0,14	0,00	0,43	0,43	0,14	0,43	0,43
	2	Frontal	0,75	9	0,67	0,33	0,67	0,50	0,50	0,67	0,67
	6	Ocular	0,65	9	0,45	0,45	0,45	0,45	0,36	0,36	0,45
	3	Ocular	0,7	9	0,57	0,43	0,57	0,57	0,57	0,57	0,57
	2	Ocular	0,75	9	1,00	0,50	0,83	0,83	0,83	0,50	1,00
	6	Lsuperior	0,65	9	0,40	0,10	0,20	0,40	0,50	0,30	0,50
	3	Lsuperior	0,7	9	0,29	0,43	0,29	0,43	0,57	0,43	0,57
	2	Lsuperior	0,75	9	0,83	0,83	0,83	0,83	0,83	1,00	1,00
	6	Linferior	0,65	9	0,09	0,27	0,09	0,27	0,18	-	0,27
	3	Linferior	0,7	9	0,43	0,29	0,00	0,29	0,14	0,29	0,43
	2	Linferior	0,75	9	0,83	0,67	0,83	0,83	1,00	1,00	1,00

Tabla 10. Valores de Exactitud (Accuracy) para los diferentes algoritmos de clasificación

Entrenamiento ML con Promedio Evaluaciones de los Médicos Dr1, Dr2, y Dr3 (Dr123)

Médico	Nro Categorías HB	Zona	pTrain	k-fold Resamples	Ida	Cart	KNN	SVM	RF	Nbayes	MAX Accuracy	AVERAGE Accuracy
DR123	6	Frontal	0,65	9	42,9%	27,3%	42,9%	42,9%	36,4%	42,9%	42,9%	39,2%
	3	Frontal	0,80	9	42,9%	42,9%	42,9%	40,0%	80,0%	80,0%	80,0%	54,8%
	2	Frontal	0,80	9	50,0%	66,7%	66,7%	66,7%	66,7%	50,0%	66,7%	61,1%
DR123	6	Ocular	0,65	9	44,4%	45,5%	44,4%	45,5%	33,3%	33,3%	45,5%	41,1%
	3	Ocular	0,80	9	60,0%	60,0%	60,0%	60,0%	60,0%	60,0%	60,0%	60,0%
	2	Ocular	0,80	9	100,0%	83,3%	83,3%	83,3%	83,3%	50,0%	100,0%	80,5%
DR123	6	Lsuperior	0,65	9	22,2%	30,0%	33,3%	30,0%	11,1%	44,4%	44,4%	28,5%
	3	Lsuperior	0,80	9	40,0%	60,0%	42,9%	60,0%	60,0%	60,0%	60,0%	53,8%
	2	Lsuperior	8,00	9	83,3%	66,7%	83,3%	83,3%	83,3%	83,3%	83,3%	80,6%
DR123	6	Linferior	0,65	9	27,3%	27,3%	44,4%	44,4%	33,3%	11,1%	44,4%	31,3%
	3	Linferior	0,80	9	40,0%	60,0%	80,0%	60,0%	60,0%	80,0%	80,0%	63,3%
	2	Linferior	0,80	9	83,3%	66,7%	66,7%	66,7%	83,3%	83,3%	83,3%	75,0%

Tabla 11. Valores de Exactitud (Accuracy) para los diferentes algoritmos de clasificación

Anexo D: Formulario de Consentimiento

QA3C, VERSIÓN 23.02.16

 HOSPITAL CLÍNICO UNIVERSIDAD DE CHILE

DOCUMENTO DE CONSENTIMIENTO INFORMADO

Evaluación automática del grado de parálisis facial usando videos.

Nombre del Investigador Institución:	Mauricio Cerda/Daniel Rappoport
Teléfono(s):	89641/88335
Servicio o Departamento:	ICBM/Cirugía Norte
Información del patrocinador (si aplica):	No aplica.

Invitación a participar: Le estamos invitando a participar en el proyecto de investigación "Validación de Algoritmo para la Cuantificación del Grado de Parálisis Facial y Medición de la Progresión de Parálisis Facial.", para evaluar una herramienta/metodología basada en algoritmos que busca cuantificar de manera más objetiva el grado de parálisis facial, considerando su aplicación durante la evolución/progresión de la patología en el tiempo.
Antes de tomar la decisión de participar lea atentamente este documento.

Introducción: En el caso de la parálisis facial unilateral, para cuantificar el grado de parálisis el facultativo evalúa la pérdida de movimiento en base a una serie de gestos que se solicita realizar al paciente. Se ha desarrollado una herramienta, la cual a partir de la grabación de los gestos que se solicita al paciente realizar, entrega el grado de parálisis correspondiente. Este estudio busca determinar si esta herramienta es adecuada y fácil de usar para la evaluación.

Objetivos: Esta investigación busca determinar si la evaluación realizada por el método basado en algoritmos es al menos igual de certera que aquella realizada por el médico (evaluación clínica). Se considera realizar evaluaciones periódicas de los pacientes post cirugía. La muestra contempla evaluar a 24 pacientes, atendidos en el Hospital Clínico de la Universidad de Chile, en un período aproximado de 6 meses.

Procedimientos: Si Ud. acepta participar, la autorización entregada a través de este documento permitirá grabar la evaluación del grado de parálisis que realiza el médico tratante. La evaluación consiste en que el médico solicita al paciente realizar una serie de gestos, siendo estos gestos el objetivo de la grabación. **Esta actividad no introduce ningún cambio en la técnica quirúrgica como tampoco en el tratamiento ofrecido por su médico tratante, tanto en la cirugía como en el seguimiento post operatorio.** Las grabaciones se realizarán en el hogar con un teléfono celular. Pudiendo grabarse uno mismo u otra persona.

Riesgos: La grabación de un video no reviste riesgos.

Costos: No hay costos asociados, sin embargo, el paciente requiere disponer de un celular que permita realizar la grabación.

Versión _____. Hospital Clínico Universidad de Chile, Fecha _____

Página 1 de 3

Como participante en este proyecto Ud. o su sistema previsional deberán financiar las hospitalizaciones, honorarios, exámenes y tratamientos habituales para el tratamiento de su enfermedad.

Beneficios: Aporte al estudio en cuestión.

Alternativas: Si Ud. decide no participar en esta investigación se mantiene el tratamiento habitual, al igual que en el caso en que decida participar.

Compensación: Ud. no recibirá ninguna compensación económica por su participación en el estudio.

Confidencialidad: Toda la información derivada de su participación en este estudio será conservada en forma de estricta confidencialidad, lo que incluye el acceso de los investigadores o agencias supervisoras de la investigación. Cualquier publicación o comunicación científica de los resultados de la investigación será completamente anónima. Las imágenes (fotografías) que se publiquen serán adecuadamente modificadas para garantizar el anonimato. La modificación de la imagen consiste en tapar/cubrir la zona de los ojos.

Información adicional: (garantía de acceso a toda información nueva relevante) Ud. o su médico tratante serán informados si durante el desarrollo de este estudio surgen nuevos conocimientos o complicaciones que puedan afectar su voluntad de continuar participando en la investigación.

Voluntariedad: Su participación en esta investigación es totalmente voluntaria y se puede retirar en cualquier momento comunicándolo al investigador y a su médico tratante, sin que ello signifique modificaciones en el estudio y tratamiento habituales de su enfermedad.

Complicaciones: La participación en este estudio no incorpora medicamentos o acciones que modifiquen en el tratamiento actual, no existiendo elementos que puedan originar complicaciones producto de la participación en este proyecto.

Derechos del participante: Usted recibirá una copia de este documento firmado. Si usted requiere cualquier otra información sobre su participación en este estudio puede comunicarse con:

Investigador: Nombres y teléfono (teléfono de contacto las 24 hrs si corresponde a un estudio internacional)

Mauricio Cerda, +569 85927543.

Otros Derechos del participante

En caso de duda sobre sus derechos comunicarse con el Comité Ético Científico o de Investigación del Hospital Clínico Universidad de Chile, Teléfono: 229789008, Email:

comiteetica@hcuch.cl, ubicado en Santos Dumont N° 999, 4 Piso Sector D, Comuna de Independencia, Santiago.

Conclusión:

Después de haber recibido y comprendido la información de este documento y de haber podido aclarar todas mis dudas, otorgo mi consentimiento para participar en el proyecto "Validación de Algoritmo para la Cuantificación del Grado de Parálisis Facial y Medición de la Progresión de Parálisis Facial."

Nombre del sujeto
Run.

Firma

Fecha y Hora

Mauricio Cerda
Run. 15.260.934-5

Firma

Fecha y Hora

Nombre del Director o Delegado
Run.

Firma

Fecha y Hora

Si se trata de un sujeto iletrado, no vidente etc, registrar nombre del sujeto y de su apoderado (Testigo).

Nombre del Testigo
Run.

Firma

Fecha y Hora

Anexo E: Datos Estadísticos Complementarios

- I. Concordancia respecto de la **Evaluación Promedio** de un cuarto médico evaluador. Los datos de su evaluación no fueron utilizados en la construcción de la **Evaluación Promedio**. Los resultados obtenidos son los siguientes:

Zona	Nro Categorías HB	Agreement %	Kappa de Cohen
Frontal	6	37,90	0,23
	3	69,00	0,50
	2	82,00	0,64
Ocular	6	41,40	0,28
	3	58,60	0,37
	2	82,80	0,61
Labio Superior	6	48,30	0,36
	3	75,90	0,61
	2	86,20	0,72
Labio Inferior	6	44,80	0,30
	3	72,40	0,53
	2	79,30	0,57
Promedios		64,88	0,48
6 categorías		43,10	
3 categorías		68,98	
2 categorías		82,58	

Tabla 12. Concordancia y Confiabilidad entre Dr4 y la Evaluación Promedio.

II. Concordancia inter-observadores n = 3.

Zona	Nro Categorías HB	Agreement DRs 1-2-3 %	Kappa de Fleiss
Frontal	6	26,70	0,33
	3	63,30	0,60
	2	66,70	0,54
Ocular	6	23,30	0,24
	3	50,00	0,48
	2	56,70	0,39
Labio Superior	6	13,30	0,19
	3	50,00	0,48
	2	63,30	0,51
Labio Inferior	6	10,00	0,12
	3	40,00	0,36
	2	60,00	0,46
promedios		43,61	0,40
6 categorías		18,33	0,18
3 categorías		50,83	0,48
2 categorías		61,68	0,47

Tabla 13. Concordancia y Confiabilidad para grupos de 3 evaluadores.

Las Concordancias entre grupos de médicos en promedio son similares; en el análisis por zona, la zona frontal presenta mayor concordancia para todas las agrupaciones de categorías.