



UNIVERSIDAD DE GUADALAJARA

VICERRECTORÍA EJECUTIVA
SISTEMA UNIVERSITARIO DE BIBLIOTECAS
Biblioteca Digital

La presente tesis es publicada a texto completo en virtud de que el autor ha dado su autorización por escrito para la incorporación del documento a la Biblioteca Digital y al Repositorio Institucional de la Universidad de Guadalajara, esto sin sufrir menoscabo sobre sus derechos como autor de la obra y los usos que posteriormente quiera darle a la misma.

UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS
DIVISIÓN DE TECNOLOGÍAS PARA LA INTEGRACIÓN CIBER-HUMANA



DETECCIÓN DE ESCLEROSIS LATERAL AMIOTRÓFICA MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO EN DATOS BIOMÉTRICOS

TESIS PARA OPTAR POR EL GRADO DE
MAESTRA EN CIENCIAS EN BIOINGENIERÍA Y CÓMPUTO INTELIGENTE

PRESENTA
ING. DANIELA SUÁREZ HERNÁNDEZ

DIRECTOR DE TESIS
DR. ISRAEL ROMÁN GODÍNEZ

CODIRECTOR
DR. STEWART RENÉ SANTOS ARCE

GUADALAJARA, JALISCO. MARZO DE 2025

| | |
|---|---------------|
| DEDICATORIA | I |
| AGRADECIMIENTOS | II |
| ABSTRACT | III |
| RESUMEN | IV |
| CAPÍTULO 1. INTRODUCCIÓN | - 1 - |
| 1.1. INTRODUCCIÓN | - 1 - |
| 1.2. ANTECEDENTES | - 2 - |
| 1.3. PLANTEAMIENTO DEL PROBLEMA | - 4 - |
| 1.4. JUSTIFICACIÓN | - 5 - |
| 1.5. HIPÓTESIS | - 5 - |
| 1.6. OBJETIVOS | - 5 - |
| 1.6.1. OBJETIVO GENERAL | - 5 - |
| 1.6.2. OBJETIVOS ESPECÍFICOS | - 5 - |
| CAPÍTULO 2. MARCO TEÓRICO | - 7 - |
| 2.1. ESCLEROSIS LATERAL AMIOTRÓFICA (ELA) | - 7 - |
| 2.1.1. CARACTERÍSTICAS DE ELA | - 7 - |
| 2.1.2. DIAGNÓSTICO DE ELA | - 7 - |
| 2.2. MARCADORES FACIALES PARA EL ANÁLISIS DEL MOVIMIENTO | - 9 - |
| 2.2.1. MARCADORES FACIALES | - 9 - |
| 2.2.2. TRANSFORMACIÓN DE COORDENADAS PARA MEDICIÓN ABSOLUTA | - 9 - |
| 2.2.3. ANÁLISIS DE LAS SIMILITUDES ENTRE MARCADORES FACIALES | - 10 - |
| 2.2.4. ANÁLISIS DEL MOVIMIENTO MEDIANTE VISIÓN COMPUTACIONAL | - 12 - |
| 2.3. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO | - 13 - |
| 2.3.1. ALGORITMOS DE AA PARA CLASIFICACIÓN | - 15 - |
| 2.3.2. MÉTRICAS DE EVALUACIÓN DEL DESEMPEÑO DE LOS MODELOS | - 16 - |
| CAPÍTULO 3. METODOLOGÍA | - 17 - |
| 3.1. EXPLORACIÓN DE BASES DE DATOS NEUROFACIALES | 19 |
| 3.1.1. BASE DE DATOS DE ELA | 19 |
| 3.1.2. DATOS AUDITIVOS | 19 |

| | |
|--|-----------|
| 3.1.3. DATOS FACIALES | 21 |
| 3.2. PREPROCESAMIENTO DE LOS DATOS | 21 |
| 3.2.1. EXTRACCIÓN DE MARCADORES FACIALES | 22 |
| 3.2.2. TRANSFORMACIÓN ESPACIAL DE COORDENADAS CARTESIANAS A ESFÉRICAS | 25 |
| 3.2.3. FILTRADO DE MARCADORES FACIALES | 25 |
| 3.3. DEFINICIÓN DE LAS CARACTERÍSTICAS FACIALES RELEVANTES | 27 |
| 3.4. MODELADO CON ALGORITMOS DE APRENDIZAJE AUTOMÁTICO | 28 |
| 3.4.1. ALGORITMOS DE APRENDIZAJE AUTOMÁTICO IMPLEMENTADOS | 29 |
| 3.4.2. ENTRENAMIENTO Y VALIDACIÓN | 30 |
| CAPÍTULO 4. RESULTADOS Y DISCUSIÓN | 31 |
| 4.1. COORDENADAS EN EL ESPACIO-TIEMPO | 31 |
| 4.2. MATRICES DE CORRELACIÓN DEL COMPORTAMIENTO DE LOS MARCADORES FACIALES | 34 |
| 4.3. HIPERPARÁMETROS RESULTANTES DEL <i>GRIDSEARCH</i> | 36 |
| 4.4. IMPLEMENTACIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO | 37 |
| 4.4.1. COMPARACIÓN DE ALGORITMOS DE CLASIFICACIÓN | 37 |
| 4.4.2. RENDIMIENTO DE LAS COORDENADAS ESFÉRICAS | 40 |
| 4.5. DISCUSIÓN | 41 |
| CAPÍTULO 5. CONCLUSIONES | 42 |
| 5.1. LIMITACIONES | 42 |
| 5.2. TRABAJO FUTURO | 43 |
| 5.3. PRODUCTOS DERIVADOS | 43 |
| 6. ANEXOS | 44 |
| REFERENCIAS BIBLIOGRÁFICAS | 47 |

Dedicatoria

A mi tía Emma Hernández, quien con fortaleza enfrentó la enfermedad que fue el centro de esta investigación.

Este trabajo es para usted, con todo mi amor, con la esperanza de que su lucha no haya sido en vano. Que cada hallazgo, cada esfuerzo y cada paso en este camino contribuyan a un futuro en el que quienes enfrentan este diagnóstico tengan más respuestas y más esperanza.

Su historia vive en cada palabra de esta tesis; este es mi pequeño homenaje a su memoria, con la certeza de que su luz sigue guiando mi camino.

Agradecimientos

Este trabajo no hubiera sido posible sin el apoyo, guía y cariño de muchas personas que han sido fundamentales en este camino. A cada una de ellas, gracias por estar.

A mis padres, Alicia y Baltazar, por su amor incondicional, su apoyo y su confianza en mí. Gracias por ser mi más grande inspiración. Cada logro en mi vida es también suyo.

Al Dr. Israel Román, mi director, por su guía constante, su paciencia y su confianza en mi trabajo; al Dr. Stewart Santos, mi co-director, por su apoyo, orientación y su tiempo; a la Dra. Sulema Torres, por ser no solo una asesora excepcional, sino también una consejera cuya orientación ha sido importante en mi; al Dr. Ricardo Salido, por su asesoría y disposición para enriquecer este proyecto; y al Dr. Roberto Carrasco, por su valioso tiempo y contribuciones como sinodal. A cada uno de ustedes, gracias por compartir su conocimiento y por haberme acompañado en este proceso.

A Daniel, por estar presente, su amor, paciencia y apoyo. Por celebrar mis logros como suyos y acompañarme en cada paso, incluso en mis momentos más estresantes. Gracias por ser.

A Liz, mi hermana del alma, por su amor y por estar siempre a mi lado, sin importar la distancia o el tiempo. A Uri, un amigo que ha estado en los momentos más difíciles dándome apoyo y escuchándome siempre. A Fer Suárez, por ser mi apoyo constante y mi amiga.

A Fer y Razo, mis amigas de la maestría y la carrera, con quienes compartí aprendizajes, risas y momentos inolvidables.

A mis compañeros de LASAID, por siempre compartir conmigo y por hacer que este proceso fuera más divertido.

A CONAHCYT, por el apoyo brindado, que hizo posible la realización de este proyecto.

A todos ustedes, gracias por su tiempo, su confianza y su cariño. Este logro también es suyo.

Abstract

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that affects motor neurons, leading to progressive paralysis that impacts mobility, communication, and swallowing. Its diagnosis is often delayed due to the absence of specific biomarkers and the variability of symptoms, which postpones the initiation of palliative treatments.

This study proposes the development of a machine learning (ML) model based on facial biometrics for the early detection of ALS. Using the Toronto NeuroFace dataset, videos of ALS patients and healthy controls are analyzed, focusing on the repetition task of the syllable sequence /pataka/. The MediaPipe® library is employed to extract 54 facial landmarks, which are then transformed into spherical coordinates (r, θ, φ) to minimize spatial errors and improve analysis stability. Through Pearson correlation matrices, the set of facial markers is reduced to 24 key points, optimizing the model without significant loss of relevant information.

For data classification, k-Nearest Neighbors (kNN), Decision Trees (DT), Random Forest (RF), and Artificial Neural Networks (ANN) algorithms were implemented. Hyperparameter optimization was conducted using GridSearch, and stratified cross-validation was applied to evaluate model performance in terms of accuracy, sensitivity, and specificity. The results indicate that the r and θ coordinates provide the best classification performance, with the ANN model achieving 72.2% accuracy and 88.9% sensitivity in detecting ALS. In contrast, the φ coordinate showed lower classification utility, possibly due to limitations in estimating the depth of facial movements.

This study demonstrates the potential of using ML and facial biometrics to complement the clinical diagnosis of ALS, particularly in environments with limited access to neurology specialists. Future studies are recommended to integrate auditory data and use advanced models such as Graph Neural Networks to enhance system accuracy. The implementation of these tools in medical practice could facilitate faster and more accessible ALS detection, optimizing diagnosis and treatment timelines.

Resumen

La esclerosis lateral amiotrófica (ELA) es una enfermedad neurodegenerativa que afecta las neuronas motoras, causando una parálisis progresiva que impacta la movilidad, la comunicación y la deglución. Su diagnóstico es tardío debido a la ausencia de biomarcadores específicos y a la variabilidad de los síntomas, lo que retrasa el inicio de los tratamientos paliativos.

Este estudio propone el desarrollo de un modelo de aprendizaje automático (AA) basado en biometría facial para la detección temprana de la ELA. Utilizando el conjunto de datos Toronto NeuroFace, se analizan videos de pacientes con ELA y controles sanos, enfocándose en la tarea de repetición de la secuencia de sílabas /pataka/. Se emplea la biblioteca MediaPipe® para extraer 54 marcadores faciales, los cuales se transforman a coordenadas esféricas (r, θ, φ) para minimizar errores espaciales y mejorar la estabilidad del análisis. A través de matrices de correlación de Pearson, se reduce el conjunto de marcadores faciales a 24 puntos clave, optimizando el modelo sin pérdida significativa de información relevante.

Para la clasificación de los datos, se implementaron los algoritmos k-Nearest Neighbors (kNN), árboles de decisión (DT), Random Forest (RF) y redes neuronales artificiales (RNA). Se llevó a cabo una optimización de hiperparámetros mediante GridSearch y una validación cruzada estratificada para evaluar el desempeño de los modelos en términos de precisión, sensibilidad y especificidad. Los resultados muestran que las coordenadas r y θ ofrecen los mejores desempeños en la clasificación, destacando el modelo de RNA con una precisión del 72.2% y una sensibilidad del 88.9% en la detección de ELA. En contraste, la coordenada φ presentó menor utilidad en la clasificación, posiblemente debido a las limitaciones en la estimación de la profundidad de los movimientos faciales.

El presente trabajo demuestra el potencial del uso de AA y biometría facial para complementar el diagnóstico clínico de la ELA, especialmente en entornos con acceso limitado a especialistas en neurología. Se recomienda en futuros estudios la integración de datos auditivos y el uso de modelos avanzados como redes neuronales gráficas para mejorar la precisión del sistema. La implementación de estas herramientas en la práctica médica podría facilitar una detección más rápida y accesible de la ELA, optimizando los tiempos de diagnóstico y tratamiento.

Capítulo 1. Introducción

1.1. Introducción

La esclerosis lateral amiotrófica (ELA) es una enfermedad neurodegenerativa caracterizada por una parálisis muscular progresiva que refleja la degeneración de las neuronas motoras en la corteza motora primaria, los tractos corticoespinales, el tronco encefálico y la médula espinal. Los síntomas están relacionados con debilidad y desgaste muscular focal, espasticidad en las extremidades, afecta la destreza manual y la marcha, así como disartria y disfagia para sólidos o líquidos. La parálisis es progresiva y conduce a la muerte debido a insuficiencia respiratoria en un plazo de 2 a 5 años aproximadamente [1]. El tratamiento de la ELA es de apoyo, paliativo y multidisciplinario, ya que involucra a diversos especialistas para abordar los múltiples efectos de la enfermedad. Neurólogos, neumólogos, fisioterapeutas, terapeutas ocupacionales, nutricionistas, psicólogos y otros profesionales trabajan en conjunto para mejorar la movilidad, respiración, comunicación, nutrición y bienestar emocional del paciente [1].

La detección temprana de la ELA es crucial, ya que permite iniciar tratamientos que pueden mejorar la calidad de vida de los pacientes y extender su expectativa de vida. Sin embargo, el diagnóstico suele ser tardío debido a la falta de herramientas diagnósticas específicas y accesibles, así como a la similitud de los síntomas de la ELA con otras patologías neurológicas como la Atrofia Muscular Espinal, la Parálisis Bulbar Progresiva y la Enfermedad de Kennedy, que también dañan las motoneuronas [2]. A diferencia de muchas, la ELA afecta tanto motoneuronas superiores como inferiores y progresa de forma irreversible.

Actualmente, el diagnóstico depende en gran medida de la experiencia de los especialistas y de pruebas clínicas que no siempre están disponibles o son concluyentes en etapas iniciales. Esto subraya la necesidad de desarrollar sistemas automáticos y no invasivos que puedan ayudar a identificar la enfermedad de manera más eficiente y precisa, facilitando la toma de decisiones clínicas en tiempo real y reduciendo el tiempo de diagnóstico [3].

Nuestra propuesta busca contribuir en el diagnóstico mediante el uso de aprendizaje automático (AA) para la detección de ELA, utilizando imágenes faciales como datos biométricos. Al adaptar y preprocesar una base de datos que contenga estos indicadores, pretendemos desarrollar un modelo que identifique patrones característicos de la enfermedad. Esta herramienta puede complementar el trabajo de

los médicos neurólogos, aliviando parte de la carga que enfrenta el sistema de salud, especialmente en un país como México, donde hay un déficit significativo de especialistas en neurología y una brecha amplia con el diagnóstico [4].

En esta propuesta, se plantea utilizar técnicas de visión computacional para extraer características relevantes de las imágenes faciales, a partir de representaciones de datos que resalten la simetría durante movimientos faciales que podrían estar alterados debido a la afectación neuromuscular. Estas técnicas permitirán obtener una caracterización de los gestos y movimientos sutiles del rostro, proporcionando datos objetivos que puedan alimentar al modelo de aprendizaje automático (AA). De esta manera, la visión computacional se convierte en una herramienta clave para automatizar la detección de signos tempranos de la ELA, posibilitando un diagnóstico más accesible y eficiente.

1.2. Antecedentes

“Esclerosis lateral” se refiere a la dureza a la palpación de las columnas laterales de la médula espinal en las muestras de autopsia, donde la gliosis sigue a la degeneración de los tractos corticoespinales. “Amiotrófica” se refiere a la atrofia muscular, la debilidad y la fasciculación que indican enfermedad de las neuronas motoras inferiores. En pacientes con ELA típica, los síntomas son principalmente de debilidad, que puede comenzar en las manos o las piernas o manifestarse por dificultad para hablar y disfagia [5]. Aunque también incluyen dificultades del habla conocidas como disartria, debilidad en la mitad inferior de la cara, contracciones musculares involuntarias llamadas fasciculaciones y atrofia de la lengua [1].

A nivel mundial, el 50% de los pacientes mueren dentro de los 30 meses posteriores al inicio de los síntomas, y aproximadamente el 20% de los individuos sobreviven entre 5 y 10 años [2]. En cambio, en México, el promedio de supervivencia es de 68.6 meses [4].

La detección temprana de la ELA es crucial para mejorar el pronóstico y calidad de vida de los pacientes. Sin embargo, lograr diagnósticos oportunos sigue siendo un desafío. Por ello, en los últimos años se ha involucrado la computación, que ha tratado de abonar al diagnóstico haciendo uso de señales biomédicas como la electromiografía, señales de voz e imágenes faciales. A través de su caracterización y el uso de inteligencia artificial, es posible detectar patrones asociados a la enfermedad. En particular, el análisis de imágenes faciales ha cobrado relevancia, ya que la ELA afecta la musculatura orofacial, lo que permite explorar biomarcadores visuales que contribuyan a su diagnóstico temprano. Por ejemplo, Bandini *et al.* [6] proponen un método innovador

para detectar ELA mediante el análisis de videos que capturan movimientos faciales durante tareas específicas de habla y no habla. Este enfoque utiliza un sensor Intel RealSense™® SR300 para registrar videos y extraer características cinemáticas como la velocidad, la simetría y el rango de movimiento de los labios. Las tareas evaluadas incluyeron la repetición de frases y movimientos faciales simples, como abrir la mandíbula y sonreír. Utilizando modelos de clasificación como máquinas de soporte vectorial (SVM) y regresión logística, el sistema logró una precisión de hasta el 88.9% en ciertas tareas, destacando la utilidad del análisis facial para detectar síntomas bulbares en etapas tempranas de ELA. Este método también ofrece ventajas como la posibilidad de monitoreo remoto y la reducción de costos clínicos.

En el año 2021 Bandini *et al.* [7] presentaron el conjunto de datos Toronto *NeuroFace* utilizado en este estudio, el primero en incluir videos de gestos orofaciales de individuos con ELA y accidentes cerebrovasculares, así como controles sanos. Este recurso incluye más de 3300 fotogramas anotados con marcadores faciales clave y métricas clínicas asociadas. Los experimentos realizados con este conjunto de datos revelaron sesgos en modelos preentrenados de alineación facial, destacando cómo estos tienden a desempeñarse peor en pacientes con disfunciones orofaciales. Sin embargo, los resultados también demostraron que ajustar los modelos con datos de las poblaciones objetivo reduce significativamente estos sesgos. Este conjunto de datos ofrece una base robusta para avanzar en el desarrollo de herramientas no invasivas y automatizadas para la evaluación clínica [6].

Jin *et al.* [8] investigaron el diagnóstico de la enfermedad de Parkinson (EP) mediante el análisis de expresiones faciales a partir de videos. El estudio utilizó características como la amplitud de los movimientos faciales y el temblor de pequeños grupos musculares, extraídas de los puntos clave faciales detectados con Face++. Empleando algoritmos de AA, como SVM y modelos de memoria a largo plazo, el sistema alcanzó una precisión destacable, con valores de F1 del 99% al usar SVM. Este enfoque resalta el potencial de los análisis faciales como una herramienta no invasiva y remota para el diagnóstico de EP, complementando otros métodos como el análisis de patrones de habla [8].

Por otra parte, Gomes *et al.* [9] introducen un enfoque basado en redes neuronales gráficas (*Graph Neural Networks*, GNN) llamado *Facial Point Graphs* para detectar ELA de manera automática. Este método utiliza puntos clave faciales seleccionados estratégicamente para construir grafos que capturan las características geométricas del rostro. El conjunto de datos Toronto *NeuroFace* fue utilizado para entrenar y validar el modelo, con tareas como el esparcimiento de labios (SPREAD) y la apertura máxima de la mandíbula (OPEN). Los resultados mostraron que el modelo

logró una precisión del 81.8% en ciertas tareas, superando otros métodos en muchos casos. Este enfoque elimina la necesidad de características diseñadas manualmente, lo que simplifica su implementación y mejora su aplicabilidad clínica [9].

1.3. Planteamiento del problema

El diagnóstico de ELA es especialmente complicado debido a la variabilidad de los hallazgos clínicos en las fases iniciales y la ausencia de marcadores biológicos o pruebas específicas que permitan su detección temprana. Entre el 10% y el 42% de los diagnósticos iniciales son erróneos, lo que retrasa el tratamiento y afecta tanto a los pacientes como a los médicos; esto se debe, en parte, al desconocimiento de la enfermedad, presentaciones clínicas atípicas, similitudes con otras patologías y errores en la interpretación de pruebas neuroimágenes y neurofisiológicas [10]. En México, hay un neurólogo certificado por cada 170,800 habitantes [11], lo cual es insuficiente para atender a más de 126 millones de personas [12], lo que contribuye a que el diagnóstico se retrase aún más. Ante esta problemática, la computación ha sido empleada para asistir en el diagnóstico mediante el análisis de señales biomédicas y modelos de inteligencia artificial. Sin embargo, los enfoques previos presentan limitaciones: los métodos basados en señales de voz o electromiografía pueden ser afectados por variabilidad en la adquisición de datos, mientras que las técnicas de procesamiento de imágenes faciales requieren enfoques computacionalmente exigentes, como la construcción de grafos de marcadores faciales, lo que dificulta su implementación en entornos clínicos con recursos limitados. Estas limitaciones resaltan la necesidad de desarrollar modelos más eficientes y accesibles para apoyar el diagnóstico temprano de la ELA [13].

En el ámbito computacional, la detección automática de ELA presenta varios desafíos clave. El uso de sistemas costosos y el preprocesamiento intensivo para extraer características faciales precisas limitan su viabilidad en entornos con recursos limitados. Además, el procesamiento cuadro por cuadro y la construcción de grafos faciales son tareas exigentes en términos de cómputo, dificultando su aplicación en tiempo real. Estos obstáculos complican la adopción de estas herramientas en la práctica clínica, especialmente en entornos con acceso restringido a tecnología avanzada.

Por ello, en este estudio se busca crear un modelo de AA para la detección de ELA con información de imágenes faciales en videos en ambientes no controlados, con el fin de proveer una herramienta que podría complementar las herramientas diagnósticas para ELA.

1.4. Justificación

Nuestra propuesta de desarrollar un modelo de AA para la detección de ELA tendrá un impacto positivo en diversas áreas. Uno de los principales beneficios de un diagnóstico temprano es que permitiría iniciar tratamientos sintomáticos y paliativos de manera más rápida, lo que mejora significativamente la calidad de vida de los pacientes y prolonga su expectativa de vida. Además, el diagnóstico anticipado ayuda a que las familias puedan planificar el cuidado del paciente y optimizar sus recursos.

Sumamos a esto los beneficios de trabajar en un entorno no controlado. Al utilizar datos capturados en entornos cotidianos, en lugar de laboratorios controlados, desarrollamos un modelo más robusto y adaptable a las condiciones reales de los pacientes. De esta manera, hacemos que nuestra herramienta diagnóstica sea más accesible y funcional en la vida diaria, donde los pacientes no siempre cuentan con tecnologías avanzadas o instalaciones especializadas. Analizamos datos en ambientes no controlados para reflejar mejor la variabilidad y complejidad de la vida real, aumentando la capacidad del modelo para generalizar y detectar patrones relevantes en condiciones menos ideales.

Además, el estado del arte en la aplicación de AA al diagnóstico de enfermedades neurológicas ha mostrado avances significativos en el uso de datos biométricos como imágenes faciales.

1.5. Hipótesis

El análisis de imágenes faciales utilizando técnicas de aprendizaje automático, permite identificar la esclerosis lateral amiotrófica en videos en ambientes no controlados.

1.6. Objetivos

1.6.1. Objetivo general

Desarrollar una metodología para la detección de ELA a partir del análisis de videos faciales en ambientes no controlados, utilizando herramientas de visión computacional y aprendizaje automático.

1.6.2. Objetivos específicos

Identificar y seleccionar una base de datos que contenga imágenes faciales de pacientes diagnosticados con ELA.

Preprocesar los datos biométricos para minimizar ruido, y adecuar los datos para su análisis mediante técnicas de AA.

Extraer características relevantes de los datos biométricos que permitan distinguir patrones asociados a la ELA.

Entrenar, validar y evaluar un modelo de AA capaz de detectar la presencia de ELA a partir de las características biométricas extraídas.

Capítulo 2. Marco Teórico

2.1. Esclerosis Lateral Amiotrófica (ELA)

ELA es una enfermedad caracterizada por la degeneración progresiva de las neuronas motoras superiores e inferiores, lo que conlleva a una pérdida gradual del control muscular [3]. Como se mencionó en la sección 1.3 esta enfermedad afecta a la población mexicana [4] que en el año 2014 se estimaba que la prevalencia era de 5000 a 7000 pacientes [14]. Su caracterización y diagnóstico es complicado ya que existe una variabilidad de los hallazgos clínicos en las fases iniciales y una falta de marcadores biológicos o pruebas específicas que permitan su detección temprana [10]. A continuación, se presentan las características de esta condición así como las técnicas de diagnóstico actuales.

2.1.1. Características de ELA

Aunque la causa exacta de ELA es desconocida, se han identificado diversas manifestaciones clínicas, entre las que se incluyen debilidad muscular asimétrica en las extremidades, disartria (dificultad para hablar), disfagia (dificultad para tragar), espasticidad (rigidez muscular), fasciculaciones (contracciones involuntarias de los músculos) y, en etapas avanzadas, dificultades respiratorias. Además, en algunos casos, se ha reportado una asociación con la demencia frontotemporal [10], por lo que el diagnóstico correcto y temprano es de vital importancia para retrasar en medida de lo posible la aparición de estos síntomas.

2.1.2. Diagnóstico de ELA

Si bien, no existe una prueba definitiva para el diagnóstico de la ELA, éste se fundamenta en la combinación de signos de disfunción tanto de las neuronas motoras superiores (NMS) como de las inferiores (NMI), la progresión de los síntomas y la exclusión de otras enfermedades a través de pruebas neurofisiológicas, de neuroimagen y de laboratorio [10].

2.1.2.1. Diagnóstico médico tradicional

Un examen fundamental para el diagnóstico médico de ELA es la electromiografía (EMG), que puede detectar la degeneración de las NMI antes de que los síntomas se manifiesten clínicamente. La EMG evalúa los potenciales de denervación crónica y activa, lo que permite diferenciar la ELA de otras enfermedades que afectan el sistema nervioso [10].

Por otro lado, la escala ALSFRS-R (Escala de Clasificación Funcional Revisada para la ELA) es una herramienta ampliamente utilizada para evaluar el deterioro funcional en pacientes con ELA. Consta de 12 ítems que abarcan cuatro dominios clave: función bulbar, motora gruesa, motora fina y respiratoria. Aunque fue diseñada originalmente para ensayos clínicos, también se emplea en la práctica clínica y en autoevaluaciones remotas para monitorear la progresión de la enfermedad. La puntuación total de esta escala permite predecir la supervivencia y ajustar los tratamientos de manera personalizada [15].

A pesar de los avances, el diagnóstico de ELA sigue siendo un desafío debido a la falta de un marcador biológico específico y a la variabilidad en los síntomas iniciales. Esto puede llevar a diagnósticos erróneos en hasta un 42% de los casos, lo que retrasa el inicio del tratamiento adecuado. En países desarrollados, el tiempo medio desde el inicio de los síntomas hasta el diagnóstico correcto es de entre 12 y 14 meses [10].

2.1.2.2. Diagnóstico basado en pruebas de diadococinesia

La diadococinesia (DDK, por sus siglas en inglés) se refiere a la capacidad de realizar movimientos repetitivos rápidos y alternantes. Estas pruebas evalúan las habilidades oromotoras, y entre ellas se incluyen tareas de tasa de movimiento alternante, que implican la repetición rápida de una sola sílaba como /pa/, /ta/ o /ka/, y tareas de tasa de movimiento secuencial, que consisten en la repetición de una secuencia de varias sílabas diferentes, como la frase /pataka/ de la cual la sílaba /pa/ requiere de la activación gesticular usando los labios, /ta/ es una sílaba que exige el movimiento de la lengua en la zona delantera, y finalmente la sílaba /ka/ motiva la gesticulación de la zona posterior de la lengua; para evaluarla, se pide a los participantes que repitan la frase /pataka/ tantas veces como sea posible en un solo respiro [16].

2.1.2.3. Datos biométricos para el diagnóstico de ELA

El diagnóstico computacional de ELA es un enfoque emergente que utiliza herramientas de inteligencia artificial y AA para identificar patrones que puedan ayudar a detectar la presencia de esta enfermedad neurodegenerativa; los enfoques principales son los siguientes:

Señales de electromiografía: Debido a que detectan la actividad eléctrica en los músculos, su análisis permite detectar anomalías en dicha actividad [17].

Voz: Como se ven afectados los músculos involucrados en el habla es posible usar técnicas de procesamiento de señales y AA para extraer características del habla; así analizar las alteraciones y predecir la enfermedad [18].

Imágenes faciales: Está basado en la identificación de cambios en el movimiento facial, simetría de distintas expresiones a través de puntos clave en el rostro como lo son el contorno de la boca, ojos, cejas, y otras partes móviles. Esto permite realizar un seguimiento de las variaciones en la expresión facial [6].

Estos enfoques ofrecen alternativas prometedoras para mejorar la precisión del diagnóstico y permitir una detección más temprana de la ELA. Por lo general, se desarrollan modelos computacionales que manejan dos clases: “sanos” y “ELA” [13].

2.2. Marcadores faciales para el análisis del movimiento

Los marcadores faciales para el análisis del movimiento son puntos de referencia clave que se colocan en la cara para realizar un seguimiento de los movimientos de un sujeto.

2.2.1. Marcadores faciales

Los marcadores faciales suelen colocarse en áreas anatómicas clave que reflejan con precisión las expresiones y movimientos, como [19]:

- Alrededor de los ojos (cejas, párpados).
- Bordes de la nariz.
- Esquinas de la boca y labios.
- Mandíbula y mejillas.
- Frente.

Estos marcadores faciales son construidos a partir de modelos de AA que permiten etiquetar imágenes individuales o una transmisión continua de imágenes, el objeto resultante contiene coordenadas cartesianas correspondientes a cada punto de referencia facial [20].

2.2.2. Transformación de coordenadas para medición absoluta

Debido a que las coordenadas cartesianas definen distancias respecto a una referencia cambiante con el movimiento del rostro, las distancias entre puntos faciales pueden cambiar de manera inconsistente, afectando estabilidad y comparación de datos, por esto se pueden representar estos mismos marcadores en coordenadas esféricas para definir distancias absolutas.

Las coordenadas esféricas, son un sistema de coordenadas para el espacio tridimensional que especifica la posición de un punto en el espacio mediante tres

números reales. Estos números son la distancia radial absoluta, el ángulo polar y el ángulo azimutal [21].

Se pueden calcular las coordenadas esféricas a partir de las coordenadas x, y, z de un punto utilizando las siguientes ecuaciones:

Distancia radial absoluta (r):

$$r = \sqrt{x^2 + y^2 + z^2} \quad (1)$$

Ángulo polar (θ):

$$\theta = \arccos\left(\frac{z}{r}\right) \quad (r \neq 0) \quad (2)$$

Ángulo azimutal (φ):

$$\varphi = \arctan(y, x) \quad (3)$$

Estas ecuaciones nos permiten transformar puntos definidos en un sistema cartesiano a su representación esférica, proporcionando una referencia absoluta en el espacio tridimensional. Esta conversión resulta fundamental para analizar relaciones espaciales de manera consistente y precisa. Esta explicación se puede entender con mayor profundidad en la representación gráfica mostrada en la Figura 1.

2.2.3. Análisis de las similitudes entre marcadores faciales

Una de las formas de analizar las similitudes existentes entre marcadores faciales, que es un concepto estadístico que mide la relación entre dos señales, indicando cómo es su comportamiento. Esto nos permite identificar los cambios y determinar la naturaleza de esa relación [22].

En la correlación positiva, ambas variables se comportan de manera similar: si una variable aumenta, la otra también lo hace, por consiguiente, si una disminuye, la otra sigue la misma tendencia. Esta relación puede ser débil, moderada o fuerte, dependiendo del valor del coeficiente de correlación; una relación fuerte significa que los puntos están mostrando un patrón claro y consistente entre las señales, una moderada indica que están algo dispersos, pero aún muestran una tendencia general, una relación débil ocurre cuando la tendencia es difícil de identificar. Este grado de relación nos da una idea de la intensidad con la que las variables están conectadas entre sí [22].

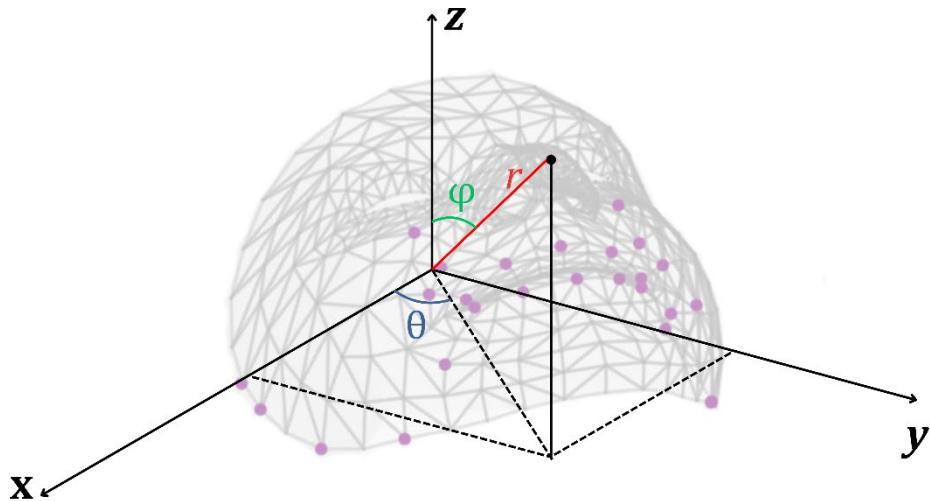


Figura 1. Representación de coordenadas esféricas utilizadas en el análisis. La variable r corresponde a la distancia radial, θ al ángulo polar y φ al ángulo azimutal. Este sistema permite describir la posición de un punto en el espacio tridimensional en relación con un origen común [21].

Adicionalmente existen otras alternativas del análisis de la correlación en función de la medición de los coeficientes. Dependiendo del tipo de datos y de la relación entre las variables, se utilizan diferentes métodos que se describen a continuación:

Coeficiente de Correlación de Pearson

Este coeficiente mide la relación lineal que existe entre dos variables continuas y requiere que ambas sigan una distribución normal. Pearson se usa principalmente cuando se espera que las variables tengan una relación aproximadamente lineal y no existen valores atípicos extremos; tiene un rango de +1 (correlación positiva perfecta) a -1 (correlación perfecta pero negativa) donde 0 denota la ausencia de una relación lineal [22].

Coeficiente de Correlación de Spearman. El coeficiente de Spearman mide la correlación no paramétrica. Utiliza los rangos de las variables en lugar de los valores reales y, por lo tanto, es adecuado para datos ordinales o para relaciones que no necesariamente son lineales. Este coeficiente es útil cuando los datos tienen valores atípicos que podrían distorsionar la correlación de Pearson [23].

Correlación de Kendall (Tau de Kendall)

Este coeficiente también mide la concordancia o discordancia entre dos variables. Es otra alternativa a Spearman y se utiliza para datos no paramétricos u ordinales. La correlación de Kendall suele proporcionar una mejor medida de la relación para muestras pequeñas [23].

El coeficiente de correlación de Pearson, que se usa comúnmente para medir la intensidad de la relación, será un valor cercano a 1 cuando la relación sea muy fuerte y ambas variables aumenten de manera similar.

En la correlación negativa, las dos variables se mueven en direcciones opuestas. Es decir, cuando una variable aumenta, la otra disminuye y viceversa. Esta relación inversa también puede variar en magnitud, dependiendo de la fuerza con la que cambien las variables.

Por otra parte, una correlación nula significa que no hay ninguna relación significativa entre las dos variables. Es decir, los cambios en una variable no tienen ningún impacto predecible en la otra.

En este estudio, usaremos la correlación de Pearson que se emplea para identificar patrones en los movimientos faciales y determinar similitudes de comportamiento a lo largo del tiempo en los videos de los pacientes. Elegimos esta medida porque nos permite evaluar de manera precisa la relación lineal entre las variables, proporcionando una herramienta clara para analizar cómo se sincronizan y se comportan los distintos marcadores faciales con relación a el tiempo.

2.2.4. Análisis del movimiento mediante visión computacional

La visión computacional, rama de la inteligencia artificial, desarrolla sistemas que interpretan imágenes del mundo real mediante algoritmos de procesamiento [24]. Permite realizar tareas como la clasificación de objetos, el reconocimiento facial y el análisis de movimientos, con aplicaciones en medicina, seguridad, industria y entretenimiento. En el ámbito médico, apoya el diagnóstico y el seguimiento de enfermedades al analizar datos visuales de manera rápida y objetiva, brindando a los especialistas información complementaria para tomar decisiones clínicas [25].

Para la detección de la ELA, es posible utilizar la visión computacional como una herramienta valiosa para la obtención de datos de las expresiones faciales y en los movimientos sutiles del rostro afectados por la degeneración de las neuronas motoras. Se aplicarán algoritmos que identifican signos de debilidad muscular, asimetrías en las expresiones o anomalías en los movimientos faciales que son difíciles de percibir a simple vista. Al automatizar este proceso, esperamos un análisis más consistente y

preciso, facilitando la detección temprana de la ELA y permitiendo un diagnóstico más accesible, especialmente en áreas con acceso limitado a especialistas.

2.3. Algoritmos de aprendizaje automático

En el contexto del diagnóstico de ELA, el análisis de movimientos faciales mediante inteligencia artificial ha emergido como una alternativa prometedora para identificar patrones asociados a la enfermedad. El uso de técnicas computacionales permite la extracción y análisis de características faciales, facilitando la detección temprana a través del reconocimiento de alteraciones motoras sutiles.

El AA es un campo de las matemáticas aplicadas que se enfoca en desarrollar e implementar algoritmos capaces de aprender a partir de datos [26]. Este campo se divide en tres enfoques principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo, los cuales han sido utilizados en estudios previos para modelar y predecir la progresión de la ELA a partir de datos biomédicos, incluyendo imágenes faciales.

En la Figura 2 vemos un diagrama que organiza estas categorías y sus métodos, mostrando cómo cada enfoque del AA aborda distintos tipos de problemas de manera específica.

El aprendizaje no supervisado tiene como objetivo aprender la estructura de los datos en ausencia de un resultado o una retroalimentación bien definidos, es decir trabaja con datos no etiquetados, buscando patrones y estructuras ocultas [27]. Por otra parte en el aprendizaje por refuerzo un agente interactúa con un entorno donde aprende a realizar acciones para maximizar una recompensa, mientras que el aprendizaje supervisado se centra en que el modelo se entrena con un conjunto de datos etiquetados donde las respuestas o salidas correctas están disponibles [16]. Dentro de los algoritmos de aprendizaje supervisado existen los de regresión y clasificación, estos últimos son utilizados en tareas donde se asignan datos a categorías predefinidas [27].

En el área médica, los algoritmos de clasificación tienen aplicaciones prometedoras de diagnóstico, pronóstico y estratificación de riesgos y por ello para los fines de este trabajo haremos uso de este tipo de algoritmos [13].

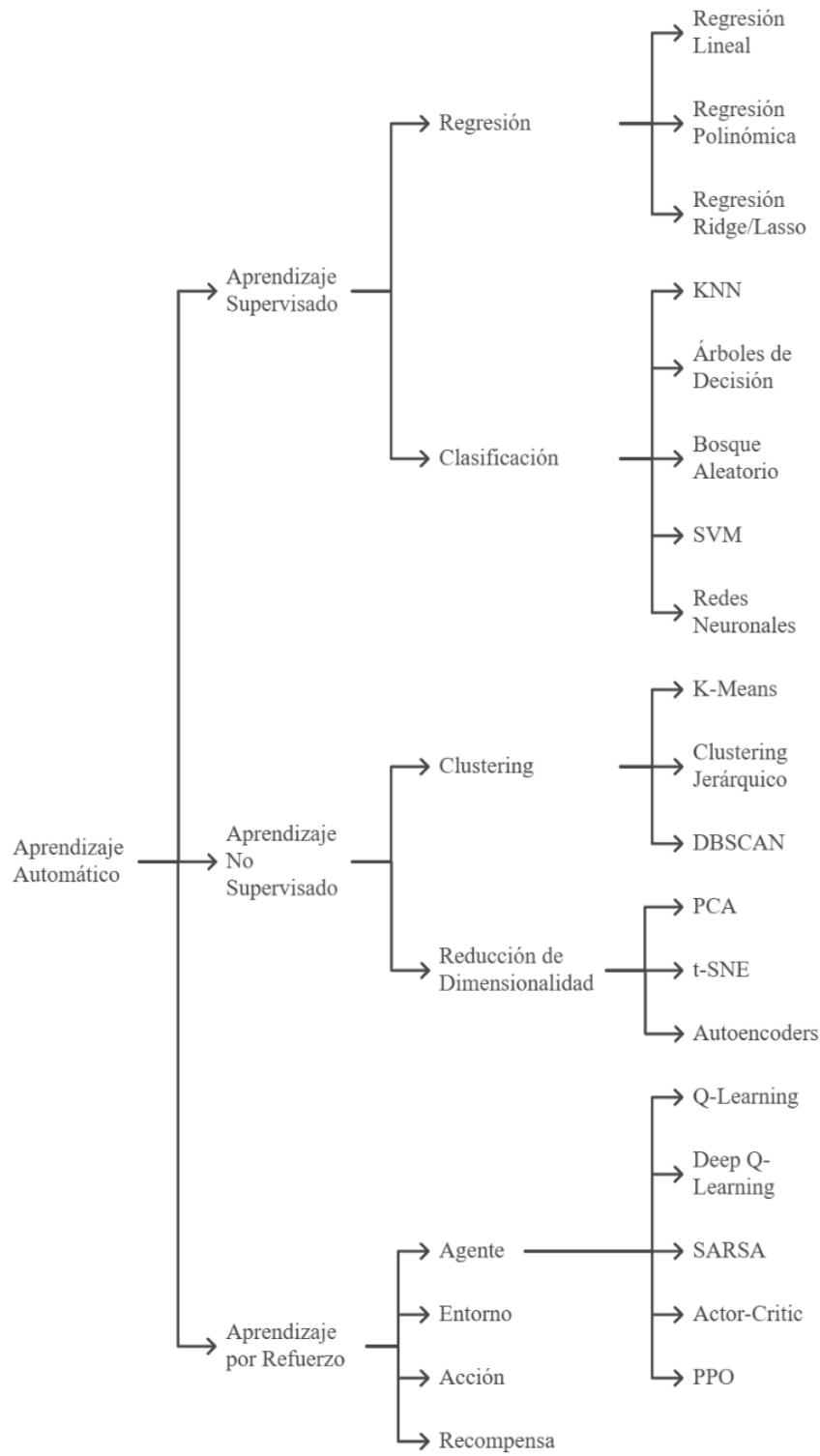


Figura 2. Diagrama de enfoques del aprendizaje automático. Se muestran las tres principales categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo, junto con sus métodos más representativos para resolver distintos tipos de problemas [27].

2.3.1. Algoritmos de AA para clasificación

Los algoritmos de clasificación tienen aplicaciones prometedoras en el área médica debido a su capacidad para identificar patrones complejos en los datos, lo que facilita el diagnóstico, el seguimiento de enfermedades y la toma de decisiones clínicas. Al procesar datos médicos, como imágenes, registros o biomarcadores, estos algoritmos pueden ayudar a los especialistas a detectar condiciones de manera más temprana y precisa, complementando su labor con análisis rápidos y objetivos [27].

Dentro de los algoritmos ampliamente conocidos y más utilizados de AA supervisado de clasificación, podemos encontrar:

Árbol de decisión (DT) es un gráfico que representa las opciones y sus resultados en forma de árbol. Los nodos del gráfico representan un evento o una opción y los bordes del gráfico representan las reglas o condiciones de decisión. Cada árbol consta de nodos y ramas. Cada nodo representa atributos en un grupo que se va a clasificar y cada rama representa un valor que el nodo puede tomar [27].

Random Forest (RF) es un método conjunto basado en DT. Cada árbol en el RF se construye sobre un subconjunto aleatorio de los datos de entrenamiento y las características disponibles. Esto aumenta la robustez a los valores atípicos y la generalización; la estimación final es el promedio o la mayoría de las estimaciones de los árboles dependiendo de si el objetivo es una tarea de regresión o clasificación [27].

Redes neuronales artificiales (RNA) están basadas en un perceptrón que es una versión simplificada de una neurona humana. Funciona tomando entradas, multiplicándolas por pesos específicos y aplicando una función no lineal a la suma de estos productos. Los perceptrones pueden conectarse en capas, formando una red neuronal multicapa, uno de los marcos más comunes en inteligencia artificial. Las RNA ajustan sus pesos mediante métodos de optimización, como el descenso de gradiente estocástico [27].

K-Nearest Neighbors (kNN) este algoritmo selecciona vecinos cercanos de un conjunto de datos para realizar clasificaciones o regresiones, usando la distancia euclíadiana. Los datos deben normalizarse antes de aplicar el algoritmo [27].

Comúnmente al implementar los modelos de AA, realiza una optimización de hiperparámetros que puede ser llevada a cabo utilizando *GridSearch*. Este proceso evalúa sistemáticamente combinaciones de valores de hiperparámetros predefinidos para optimizar el rendimiento de los modelos [28]. El objetivo principal de esta optimización es identificar la mejor configuración para cada modelo, asegurando una alta precisión, una generalización robusta y eficiencia computacional.

GridSearch utiliza validación cruzada para evaluar cada combinación de hiperparámetros dentro del espacio de búsqueda especificado [28].

2.3.2. Métricas de evaluación del desempeño de los modelos

Las métricas de evaluación son fundamentales para medir el rendimiento de un modelo y se utilizan para evaluar diferentes aspectos del rendimiento. Tres métricas relevantes son la precisión, la especificidad y la sensibilidad:

La precisión se utiliza para evaluar la eficacia y se define como la proporción de resultados verdaderos entre el total de casos. La fórmula es:

$$Exactitud = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

donde los verdaderos positivos (TP) representan los casos positivos que fueron predichos correctamente por el modelo, mientras que los verdaderos negativos (TN) corresponden a los casos negativos que también fueron clasificados de manera correcta. Por otro lado, los falsos positivos (FP) son aquellos casos que fueron incorrectamente identificados como positivos, y los falsos negativos (FN) son los casos negativos que el modelo clasificó erróneamente como positivos.

La especificidad se centra en la capacidad del modelo para identificar correctamente los casos negativos, la fórmula es:

$$Especificidad = \frac{TN}{TN + FP} \quad (5)$$

Y la capacidad del modelo para predecir correctamente los casos positivos es la sensibilidad:

$$Sensibilidad = \frac{TP}{TP + FN} \quad (6)$$

Estas métricas proporcionan una visión del rendimiento del modelo, lo que nos permite identificar tanto su capacidad para detectar correctamente (sensibilidad) así como su precisión para predecir sujetos sanos (especificidad) [26].

Capítulo 3. Metodología

El objetivo de este trabajo es desarrollar un modelo de AA para la detección ELA mediante el análisis de datos biométricos, obtenidos a partir de imágenes faciales, utilizando el *conjunto de datos Toronto NeuroFace* [7]. La Figura 3 muestra el diagrama metodológico de nuestra propuesta. Esta se basa en la extracción de puntos de referencia faciales, su procesamiento y la implementación de algoritmos de AA para identificar patrones asociados con la enfermedad. Teniendo como fin ser un auxiliar en el diagnóstico y complementar las evaluaciones clínicas tradicionales.

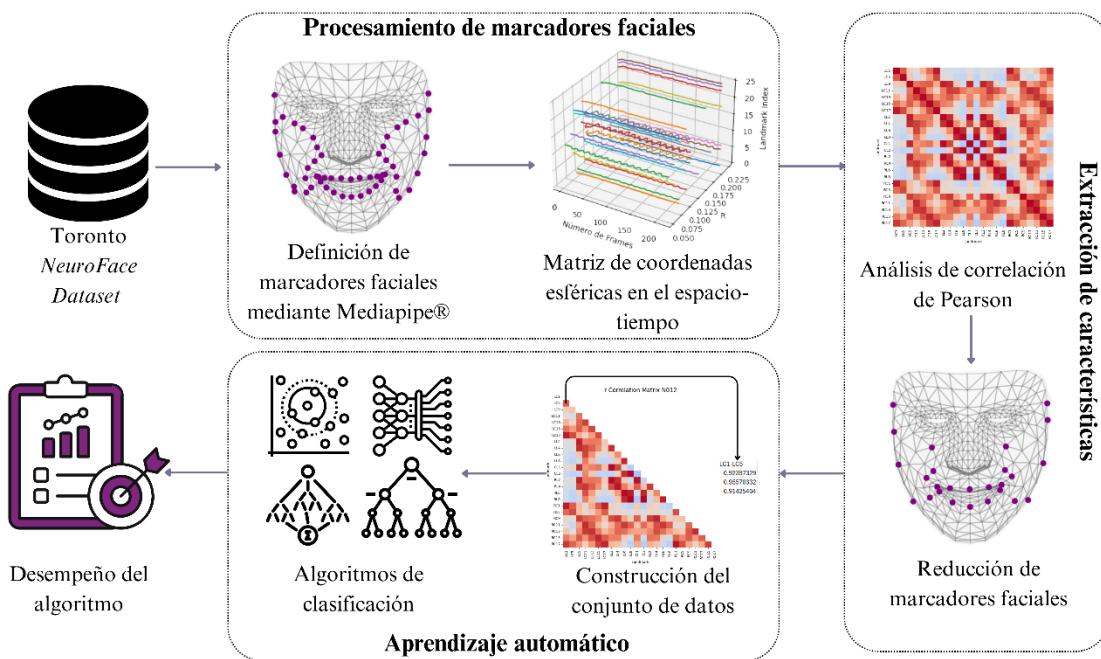


Figura 3: Diagrama metodológico para detección de ELA a través de marcadores faciales.

Además del diagrama metodológico, a continuación, en la Figura 4, presentamos el pseudocódigo de la metodología que proponemos en este trabajo. El pseudocódigo se divide en tres bloques principales: 1) **Entrada**, donde presentamos las variables y sus valores asignados que serán utilizados en toda la metodología propuesta; 2) **Procedimiento**, aquí presentamos todos los pasos de la metodología propuesta, indicando, a través de funciones, los procesamientos que se hacen con los datos; 3) **Salida**, en esta última parte, indicamos la salida de la metodología propuesta.

En las secciones posteriores al pseudocódigo, se detalla cada paso del procesamiento de los datos con mayor detalle.

Entrada:

Conjunto de videos $V = \{v_1, v_2, \dots, v_n\}$, donde cada v_i ha sido previamente segmentado para contener solo la parte relevante y $n = 18$.

Definición de umbral de correlación $\tau = 0.95$.

Número de marcadores faciales seleccionados $L = 24$.

Seleccionar un conjunto de algoritmos de clasificación $A = \{kNN, RNA, RF, DT\}$.

Considerar los métodos de validación $V_m = k - CV$.

Procedimiento:

Paso 1: Extracción de marcadores faciales con MediaPipe®:

$$\forall v_i \in V, [\mathbf{M}_x, \mathbf{M}_y, \mathbf{M}_z] = \text{MediaPipe}(v_i) \quad (7)$$

donde $i = 1, \dots, 18$ videos es un conjunto de 54 marcadores faciales en coordenadas cartesianas (x, y, z), generando tres matrices temporales por sujeto:

\mathbf{M}_x para la coordenada x .

\mathbf{M}_y para la coordenada y .

\mathbf{M}_z para la coordenada z .

Paso 2: Conversión a coordenadas esféricas:

$$(r, \phi, \theta) = f(M_x, M_y, M_z) \quad (8)$$

Lo que genera tres nuevas matrices en el espacio-tiempo:

R para la coordenada radial r .

ϕ para el ángulo de elevación ϕ .

θ para la coordenada polar θ .

Paso 3: Cálculo de las matrices de correlación de Pearson:

$$Cr = \text{Pearson}(R), \varphi = \text{Pearson}(\phi), \theta = \text{Pearson}(\theta) \quad (9)$$

Donde cada C es una matriz 54×54 que representa la correlación entre pares de marcadores faciales.

Paso 4: Umbralización y binarización:

$$Cr, \varphi, \theta'(j, k) = \begin{cases} 1, & \text{si } |Cr, \varphi, \theta(j, k)| \geq \tau \\ 0, & \text{si } |Cr, \varphi, \theta(j, k)| < \tau \end{cases} \quad (10)$$

Paso 5: Promediación de las matrices de todos los sujetos para obtener tres matrices combinadas por coordenada.

Donde los índices j y k corresponden a los índices de la filas y columnas, respectivamente, de la matriz C , y “ $|*$ ” corresponde al valor absoluto.

Paso 6: Reducción de marcadores faciales:

Se identifican marcadores faciales redundantes.

Se seleccionan los $L = 24$ marcadores faciales que coinciden en las tres matrices combinadas

$$(C_r', C_\varphi', C_\theta'). \quad (11)$$

Paso 7: Construcción de un archivo separado por comas (*.csv) con correlaciones por coordenada: Para cada matriz de correlación

$$Cr, \varphi, \theta: \forall j, k \in \{1, \dots, 54\}, j > k \quad (12)$$

Se eliminan:

La mitad diagonal superior ($j < k$) para evitar información redundante, ya que C es simétrica.

La diagonal principal ($j = k$) porque todos los valores son 1.

Los valores restantes $C_r, C_\varphi, C_\theta(j, k)$ se almacenan en un archivo CSV separado por coordenada, donde:

El encabezado indica el par de marcadores faciales correspondientes.

Cada columna representa un sujeto.

Paso 8: Etiquetado, se agrega la etiqueta y_i que indica si el sujeto es sano o con ELA.

Búsqueda de hiperparámetros:

$$B = \text{argmax} \Theta \in Hscore(A, D, V_m) \quad (13)$$

donde:

H es el espacio de hiperparámetros.

D es el conjunto de datos.

Paso 9: Evaluación, se evalúan k-fold cross-validation ($k = 3, 6$) y leave-one-out.

Entrenamiento y evaluación de modelos:

$$\forall A \in A, A * = A(D, \Theta *) \quad (14)$$

Paso 10: Cálculo de desempeño, se mide el desempeño con métricas de exactitud, sensibilidad y especificidad.

Salida:

Modelos entrenados $A *$ con los mejores hiperparámetros. Evaluación del rendimiento de cada algoritmo sobre el conjunto de datos.

Figura 4: Pseudocódigo de la metodología propuesta para detección de ELA.

3.1. Exploración de bases de datos neurofaciales

3.1.1. Base de datos de ELA

Para este estudio se utilizó el Toronto *NeuroFace Dataset*, recopilado por el equipo de *Yana Yunusova en el UHN-Toronto Rehabilitation Institute* en Canadá. Esta base de datos se divide en tres grupos, sujetos con ELA, sujetos de control sanos y sujetos post-accidente cerebrovascular (*Stroke*); para acceder a esta base de datos, es necesario completar formularios y acuerdos de uso que regulan su disponibilidad y aplicación en investigaciones [7].

Cada grupo contiene de cada sujeto:

- **Videos:** Tareas orofaciales grabadas en formato *.avi*.
- **Frames:** Imágenes en *.jpg* extraídas de los videos.
- **Marcadores faciales:** Archivos *.txt* con 68 puntos de referencia faciales para cada cuadro.
- **Bounding boxes:** Archivos *.txt* con coordenadas de las cajas delimitadoras faciales.
- **Información clínica:** Datos demográficos y duración de la enfermedad para pacientes con ELA.
- **Evaluaciones clínicas:** Calificaciones de simetría, rango de movimiento, variabilidad y fatiga por dos patólogos del habla en una escala de 1 (normal) a 5 (severa).

Las grabaciones se realizaron usando una cámara Intel RealSense® SR300 a una distancia de entre 30 y 60 cm, con una fuente de luz continua para iluminar uniformemente el rostro. Los videos se almacenaron con una resolución de 640x480 píxeles y una frecuencia de cuadro de aproximadamente 50 fps [7]. En la Figura 5, podemos observar de manera gráfica un resumen del contenido de la base de datos utilizada. Cabe mencionar que en este trabajo solo utilizamos los datos correspondientes al grupo de ELA y control.

3.1.2. Datos auditivos

Los datos auditivos proporcionados por la base de datos se adquirieron mediante la grabación de tareas específicas diseñadas para evaluar el control motor del habla y el desempeño orofacial. Estas tareas incluyeron:

- **Repetición de frases:** Una de las frases utilizadas fue "*Buy Bobby a Puppy*", elegida por su complejidad fonética y capacidad para resaltar problemas de articulación.
- **Repetición de sílabas:** Los participantes realizaron ejercicios como la repetición aislada de /pa/ lo más rápido posible en una respiración y otra prueba de diadococinesia (*DDK*) con la secuencia trisilábica /pataka/ que está diseñada para medir la coordinación motora, la capacidad de producir secuencias rápidas de movimientos orales y la habilidad para cambiar rápidamente entre posiciones articulatorias (labial, alveolar y velar) siendo este un indicador clave del control motor del habla [16].

Para este estudio en particular, se seleccionaron 18 videos de participantes realizando la tarea de *DDK* con /pataka/. Estos videos se grabaron mientras los individuos pronunciaban la secuencia lo más rápido posible en una sola respiración. De los 18 videos, 10 correspondieron a sujetos de control y 8 a pacientes con ELA. Dos videos adicionales fueron descartados debido a errores de codificación.

El análisis de los datos auditivos nos permite evaluar aspectos como:

- **Velocidad de pronunciación:** Qué tan rápido pueden repetir las secuencias.
- **Precisión:** Detección de errores en la articulación o sustitución de fonemas.
- **Control motor:** Variaciones en la coordinación de los movimientos orales durante la pronunciación.

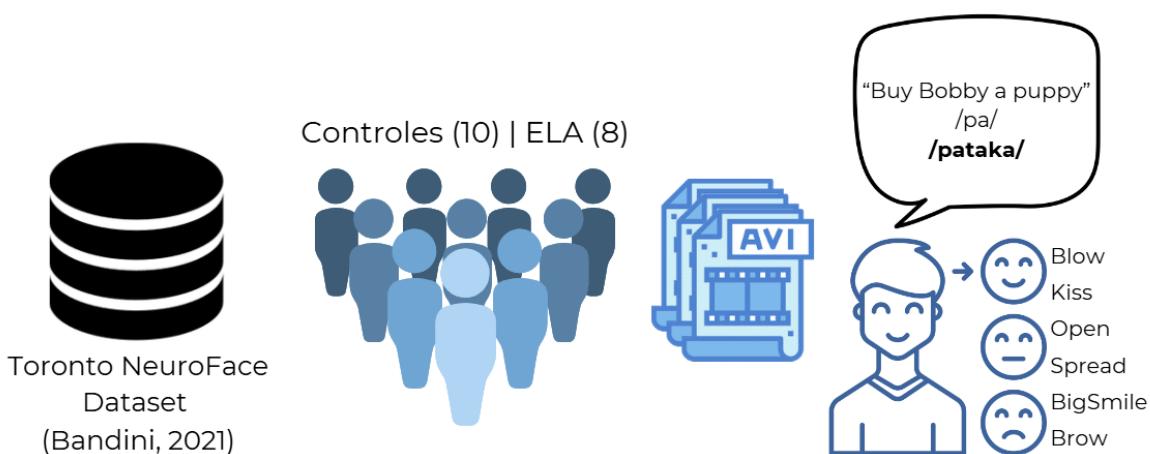


Figura 5: Resumen gráfico del Toronto NeuroFace Conjunto de datos utilizado en el estudio: conjunto de datos que incluye videos y anotaciones faciales de participantes con ELA y controles sanos, enfocado en tareas orofaciales como gestos y repetición de frases.

3.1.3. Datos faciales

Los datos faciales en el conjunto de datos incluyen una variedad de tareas orofaciales diseñadas para evaluar el movimiento y la coordinación de los músculos faciales. Las tareas incluyen:

1. Gestos verbales:

- Repetición de frases: "*Buy Bobby a Puppy*".
- Repetición de sílabas: /pa/ y /pataka/.

2. Gestos no verbales:

- **Fruncir los labios:** Simular apagar una vela (Blow) o simular dar un beso (Kiss).
- **Sonrisa:** Con labios apretados (Spread) o una sonrisa amplia (BigSmile).
- **Movimiento mandibular:** Apertura máxima de la mandíbula (Open).
- **Movimiento de las cejas:** Levantarlas (Brow) [7].

Los datos faciales que brindan las tareas orofaciales permiten evaluar de manera objetiva características claves asociadas con la disfunción motora de los músculos orofaciales, como la simetría, el rango de movimiento, la variabilidad y la fatiga; además de ser una alternativa no invasiva y de fácil acceso. Es por lo que, en este estudio utilizamos estos datos para desarrollar un modelo que permita identificar las características asociadas con la ELA.

Para este propósito, nos enfocamos exclusivamente en la tarea /pataka/, ya que ha demostrado ser particularmente efectiva en la evaluación de la coordinación y velocidad articulatoria, dos aspectos fuertemente afectados en pacientes con ELA. Además, esta tarea permite capturar patrones de disfunción motora de manera más consistente y reproducible que otras tareas orofaciales, facilitando así un análisis más preciso de las alteraciones asociadas con la enfermedad [16].

3.2. Preprocesamiento de los datos

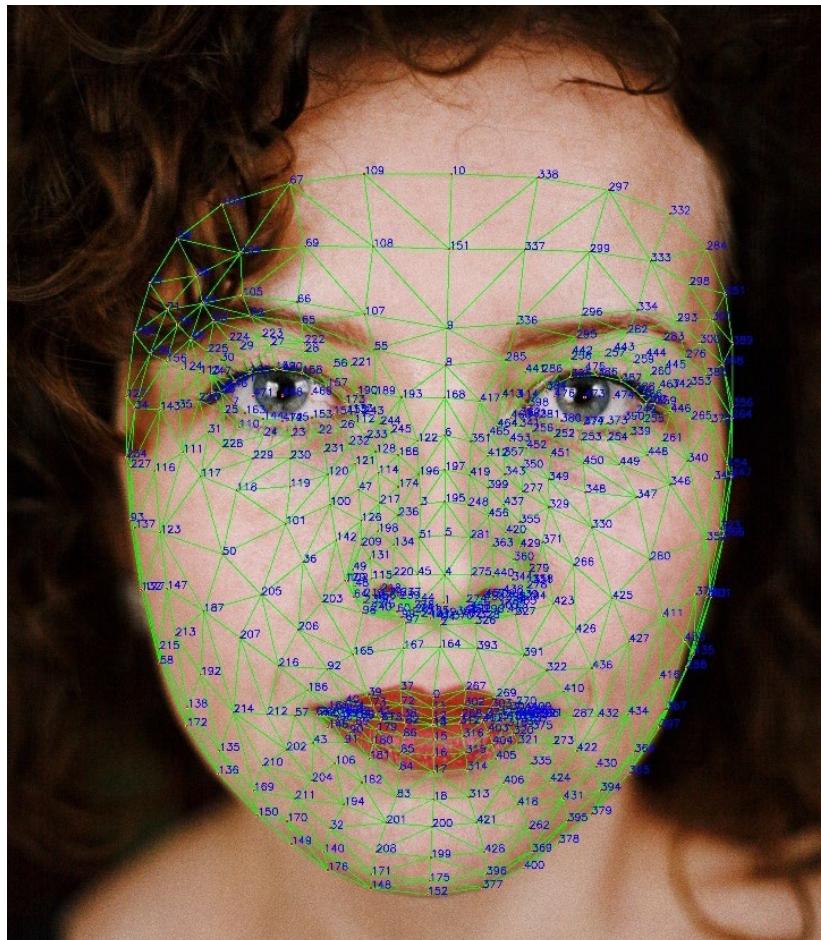
En nuestro proceso de análisis, realizamos un preprocesamiento en los videos para garantizar que los datos utilizados sean consistentes y relevantes. Cortamos los videos originales para dejar únicamente los segmentos donde la persona está realizando la tarea que en este caso es repetir el conjunto de silabas /pataka/ tantas veces como sea posible en una sola respiración; eliminando así cualquier información innecesaria o ruido que pueda afectar el análisis. Este paso nos permite enfocarnos

exclusivamente en los momentos de interés, mejorando la calidad de los datos y reduciendo el tiempo de procesamiento.

Al acotar los videos a las secciones relevantes, no sólo optimizamos los recursos computacionales, sino que también incrementamos la precisión de los algoritmos de visión computacional, ya que trabajamos con datos más claros y específicos. Además, este preprocesamiento asegura que los modelos de AA reciban información que represente de manera más precisa los patrones de interés, como movimientos faciales o expresiones sutiles, que son fundamentales para nuestra investigación. Este enfoque nos permite avanzar de manera más eficiente hacia la detección y análisis de la enfermedad.

3.2.1. Extracción de marcadores faciales

Para identificar los marcadores faciales, se utilizó la biblioteca MediaPipe® [20], que permite la detección automática de 468 puntos de referencia en el rostro, distribución que podemos ver en la Figura 6.



En este estudio, nos enfocamos solamente en 54 marcadores faciales que están distribuidos simétricamente desde la nariz hasta el mentón y alrededor de los músculos faciales, (ver la Figura 7); esta elección se centró en capturar las áreas del rostro afectadas por la ELA, especialmente los músculos faciales involucrados en el habla y la masticación. Entre las áreas clave se encuentran:

- Músculos maseteros: responsables de la masticación, cuya debilidad se manifiesta en pacientes con ELA.
 - Mandíbula: que sufre atrofia progresiva en las etapas avanzadas de la enfermedad.
 - Labios: que muestran una pérdida de control motor en los pacientes con disartria (dificultad para hablar).

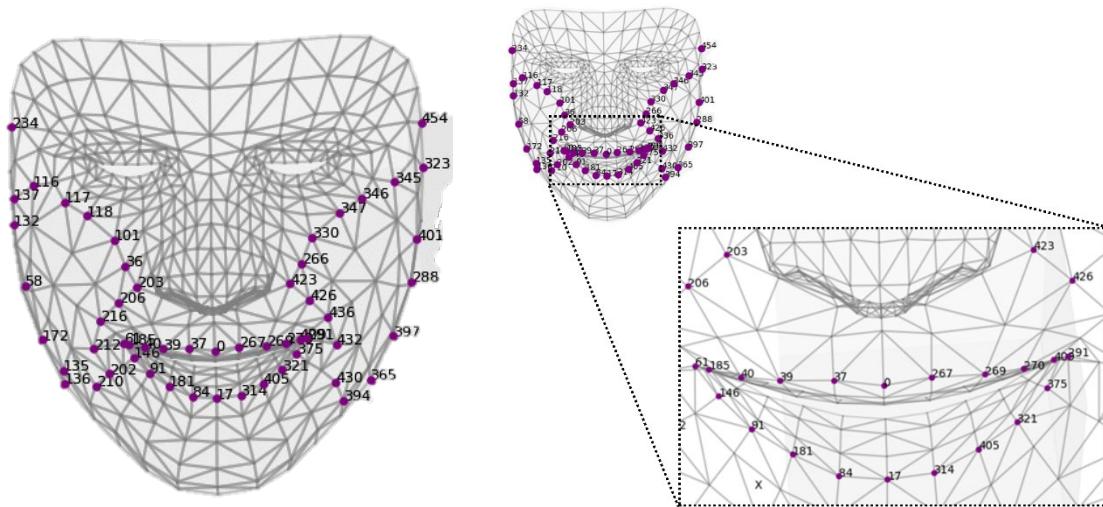


Figura 7: Representación de los marcadores faciales clave para analizar los músculos maseteros, la mandíbula y la zona perioral, áreas afectadas en pacientes con ELA y disartria.

Debido a que Mediapipe® entrega números como identificadores, asignamos nuevas etiquetas para cada uno de los marcadores faciales definidos por Mediapipe®, se colocaron con el fin de poderlos identificar de manera más fácil identificarlos: la primera es una letra que indica el lado del rostro en donde se encuentra el marcador (L: *left*, C: *central* y R: *right*), la segunda letra indica a que parte del rostro corresponde (C: *cheek* y L: *lip*) y el tercer carácter es un número consecutivo de identificación (ver Tabla 1).

Los marcadores faciales se extrajeron automáticamente para cada una de las tramas dentro del video, lo que permitió la creación de un conjunto de datos tridimensionales (x, y, z) para cada sujeto en cada *frame*, proporcionando información sobre la posición y el movimiento de estos puntos clave durante la tarea.

Tabla 1. Asignación de etiquetas a los puntos de referencia faciales detectados con MediaPipe®, para el análisis de regiones clave en estudios relacionados con ELA.

| | | | | | | | | | |
|--------------------------|------|------|------|------|------|------|------|------|------|
| MediaPipe® | 234 | 116 | 117 | 118 | 101 | 36 | 203 | 206 | 216 |
| Etiqueta Definida | LC1 | LC2 | LC3 | LC4 | LC5 | LC6 | LC7 | LC8 | LC9 |
| MediaPipe® | 212 | 202 | 210 | 136 | 172 | 58 | 132 | 137 | 37 |
| Etiqueta Definida | LC10 | LC11 | LC12 | LC13 | LC14 | LC15 | LC16 | LC17 | LL1 |
| MediaPipe® | 39 | 40 | 185 | 61 | 146 | 91 | 181 | 84 | 0 |
| Etiqueta Definida | LL2 | LL3 | LL4 | LL5 | LL6 | LL7 | LL8 | LL9 | CL1 |
| MediaPipe® | 17 | 314 | 405 | 321 | 375 | 291 | 409 | 270 | 269 |
| Etiqueta Definida | CL2 | RL9 | RL8 | RL7 | RL6 | RL5 | RL4 | RL3 | RL2 |
| MediaPipe® | 267 | 323 | 401 | 288 | 397 | 365 | 394 | 430 | 432 |
| Etiqueta Definida | RL1 | RC17 | RC16 | RC15 | RC14 | RC13 | RC12 | RC11 | RC10 |
| MediaPipe® | 436 | 426 | 423 | 266 | 330 | 347 | 346 | 345 | 454 |
| Etiqueta Definida | RC9 | RC8 | RC7 | RC6 | RC5 | RC4 | RC3 | RC2 | RC1 |

*Etiqueta izquierda L: *left*, C: *central* y R: *right*, etiqueta derecha C: *cheek* y L: *lip*

Para visualizar el comportamiento del comportamiento de los marcadores faciales creamos una representación del desplazamiento de cada marcador a través del tiempo durante la tarea, a la que le llamaremos Matriz de coordenadas en el Espacio-Tiempo. En la Figura 8 podemos observar este gráfico, que muestra las oscilaciones que tienen los marcadores faciales de un sujeto a lo largo de la ejecución de la tarea en función del tiempo.

r con respecto al número de frames

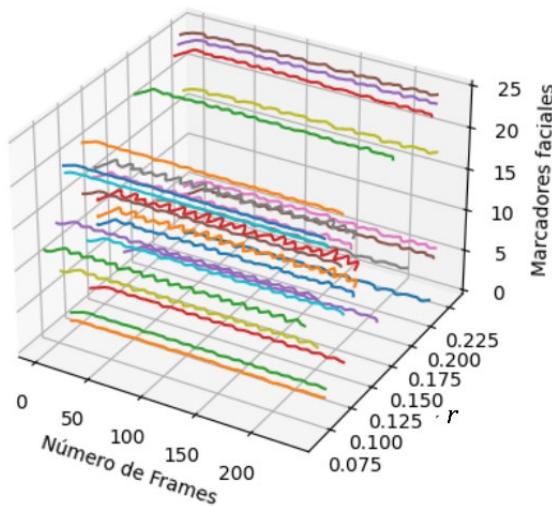


Figura 8: Representación tridimensional del cambio del marcador facial respecto al tiempo (número de frames).

3.2.2. Transformación espacial de coordenadas cartesianas a esféricas

Los marcadores faciales extraídos inicialmente se encuentran representados en coordenadas cartesianas (x, y, z), relativas al centro de masa de la cara. Para reducir el error espacial en el análisis de los movimientos faciales, las coordenadas fueron transformadas a un sistema de coordenadas esféricas.

La transformación a coordenadas esféricas implica describir la posición de cada marcador facial mediante:

r : la distancia radial, que mide la distancia entre cada punto de referencia y el centro de masa de la cara (ubicado cerca del punto central de la nariz).

θ : el ángulo azimutal, que mide la rotación horizontal alrededor del eje vertical de la cabeza.

φ : el ángulo de inclinación en el plano vertical, que describe la posición del marcador facial en relación con el eje vertical de la cabeza.

Esto permite una representación más precisa de la simetría facial y la minimización de los errores causados por variaciones en la posición de la cámara o del movimiento facial en un espacio tridimensional. El uso de coordenadas esféricas asegura que las características extraídas sean más consistentes, independientemente de las pequeñas diferencias en la postura o la posición de los pacientes durante la grabación. Al final, seguimos obteniendo las mismas matrices de características, pero en lugar de estar en un espacio cartesiano, ahora están representadas en coordenadas esféricas, lo que facilita su análisis y comparación dentro del modelo.

3.2.3. Filtrado de marcadores faciales

Una vez que transformamos los datos de los marcadores a coordenadas esféricas, creamos matrices de correlación de Pearson para identificar patrones de comportamiento redundantes entre los marcadores faciales. Utilizamos la correlación de Pearson, una técnica estadística que mide la relación lineal entre dos variables, para detectar marcadores faciales cuyos movimientos están altamente correlacionados entre sí, lo que indica que podrían estar proporcionando información redundante.

Nuestro proceso de filtrado consiste en los siguientes pasos:

- **Cálculo de matrices de correlación:** Calculamos tres matrices de correlación, una para cada dimensión esférica (r, θ, φ). Estas matrices nos permiten observar la relación de movimiento entre cada par de marcadores faciales. Organizamos la matriz de correlación siguiendo un

orden lexicográfico de los puntos de referencia (marcadores faciales). Cada valor dentro de la matriz corresponde a un número que oscilaba entre -1 y 1, representando el grado de relación que existía entre cada par de puntos.

- **Umbral de correlación:** Establecemos un umbral de correlación (0.95) para identificar marcadores faciales con comportamientos altamente correlacionados. Consideramos como redundantes aquellos marcadores faciales cuya correlación supera este umbral.
 - **Reducción del número de marcadores faciales:** Eliminamos del análisis los marcadores redundantes, dejando un conjunto reducido de 24 marcadores faciales.

En la Figura 9, se muestra de manera gráfica el proceso de filtrado de marcadores faciales de interés. Una vez llevado a cabo este proceso, los 24 marcadores faciales seleccionados se pueden observar en la Figura 10 y en la Tabla 2.

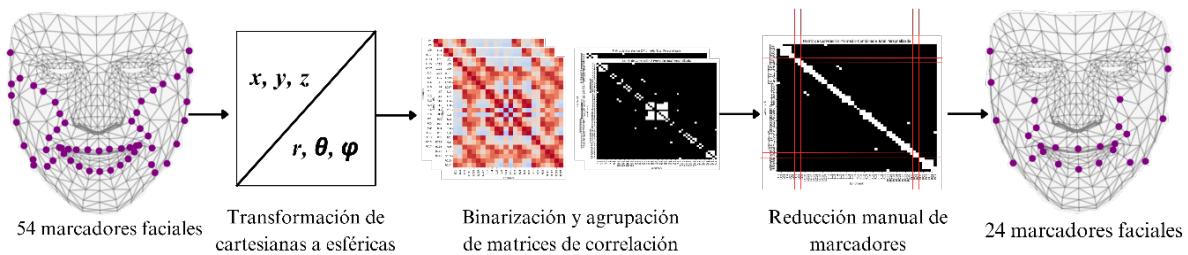


Figura 9: Panorama metodológico para reducción de marcadores faciales en el análisis de ELA, mediante MediaPipe®.

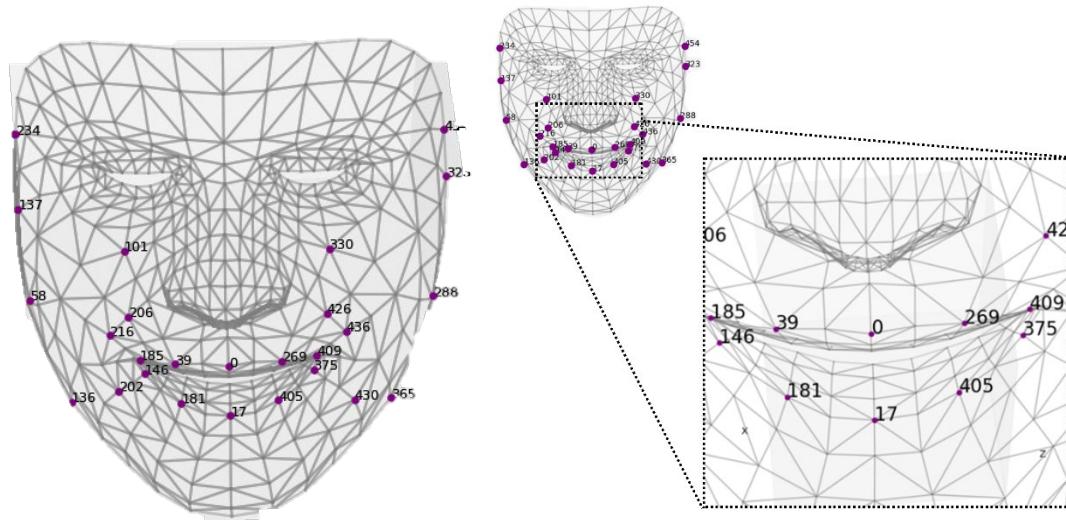


Figura 10: Conjunto optimizado de 24 marcadores faciales seleccionados para análisis de ELA, eliminando redundancias y enfocándose en puntos clave que aportan información relevante sobre los movimientos faciales.

Tabla 2. Asignación de etiquetas a los marcadores faciales reducidos: contiene los 24 puntos seleccionados con sus números en MediaPipe® y etiquetas, optimizados para análisis específicos de movimientos faciales en el estudio.

| | | | | | | |
|--------------------------|------|-----|------|------|------|------|
| MediaPipe® | 234 | 101 | 216 | 202 | 136 | 58 |
| Etiqueta Definida | LC1 | LC5 | LC9 | LC11 | LC13 | LC15 |
| MediaPipe® | 137 | 39 | 185 | 146 | 181 | 0 |
| Etiqueta Definida | LC17 | LL2 | LL4 | LL6 | LL8 | CL1 |
| MediaPipe® | 17 | 269 | 409 | 375 | 405 | 454 |
| Etiqueta Definida | CL2 | RL2 | RL4 | RL6 | RL8 | RC1 |
| MediaPipe® | 330 | 436 | 430 | 365 | 288 | 323 |
| Etiqueta Definida | RC5 | RC9 | RC11 | RC13 | RC15 | RC17 |

*Etiqueta izquierda L: left, C: central y R: right, etiqueta derecha C: cheek y L: lip

Este proceso de reducción de marcadores faciales es fundamental para simplificar el procesamiento y acelerar los algoritmos de AA sin sacrificar precisión. Al mantener sólo los marcadores faciales más informativos, se facilita el análisis de la simetría facial y se mejora el rendimiento del modelo de clasificación en términos de exactitud, sensibilidad y especificidad. Además, al reducir la dimensionalidad del conjunto de datos, se previene el sobreajuste, especialmente en modelos con un número limitado de muestras.

3.3. Definición de las características faciales relevantes

Ahora que contamos con nuestros 24 marcadores faciales seleccionados, los cuales aportan información no redundante, procedemos a construir nuestros conjuntos de datos. Organizamos los datos en tres conjuntos distintos, correspondientes a las tres dimensiones esféricas: r , θ y φ . Cada conjunto de datos está compuesto de 4968 datos que son los valores numéricos de las correlaciones calculadas entre cada par de marcadores faciales. En cada caso, registramos el coeficiente de correlación que describe la relación lineal entre los movimientos de los dos marcadores en la dimensión correspondiente. Estos coeficientes nos ofrecen una medida de cómo interactúan los marcadores faciales en cada eje, ayudando a identificar patrones y comportamientos relevantes. Cada instancia en los conjuntos de datos representa un sujeto, mientras que cada columna corresponde a un par de marcadores faciales. A continuación, en la Figura 11 nos muestra una matriz de correlación correspondiente al conjunto de datos r , donde se destacan los valores de correlación para el par de marcadores faciales LC1-

LC5, valores que forman parte de los conjuntos de datos que conformaron nuestras entradas a los AA, cada instancia corresponde a un sujeto y cada columna a un par de marcadores faciales.

Adicionalmente, incluimos en los conjuntos de datos la etiqueta de la clase a la que pertenece cada muestra, lo que nos permite relacionar estos datos con el contexto del análisis. Esta estructura de los conjuntos de datos nos facilita integrarlos posteriormente en los modelos de AA, asegurando que contamos con una representación completa y organizada de los datos necesarios para entrenar y evaluar nuestros algoritmos.

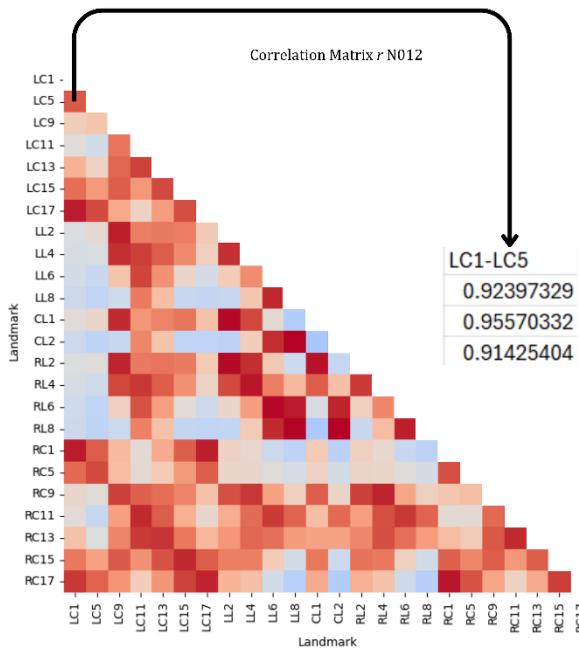


Figura 11: Diagrama correspondiente a la construcción del conjunto de datos r , se muestran los valores de correlación entre el par LC1-LC5.

3.4. Modelado con algoritmos de Aprendizaje Automático

El modelado con algoritmos de AA es el proceso de entrenar modelos para aprender de los datos y hacer predicciones basadas en patrones identificados. En este estudio, se aplicaron varios algoritmos de AA a los datos faciales caracterizados con el objetivo de clasificar a los pacientes entre aquellos con ELA y controles sanos.

El enfoque metodológico se centró en utilizar un conjunto de algoritmos supervisados, dado que las etiquetas de clase (pacientes con ELA y controles) están disponibles. La implementación de varios modelos permite comparar su rendimiento y elegir el más adecuado para el diagnóstico temprano de la ELA.

3.4.1. Algoritmos de Aprendizaje Automático implementados

Se seleccionaron cuatro algoritmos de AA que cubren un espectro diverso de enfoques para la clasificación de los datos biométricos faciales. Cada uno de estos algoritmos tiene sus propias fortalezas y debilidades, lo que justifica su inclusión para evaluar cuál puede ofrecer el mejor rendimiento en términos de exactitud, especificidad y sensibilidad. Además, se realizó un proceso de búsqueda de hiperparámetros utilizando *GridSearch* para optimizar el rendimiento de cada modelo:

- kNN: elegido por su simplicidad y capacidad para clasificar datos basados en la cercanía en el espacio de características.
- DT: por su interpretabilidad y capacidad para modelar reglas de decisión claras.
- RF: por su robustez al reducir el sobreajuste y combinar múltiples DT.
- RNA: por su capacidad para modelar patrones complejos no lineales.

En este estudio, la métrica de evaluación seleccionada es la precisión, y se utilizó validación cruzada estratificada para garantizar una evaluación balanceada entre las clases.

Se exploraron las siguientes que están en la Tabla 3 donde se presentan las combinaciones de hiperparámetros para cada algoritmo y los conjuntos de datos:

Tabla 3. Resumen de los principales hiperparámetros y sus valores para cada modelo de aprendizaje automático analizado.

| Algoritmo | Hiperparámetro | Valores |
|------------------|-------------------------------|---------------------------------|
| RF | Criterio de división | Gini, Entropy |
| | Profundidad máxima | None, 1, 2, 10, 20 |
| | Muestras mínimas por división | 1, 2, 5 |
| | Selección de características | True, False |
| DT | Criterio de división | Gini, Entropy |
| | Profundidad máxima | 2, 5, 10 |
| | Estrategia de división | Random, Best |
| kNN | Métrica de distancia | Euclidean, Manhattan, Minkowski |
| | Número de vecinos | 3, 5, 7 |
| RNA | Método de ponderación | Uniform, Distance |
| | Función de activación | Tanh, Relu, Logistic |
| | Tasa de aprendizaje | 0.0001, 0.001, 0.01 |
| | Número de capas ocultas | 2, 50 y 100 |
| | Optimizadores | Adam, LBFGS |

3.4.2. Entrenamiento y validación

Para garantizar la validez de los modelos, se utilizó un esquema de validación cruzada estratificada (*k-fold*) con $k=3$ y $k=6$, así como el método LOO.

En el caso de la validación *k-fold*, el grupo de datos se divide en k subconjuntos (*folds*). En cada iteración, un subconjunto se utiliza como conjunto de prueba, mientras que los otros ($k-1$) subconjuntos se emplean para entrenar el modelo. El proceso se repite k veces, asegurando que cada subconjunto se utilice al menos una vez como conjunto de prueba.

Por otro lado, con el enfoque LOO, se utiliza un único sujeto como conjunto de prueba, mientras que el resto se usa para el entrenamiento. Este método maximiza la cantidad de datos disponibles para entrenar el modelo, aunque con un mayor costo computacional.

Este esquema es particularmente útil en conjuntos de datos pequeños como es el caso, ya que garantiza que todos los datos se utilicen tanto para el entrenamiento como para la validación. La estratificación asegura que en cada uno de los subconjuntos haya una proporción equilibrada de pacientes con ELA y controles sanos, evitando sesgos en las predicciones.

Capítulo 4. Resultados y discusión

En esta sección, presentamos los resultados obtenidos a partir del análisis de los marcadores faciales en el contexto de la ELA. La exposición de los resultados se divide en cuatro partes, cada una abordando un aspecto clave del estudio.

Primero, se analizan las coordenadas en el espacio-tiempo, permitiendo identificar patrones característicos en los sujetos con ELA a partir de la posición y el movimiento de los marcadores faciales en función del tiempo. Luego, se presentan las matrices de correlación calculadas entre cada par de marcadores en las tres dimensiones esféricas, lo que ayuda a comprender la relación entre los movimientos faciales.

Posteriormente, se detallan los hiperparámetros resultantes de *GridSearch*, mostrando los valores óptimos seleccionados para mejorar el rendimiento de los modelos de clasificación. Finalmente, se realiza una comparación de algoritmos de clasificación, evaluando su desempeño en la diferenciación entre sujetos con ELA y controles, destacando la eficacia de cada enfoque.

Esta estructura permite una visión integral del análisis, desde la representación y caracterización del movimiento facial hasta la aplicación de técnicas de clasificación para la detección de la ELA.

4.1. Coordenadas en el espacio-tiempo

En este análisis, se estudiaron los movimientos faciales asociados a la tarea de habla repetitiva "pataka", evaluando tres parámetros principales: r (distancia radial), φ y θ . Estos parámetros fueron analizados a través de gráficos tridimensionales que muestran su evolución con respecto al número de *frames* y los marcadores faciales seleccionados.

La coordenada r , que representa la distancia de los marcadores faciales a un punto de referencia, se observó que en casi todos los gráficos, los valores de r muestran una tendencia descendente suave o una estabilización a lo largo del tiempo, presentan patrones que parecen tener patrones rítmicos, especialmente en los sujetos sanos, ya que realizaron movimientos cíclicos. En la Figura 12 vemos de lado izquierdo coordenadas en el espacio-tiempo de un sujeto con ELA y de lado derecho de un sujeto control sano; vemos que ambas gráficas muestran una tendencia descendente en general en r , acompañada de oscilaciones periódicas claras, en la segunda gráfica, las oscilaciones de r son más amplias, lo que sugiere que el sujeto sano tiene movimientos faciales más marcados que el sujeto con ELA.

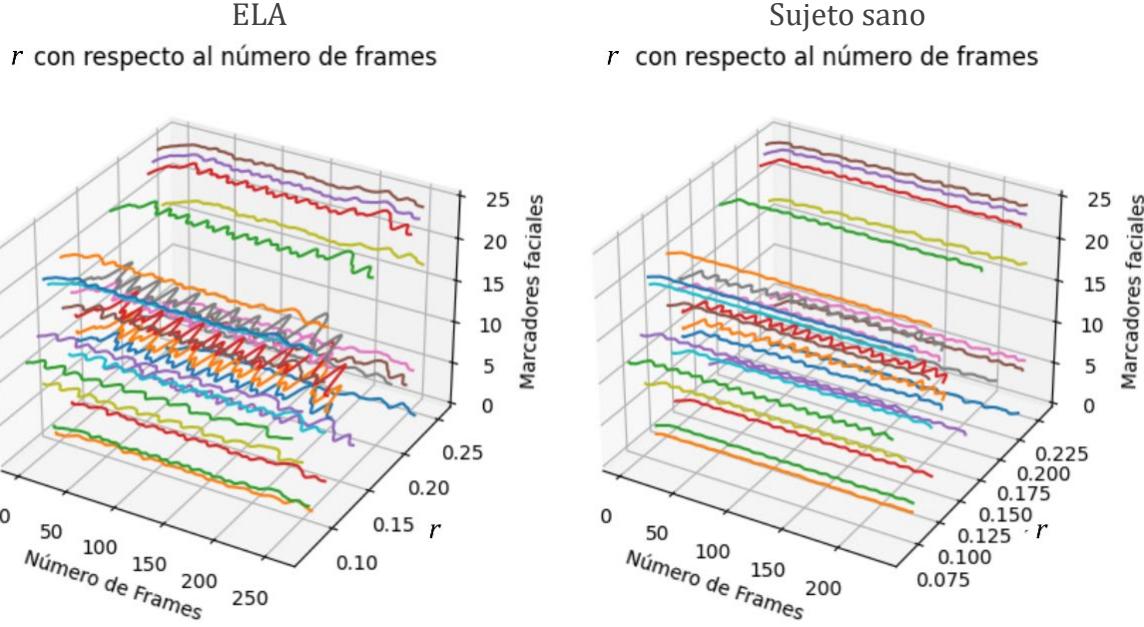


Figura 12: Coordenadas en el espacio-tiempo de r en sujeto con ELA (izquierda) y sujeto sano control (derecha).

La coordenada φ mide las oscilaciones angulares en el plano vertical, reflejando inclinaciones asociadas con movimientos de apertura y cierre de la mandíbula, las variaciones fueron más amplias en algunos marcadores faciales y, en algunos gráficos, se detectaron transiciones bruscas, posiblemente asociadas con movimientos rápidos o el hecho de la estimación de la coordenada z . En la Figura 13, vemos de lado izquierdo coordenadas en el espacio-tiempo de un sujeto con ELA y de lado derecho de un sujeto control sano; vemos que el sujeto con la enfermedad presenta menor amplitud, mientras que de lado derecho vemos más fluctuaciones lo que sugiere una mayor participación global del rostro en el movimiento.

Para la coordenada θ la cual es correspondiente al ángulo azimutal esto se refleja en los movimientos laterales significativos, como lo son los movimientos de la mandíbula; en algunos gráficos, las oscilaciones son menos regulares, lo que podría indicar una menor coordinación o fluctuaciones en la intensidad del movimiento. En la Figura 14, vemos de lado izquierdo coordenadas en el espacio-tiempo de un sujeto con ELA y de lado derecho de un sujeto control sano; vemos que el sujeto del lado derecho mostró oscilaciones en un rango más amplio y dinámico que el sujeto con ELA, con cambios más abruptos en ciertos marcadores faciales.

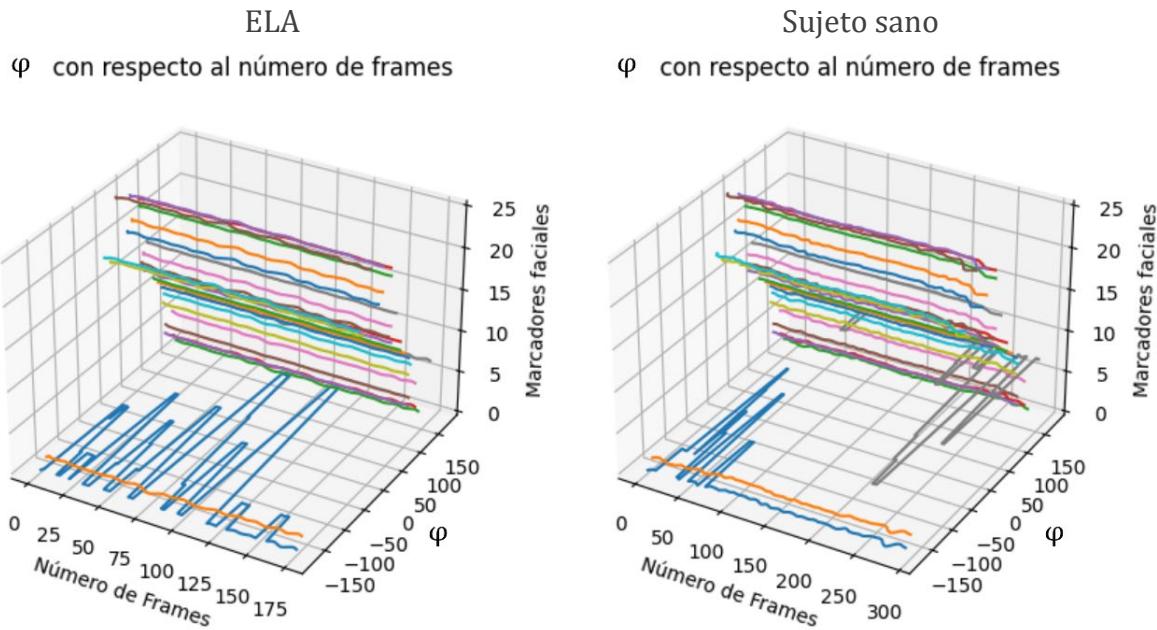


Figura 13: Coordenadas en el espacio-tiempo de φ en sujeto con ELA (izquierda) y sujeto sano control (derecha).

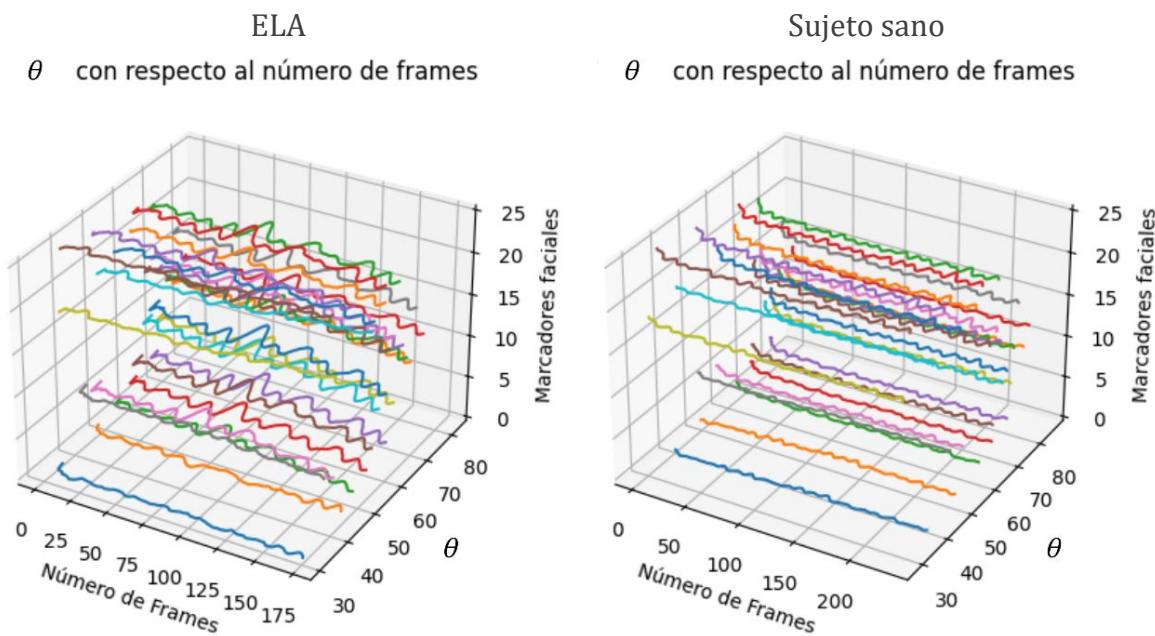


Figura 14: Coordenadas en el espacio-tiempo de θ en sujeto con ELA (izquierda) y sujeto sano control (derecha).

4.2. Matrices de correlación del comportamiento de los marcadores faciales

De cada uno de los $v = 18$ videos (uno por cada sujeto), se obtuvieron tres matrices de correlación, r , θ y φ , asociadas a cada tipo de parámetro esférico: r , θ y φ , como se muestra en la Figura 15.

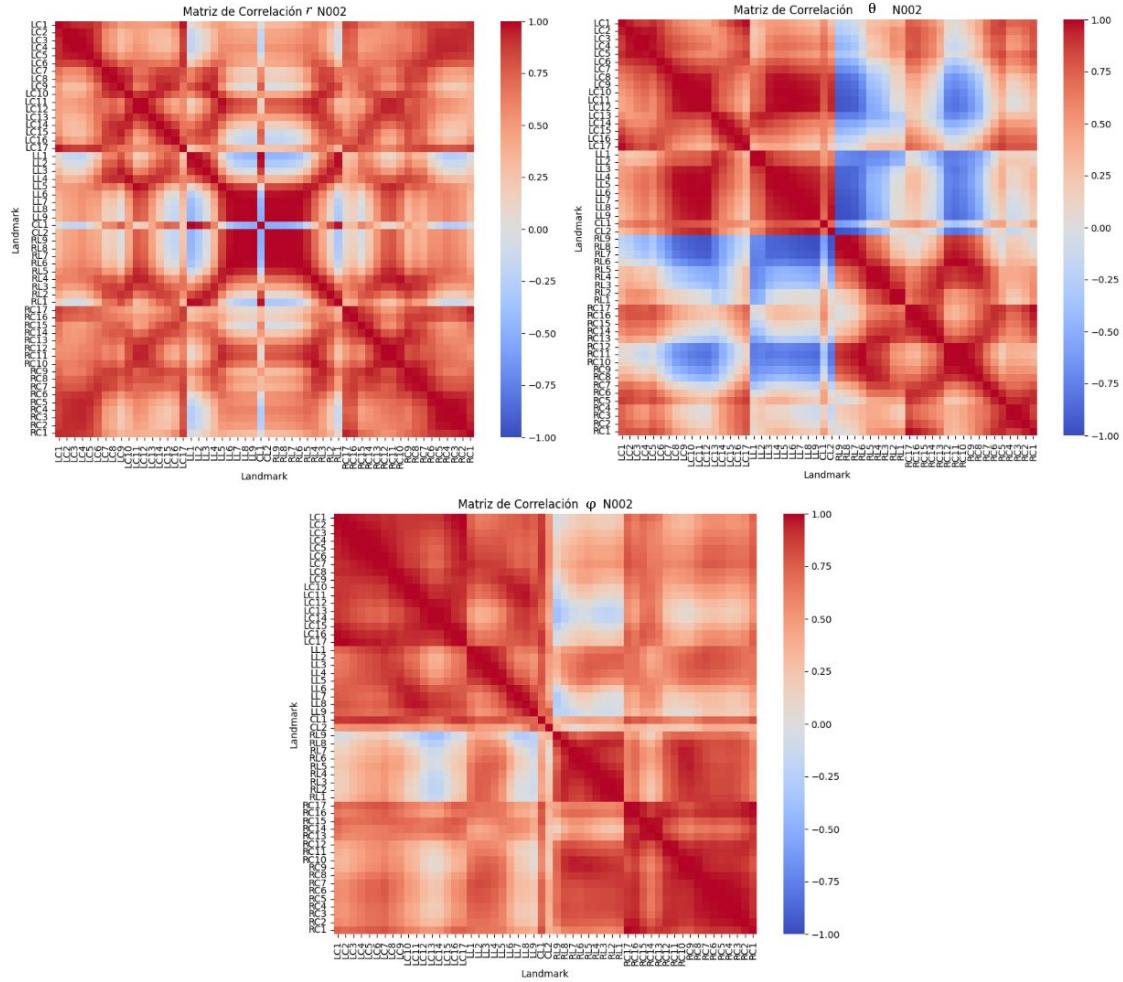


Figura 15. Matrices de correlación de Pearson de los desplazamientos de los marcadores faciales en la tarea /pataka/ para el participante N002(control sano).

La matriz Φ presenta una mayor consistencia en correlaciones positivas, lo que sugiere que varios puntos de referencia tienden a mantener movimientos sincronizados en esta dimensión, reflejando patrones estables durante la repetición de la frase. Aunque el análisis incluye información tridimensional, la profundidad real de los movimientos no está disponible. En su lugar, utilizamos estimaciones generadas mediante MediaPipe®, lo que podría introducir limitaciones en la precisión de los datos.

En el caso de la matriz Θ , las correlaciones negativas observadas indican que ciertos puntos de referencia tienden a moverse en direcciones opuestas en esta coordenada angular, posiblemente reflejando ajustes articulatorios o dinámicas estilísticas propias de la frase repetida [29].

Por otro lado, la matriz R exhibe una estructura bien definida de correlaciones positivas, evidenciando la consistencia en las distancias relativas entre puntos clave, lo que podría estar vinculado a la preservación de patrones rítmicos y estructurales durante las repeticiones [29].

Después de haber realizado la reducción de marcadores faciales, conservamos las mismas regiones, solo que sin redundancia de datos y así volvemos a calcular nuestras matrices de correlación, que a continuación podemos observar en la Figura 16.

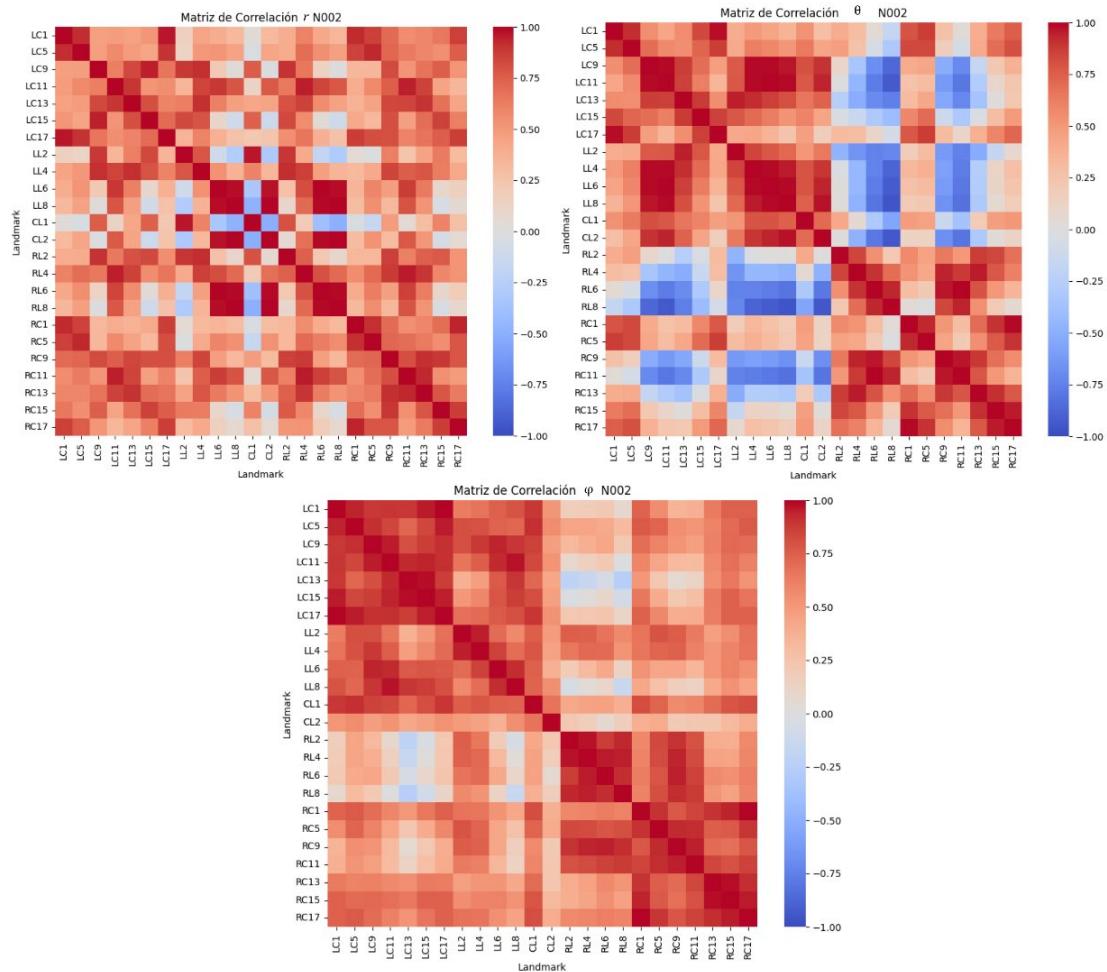


Figura 16. Matrices de correlación de Pearson de los desplazamientos de los marcadores faciales reducidos (24) en la tarea /pataka/ para el participante N002(control sano).

La matriz R (arriba izquierda) muestra una estructura clara de correlaciones positivas (rojo) entre varios puntos de referencia, lo que indica consistencia en las distancias relativas entre estos marcadores faciales durante la repetición de la frase. Algunas regiones muestran correlaciones débiles o negativas (azul), lo que podría reflejar variaciones en la manera en que se articulan ciertas partes de la frase. Estos patrones parecen mantener una simetría que sugiere estabilidad en la configuración espacial de los marcadores faciales durante el proceso.

La matriz Θ (arriba derecha) evidencia una mayor variabilidad en las correlaciones entre marcadores faciales, con zonas significativas de correlaciones negativas (azul), lo que indica que ciertos puntos de referencia tienden a moverse en direcciones opuestas en esta dimensión angular. Este comportamiento puede estar relacionado con ajustes específicos en la dinámica de pronunciación de la frase, como variaciones prosódicas o diferencias en el énfasis articulatorio [29]. Las correlaciones positivas en ciertas regiones también sugieren sincronía entre algunos grupos de marcadores faciales.

La matriz Φ (abajo) muestra una predominancia de correlaciones positivas, lo que indica movimientos más sincronizados en esta dimensión azimutal. Esto sugiere que, en esta dimensión, los marcadores faciales tienden a comportarse de manera más consistente entre repeticiones, lo que podría estar relacionado con patrones globales de articulación[29]. Sin embargo, hay pequeñas áreas de correlaciones negativas, lo que podría reflejar diferencias individuales o sutiles ajustes en ciertos puntos durante la repetición.

4.3. Hiperparámetros resultantes del *Gridsearch*

Este proceso evaluó sistemáticamente combinaciones de valores de hiperparámetros predefinidos para optimizar el desempeño de los modelos de AA. El objetivo principal de la optimización de hiperparámetros fue identificar la mejor configuración para cada modelo, garantizando alta precisión, generalización robusta y eficiencia computacional.

GridSearch utilizó validación cruzada para evaluar cada combinación de hiperparámetros dentro de la cuadrícula especificada. En este estudio, la métrica de evaluación fue la exactitud, y se empleó validación cruzada estratificada *k-fold* para garantizar una evaluación equilibrada entre las clases; las configuraciones resultantes fueron las siguientes y las podemos ver en la Tabla 4.

Tabla 4: Resultados de las métricas de validación aplicadas a los modelos de clasificación.

| | LOO | | | K = 3 | | | K = 6 | | |
|-----|---------|---------|---------|-----------|-----------|-----------|-----------|-----------|-----------|
| | R | Θ | Φ | R | Θ | Φ | R | Θ | Φ |
| kNN | auto | auto | auto | euclidian | euclidian | euclidian | euclidian | euclidian | euclidian |
| | 7 | 3 | 6 | 5 | 7 | 4 | 3 | 5 | 7 |
| | uniform | uniform | uniform | uniform | uniform | uniform | uniform | distance | uniform |
| RNA | tanh | tanh | tanh | relu | tanh | tanh | tanh | tanh | tanh |
| | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| | adam | adam | adam | adam | adam | lbfgs | adam | adam | lbfgs |
| RF | false | true | true | true | false | true | false | false | true |
| | gini | gini | gini | gini | gini | gini | gini | gini | gini |
| | none | none | none | none | none | none | none | none | none |
| | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 |
| | 2 | 10 | 5 | 10 | 2 | 2 | 10 | 10 | 2 |
| DT | 10 | 20 | 5 | 5 | 5 | 5 | 20 | 5 | 5 |
| | gini | gini | gini | gini | gini | gini | gini | gini | gini |
| | none | none | none | none | none | none | none | none | none |
| | 2 | 2 | 5 | 2 | 2 | 10 | 2 | 2 | 2 |
| | random | Random | best | random | random | best | random | best | best |

4.4. Implementación de algoritmos de Aprendizaje Automático

Posteriormente, se aplicaron varios modelos de clasificación para obtener los resultados de nuestro análisis (ver Tabla 5).

4.4.1. Comparación de Algoritmos de clasificación

En el modelo de kNN, la variable r mostró un rendimiento con una sensibilidad máxima del 60% y una exactitud de 55.5% al utilizar $k = 6$. Sin embargo, la especificidad se mantuvo baja (50%), lo que sugiere una tasa considerable de falsos positivos. La variable θ obtuvo sus mejores resultados con la validación LOO, alcanzando una precisión del 61.1%, sensibilidad del 60% y especificidad de 62.5%, destacando como la variable con mayor utilidad en este modelo. Por otra parte, la variable φ presentó un rendimiento deficiente en todas las configuraciones, con una precisión y sensibilidad en torno al 40% y una especificidad que rara vez superó el 37.5%, lo cual limita su efectividad para este modelo.

En el caso de los DT, el mejor desempeño se observó en la variable r con $k = 6$, donde se alcanzó una alta sensibilidad del 80%. Sin embargo, esta configuración mostró una precisión de 55.5% y una especificidad de solo 25%, lo que implica una tasa elevada de falsos positivos. La variable θ no arrojó resultados destacables en este modelo, con una precisión y sensibilidad bajas, alcanzando una especificidad máxima de 62.5% únicamente con LOO. La variable φ mostró resultados estables pero no muy altos, con una precisión de 50% y una especificidad de 62.5% en LOO, mientras que en otras configuraciones el rendimiento disminuyó significativamente.

Tabla 5: Resultados de las métricas de validación aplicadas a los modelos de clasificación.

| Validación | Variable | Algoritmo | Exactitud | Sensibilidad | Especificidad | |
|-------------------|-----------------|------------------|------------------|---------------------|----------------------|--|
| LOO | r | RNA | 77.78 | 44.44 | 33.33 | |
| k=3 | | | 72.2 | 66.67 | 72.2 | |
| k=6 | | | 44 | 41.67 | 33 | |
| LOO | θ | | 72.22 | 33.33 | 38.89 | |
| k=3 | | | 66.67 | 88.89 | 50 | |
| k=6 | | | 61.11 | 58.33 | 66.67 | |
| LOO | φ | | 38.89 | 16.67 | 22.22 | |
| k=3 | | | 50 | 61.11 | 38.89 | |
| k=6 | | | 50 | 58.33 | 41.67 | |
| LOO | r | KNN | 50 | 50 | 50 | |
| k=3 | | | 50 | 50 | 50 | |
| k=6 | | | 55.5 | 60 | 50 | |
| LOO | θ | | 61.1 | 60 | 62.5 | |
| k=3 | | | 55.5 | 60 | 50 | |
| k=6 | | | 27.7 | 20 | 37.5 | |
| LOO | φ | | 38.8 | 50 | 25 | |
| k=3 | | | 38.8 | 40 | 37.5 | |
| k=6 | | | 61.1 | 50 | 75 | |
| LOO | r | DT | 55.5 | 60 | 50 | |
| k=3 | | | 55.5 | 70 | 37.5 | |
| k=6 | | | 55.5 | 80 | 25 | |
| LOO | θ | | 50 | 50 | 50 | |
| k=3 | | | 38.8 | 30 | 50 | |
| k=6 | | | 44.4 | 40 | 50 | |
| LOO | φ | | 50 | 40 | 62.5 | |
| k=3 | | | 50 | 40 | 62.5 | |
| k=6 | | | 33.3 | 40 | 25 | |
| LOO | r | RF | 55 | 60 | 50 | |
| k=3 | | | 50 | 40 | 62.5 | |
| k=6 | | | 61.1 | 70 | 50 | |
| LOO | θ | | 44.4 | 30 | 62.5 | |
| k=3 | | | 44.4 | 50 | 37.5 | |
| k=6 | | | 50 | 50 | 50 | |
| LOO | φ | | 50 | 70 | 25 | |
| k=3 | | | 38.8 | 40 | 37.5 | |
| k=6 | | | 33.3 | 40 | 25 | |

El modelo de RF evidenció su mejor desempeño para la variable r con $k = 6$, alcanzando una precisión de 61.1% y una sensibilidad de 70%. Este resultado indica una adecuada capacidad de detección de verdaderos positivos, aunque la especificidad se mantuvo en 50%, sugiriendo algunos falsos positivos. La variable θ no logró un rendimiento destacable en ninguna configuración, con una precisión y sensibilidad máximas de 50% en LOO. La variable φ , por su parte, mostró una sensibilidad de 70% con LOO, aunque la especificidad fue baja (25%), indicando una alta tasa de falsos positivos, lo cual limita su aplicabilidad en este modelo.

En el caso de las RNA, la variable r obtuvo los mejores resultados con $k = 3$, alcanzando una precisión y especificidad de 72.2% y una sensibilidad de 66.67%, lo cual refleja un buen equilibrio entre las métricas y destaca la utilidad de esta variable en este modelo. La variable θ mostró una alta sensibilidad (88.89%) con $k = 3$, lo cual es ideal para aplicaciones donde la detección de verdaderos positivos es crucial. Sin embargo, la precisión de θ fue moderada (66.67%), y la especificidad alcanzó solo el 50%, lo que podría reflejar una tasa de falsos positivos considerable. Por último, la variable φ mostró un rendimiento bajo en todas las configuraciones, destacando una sensibilidad del 61.1% con $k = 3$ y $k = 6$, pero con una precisión y especificidad generalmente bajas, lo cual limita su aplicabilidad en RNA.

En general, el desempeño de los modelos varía significativamente dependiendo de la variable y el algoritmo utilizado. Para la variable r , el modelo RNA muestra una precisión más alta bajo la validación LOO (77.78%), pero esta disminuye conforme aumenta el valor de k , lo que sugiere que su rendimiento está influenciado por el tamaño de los vecinos utilizados. Por otro lado, los algoritmos KNN, DT y RF mantienen una precisión moderada alrededor del 55% en general, con una ligera mejora en $k=6$ para RF (61.1%).

En el caso de θ , los modelos muestran una mayor sensibilidad en general, especialmente con RNA ($k=3$), donde alcanza un 88.89%. Sin embargo, los valores de especificidad se mantienen más bajos, indicando una menor capacidad para identificar correctamente los casos negativos. Los modelos RF y DT muestran un comportamiento más balanceado, pero con menor precisión general en comparación con RAN.

Para φ , el rendimiento es más variable. Mientras que RNA alcanza una sensibilidad del 61.11% con $k=3$, los algoritmos RF y DT presentan valores bajos de especificidad en la mayoría de los casos, sugiriendo que tienen dificultades para distinguir entre las clases en esta dimensión. El KNN con $k=6$ es una excepción, alcanzando una especificidad del 75%.

Aunque los datos incluyen múltiples variables y algoritmos, la validación LOO parece destacar la capacidad máxima de los modelos, especialmente en términos de precisión para RNA con la variable r . Sin embargo, es importante considerar que estas métricas pueden no ser suficientes para determinar un modelo óptimo sin una evaluación adicional que contemple otros aspectos, como el balance entre sensibilidad y especificidad.

Los resultados obtenidos a través de diferentes modelos de clasificación muestran diferencias en el rendimiento de las variables r , θ y φ , así como en los algoritmos empleados.

4.4.2. Rendimiento de las coordenadas esféricas

En general, la variable r mostró un desempeño consistente en varios modelos, destacando en RNA y RF. En RNA con validación $k = 3$, r alcanzó un buen balance entre precisión, sensibilidad y especificidad, lo que indica que esta variable presenta una capacidad notable para capturar patrones relevantes en los datos. Este resultado sugiere que r contiene información importante que puede ayudar a los modelos a distinguir entre ambas clases.

La variable θ mostró un rendimiento destacado en términos de sensibilidad en RAN, especialmente con $k = 3$, donde alcanzó una sensibilidad de 88.89%. Este resultado indica que θ es particularmente útil para maximizar la detección de verdaderos positivos, lo cual puede ser beneficioso en contextos en los que la identificación de todos los casos positivos es prioritaria. Sin embargo, su especificidad y precisión fueron moderadas, lo que implica una tasa considerable de falsos positivos. Esto sugiere que, aunque θ aporta información valiosa, su capacidad para distinguir entre clases con precisión es limitada y puede requerir ajustes o combinaciones con otras variables para mejorar su rendimiento global.

La variable φ mostró un desempeño bajo en todos los modelos y configuraciones, con valores de precisión y especificidad consistentemente bajos. Aunque en algunos casos, como en RF y RNA, se observó una sensibilidad aceptable, la baja especificidad sugiere una alta tasa de falsos positivos. Estos resultados indican que φ aporta poca información relevante para la discriminación de clases en este contexto, lo cual podría deberse a una falta de relación significativa con el objetivo de clasificación o a la presencia de ruido en los datos que afecta negativamente su utilidad.

4.5. Discusión

Los resultados de φ plantean dudas sobre su relevancia en este contexto. La baja utilidad de φ en todos los modelos sugiere que podría ser prescindible o que requiere un preprocesamiento adicional para mejorar su utilidad en el modelo de clasificación.

En el análisis de los resultados obtenidos en este estudio y su comparación con trabajos previos en el campo, como los estudios de Bandini [6] y Gomes [9], surgen diversas similitudes y diferencias importantes en los métodos y hallazgos para la detección de ELA mediante el análisis de movimientos faciales.

El estudio de Bandini [6] se usó un enfoque basado en el análisis de video de movimientos faciales de pacientes con ELA, utilizando características cinemáticas (como rango de movimiento facial, velocidad de movimiento, simetría y forma) y logrando una precisión de hasta 88.9% en ciertas tareas como la repetición de frases (“Buy Bobby a puppy”). Utilizaron técnicas de clasificación como la regresión logística y SVM, y emplearon una cámara de profundidad para captar movimientos en tres dimensiones, lo cual les permitió obtener un conjunto de datos robusto en cuanto a las características faciales relevantes para la detección de ELA.

En comparación, nuestro estudio utilizó múltiples modelos de clasificación, incluyendo RNA, kNN y RF, para evaluar diferentes variables. Aunque no se usaron características tridimensionales debido a limitaciones en los datos usado, los resultados muestran que el modelo de redes neuronales fue capaz de obtener una alta precisión en la clasificación en las variables r y θ , indicando una capacidad efectiva para discriminar entre patrones asociados a ELA y controles sanos.

Gomes [9] introdujeron el uso de GNN para la identificación de ELA mediante marcadores faciales en el conjunto de datos Toronto *Neuroface*. Este enfoque permite representar relaciones espaciales entre marcadores faciales a través de grafos, capturando patrones de movimiento en tareas de habla y no habla. Ellos destacaron que en su modelo se alcanzó una precisión superior al 80% en tareas de gesticulación como “SPREAD” (sonreír con labios cerrados), donde observaron que la sensibilidad de los músculos labiales era crítica para la detección de ELA. Esta técnica innovadora de representación gráfica mejora la captura de movimientos faciales característicos de la ELA.

Ambos estudios comparativos ofrecen ideas relevantes para mejorar los métodos actuales de detección de ELA basados en movimientos faciales. Bandini *et al.* [6] demostraron la efectividad de las características cinemáticas en tareas específicas, mientras que Gomes *et al.* [9] introdujeron GNN para capturar patrones espaciales complejos en el rostro.

Capítulo 5. Conclusiones

Este estudio se centró en el desarrollo de un modelo de AA para la detección de ELA mediante el análisis de imágenes faciales. A diferencia de estudios previos, esta investigación utilizó imágenes faciales bidimensionales y estimaciones de la dimensión z generadas a través de la biblioteca MediaPipe®, junto con modelos de AA relativamente simples. A pesar de estas limitaciones, los resultados obtenidos son prometedores, especialmente en ciertas dimensiones, destacando la utilidad del enfoque adoptado.

Las coordenadas r y θ demostraron un desempeño superior en comparación con la coordenada φ , que presentó resultados significativamente inferiores. En el caso de r , las RNA alcanzaron una precisión del 72.2%, una especificidad del 72.2% y una sensibilidad del 66.67% cuando se utilizó $k=3$. Estos resultados subrayan la capacidad del modelo para identificar casos de ELA de manera confiable, manteniendo al mismo tiempo una adecuada capacidad para distinguir entre controles sanos. Por su parte, la dimensión θ se destacó por su alta sensibilidad, alcanzando un 88.89% con RNA bajo la misma configuración. Esto indica que esta dimensión es especialmente útil para maximizar la detección de verdaderos positivos, aunque presenta un equilibrio menos favorable en términos de especificidad. En contraste, la coordenada φ tuvo un desempeño bajo en todos los modelos, con baja precisión y especificidad, lo que podría atribuirse a las limitaciones inherentes de la estimación de la dimensión z , generando datos menos representativos y ruidosos.

Este trabajo resalta la importancia de una selección cuidadosa de las dimensiones y métodos de preprocesamiento para optimizar el rendimiento de los modelos de AA. El proceso de transformación a coordenadas esféricas, combinado con un análisis de correlaciones para la reducción de marcadores faciales redundantes, resultó clave para mejorar la calidad de los datos y reducir la complejidad del modelo. Asimismo, los resultados obtenidos evidencian que las RNA son la metodología más prometedora dentro de los algoritmos evaluados, debido a su capacidad para modelar patrones complejos y generalizar mejor frente a un conjunto de datos limitado.

5.1. Limitaciones

Aunque los resultados son alentadores, este estudio presenta ciertas limitaciones que deben ser abordadas en investigaciones futuras. Este análisis presenta algunas limitaciones que deben ser consideradas. La elección de valores específicos de k y el uso de LOO en algunos modelos pueden no capturar todas las variaciones posibles en el rendimiento. Asimismo, la falta de características tridimensionales completas

pudo limitar la capacidad de los modelos para capturar patrones cinemáticos más detallados, como los observados en estudios previos que emplearon cámaras de profundidad. Finalmente, el tamaño relativamente pequeño del conjunto de datos puede haber afectado la capacidad de generalización de algunos modelos.

5.2. Trabajo futuro

Para futuros estudios, se recomienda explorar configuraciones más complejas de RNA, como arquitecturas profundas, y considerar el uso de técnicas de preprocesamiento para la variable φ .

Este enfoque establece una base sólida para futuras aplicaciones en la detección de ELA, con el potencial de mejorar la precisión y rapidez en el diagnóstico clínico. Investigaciones futuras podrían explorar arquitecturas de RNA más avanzadas, como modelos profundos, para maximizar la capacidad predictiva. Además, sería beneficioso considerar el uso de técnicas innovadoras como RNA basadas en grafos para capturar relaciones espaciales complejas entre los marcadores faciales. La integración de tareas de habla y no habla, audios y la combinación de ambas fuentes de datos, junto con métodos más avanzados de adquisición de datos, podría optimizar aún más el rendimiento de los modelos. Estas mejoras podrían sentar las bases para herramientas diagnósticas más precisas, rápidas y accesibles en el ámbito clínico.

5.3. Productos derivados

A continuación se presentan los productos derivados de este trabajo.

- Presentación del trabajo “Amyotrophic Lateral Sclerosis Detection Using Facial Symmetry Analysis with Machine Learning Techniques” en el XLVII Congreso Nacional de Ingeniería Biomédica (CNIB) 2024, ver Anexo 1.
- Publicación del artículo “Suárez-Hernández, D., Santos-Arce, S. R., Torres-Ramos, S., Salido-Ruiz, R. A., & Román-Godínez, I. (2025). Amyotrophic Lateral Sclerosis Detection Using Facial Symmetry Analysis with Machine Learning Techniques. In *IFMBE proceedings* (pp. 239–248). https://doi.org/10.1007/978-3-031-82123-3_23, ver Anexo 2.
- Redacción del artículo científico “Automatic Detection of Amyotrophic Lateral Sclerosis Through Facial Image Analysis and Machine Learning” que contiene los resultados de esta tesis, y será enviado a revista indexada.

6. Anexos

Anexo 1



Anexo 2



Amyotrophic Lateral Sclerosis Detection Using Facial Symmetry Analysis with Machine Learning Techniques

Daniela Suárez-Hernández¹ , Stewart R. Santos-Arce¹ ,
Sulema Torres-Ramos¹ , Ricardo A. Salido-Ruiz¹ ,
and Israel Román-Godínez¹

¹ Division of Technologies for Cyber-Human Integration, University of Guadalajara,
Guadalajara, Jalisco, Mexico
daniela.suarez2921@alumnos.udg.mx

Abstract. Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder that causes muscle paralysis, visible fasciculations, speech difficulties (dysarthria), and tongue atrophy. 50% of ALS-affected individuals die within 30 months since diagnosis, and about 20% survive 5–10 years. In Mexico, the survival average is 68.6 months. Diagnosis of ALS is often delayed due to minimal initial symptoms and outdated diagnostic protocols, reducing the time available for symptom management. Current studies focus on using biomedical signals and machine learning (ML) to aid in diagnosing ALS. This work proposes to identify ALS through facial-symmetry analysis using simple supervised ML techniques. Using the Toronto NeuroFace dataset, key facial landmarks were extracted and converted to spherical-coordinates to minimize spatial error. Pearson correlation matrices were used to reduce the landmark numbers to 24, simplifying data processing. ML models, were evaluated using accuracy, specificity, and sensitivity metrics. Our findings underline the importance of choosing appropriate coordinates and an effective methodology for data acquisition, demonstrating that high throughput can be achieved with more accessible equipment. This work presents 66.7%, 50.0% and 87.5% in accuracy, specificity and sensitivity; respectively. This methodology has the potential to contribute to earlier and more accurate detection of ALS, facilitating a more accessible and efficient diagnosis.

Keywords: Amyotrophic Lateral Sclerosis · Machine Learning · Facial Landmarks · Facial Symmetry

1 Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder characterized by progressive muscle paralysis, leading to weakness in the affected muscles and visible fasciculations [1]. Typical symptoms include speech difficulties known as dysarthria, weakness in the lower half of the face, involuntary muscle contractions called

José de Jesús Agustín Flores Cuautle ·
Balam Benítez-Mata · José Javier Reyes-Lagos ·
Humiko Yahaira Hernandez Acosta · Gerardo Ames Lastra ·
Esmeralda Zuñiga-Aguilar · Edgar Del Hierro-Gutierrez ·
Ricardo Antonio Salido-Ruiz *Editors*

XLVII Mexican Conference on Biomedical Engineering

Proceedings of CNIB 2024, November 7–9, 2024,
Hermosillo, Sonora, México - Volume 1: Signal
Processing And Bioinformatics



MOREMEDIA 

 Springer

Referencias bibliográficas

- [1] L. C. Wijesekera y P. N. Leigh, "Amyotrophic lateral sclerosis", *Orphanet J Rare Dis*, vol. 4, núm. 1, 2009, doi: 10.1186/1750-1172-4-3.
- [2] K. Talbot, "Motor neuron disease", *Pract Neurol*, vol. 9, núm. 5, pp. 303–309, oct. 2009, doi: 10.1136/jnnp.2009.188151.
- [3] P. Masrori y P. Van Damme, "Amyotrophic lateral sclerosis: a clinical review", *Eur J Neurol*, vol. 27, núm. 10, pp. 1918–1929, oct. 2020, doi: 10.1111/ene.14393.
- [4] I. Rodriguez-Leyva, "ALS Policy in Mexico", en *Public Policy in ALS/MND Care*, Singapore: Springer Singapore, 2021, pp. 191–199. doi: 10.1007/978-981-15-5840-5_13.
- [5] L. Ewis, P. R. Owland, y E. A. S. Hneider, "Medical Progress 1688 .", 2001. [En línea]. Disponible en: www.nejm.org
- [6] A. Bandini, J. R. Green, B. Taati, S. Orlandi, L. Zinman, y Y. Yunusova, "Automatic detection of amyotrophic lateral sclerosis (ALS) from video-based analysis of facial movements: Speech and non-speech tasks", en *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, Institute of Electrical and Electronics Engineers Inc., jun. 2018, pp. 150–157. doi: 10.1109/FG.2018.00031.
- [7] A. Bandini *et al.*, "A New Dataset for Facial Motion Analysis in Individuals with Neurological Disorders", *IEEE J Biomed Health Inform*, vol. 25, núm. 4, pp. 1111–1119, abr. 2021, doi: 10.1109/JBHI.2020.3019242.
- [8] B. Jin, Y. Qu, L. Zhang, y Z. Gao, "Diagnosing parkinson disease through facial expression recognition: Video analysis", *J Med Internet Res*, vol. 22, núm. 7, jul. 2020, doi: 10.2196/18697.
- [9] N. B. Gomes, A. Yoshida, M. Roder, G. C. de Oliveira, y J. P. Papa, "Facial Point Graphs for Amyotrophic Lateral Sclerosis Identification", jul. 2023, [En línea]. Disponible en: <http://arxiv.org/abs/2307.12159>
- [10] Revista Ecuatoriana de Neurologia. (2015, June 2). *Esclerosis Lateral Amiotrófica: Criterios de El Escorial y la Electromiografía en su Temprano Diagnóstico. - revecuatneurol - Revista Ecuatoriana de Neurología*. Revecatneurol - Revista Ecuatoriana De Neurología. https://revecuatneurol.com/magazine_issue_article/esclerosis-lateral-amiotrofica-criterios-en-temprano-diagnostico/

- [11] Consejo Mexicano de Neurología a.c, “Neurólogos en México Certificados”. (n.d.). <https://www.consejomexicanodeneurologia.org/neurologos-certificados.html>
- [12] De Estadística Y Geografía, I. N. (n.d.). *Población*. <https://www.inegi.org.mx/temas/estructura/>
- [13] V. Grollemund *et al.*, “Machine learning in amyotrophic lateral sclerosis: Achievements, pitfalls, and future directions”, el 1 de febrero de 2019, *Frontiers Media S.A.* doi: 10.3389/fnins.2019.00135.
- [14] M. R. Héctor, P.-G. Juan Didier, M. María Elena, G.-G. María Teresa, y M.-C. E. Jorge, “Esclerosis lateral amiotrófica. Contribución de la Neurología Mexicana”, 2014.
- [15] A. Leiva, F. González, R. Fernandez, M. Ostolaza, y M. Sívori, “Artículo original escala revisada de valoración funcional de esclerosis lateral amiotrófica: adaptación transcultural”.
- [16] K. M. Allison, I. S. B. Nip, y P. Rong, “Use of Automated Kinematic Diadochokinesis Analysis to Identify Potential Indicators of Speech Motor Involvement in Children With Cerebral Palsy”, *Am J Speech Lang Pathol*, vol. 31, núm. 6, pp. 2835–2846, nov. 2022, doi: 10.1044/2022_AJSLP-21-00241.
- [17] K. K. Makam, V. K. Singh, y R. B. Pachori, “ALS Detection Framework Based on Automatic Singular Spectrum Analysis and Quantum Convolutional Neural Network From EMG Signals”, *IEEE Sens Lett*, vol. 8, núm. 9, pp. 1–4, sep. 2024, doi: 10.1109/LSENS.2024.3449369.
- [18] L. Simmatis, J. Robin, M. Spilka, y Y. Yunusova, “Detecting bulbar amyotrophic lateral sclerosis (ALS) using automatic acoustic analysis”, el 6 de septiembre de 2023. doi: 10.21203/rs.3.rs-3306951/v1.
- [19] A. C. Correa, A. E. S. Jimenez, y F. A. P. Ortiz, “Reconocimiento de rostros y gestos faciales mediante un análisis de relevancia con imágenes 3D”, *REVISTA DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN*, vol. 4, núm. 1, p. 7, dic. 2013, doi: 10.19053/20278306.2563.
- [20] Google, “Guía de detección de puntos de referencia faciales”, https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker?hl=es-419.
- [21] R. L. Kolar, W. G. Gray, J. J. Westerink, y R. A. Luettich, “Shallow water modeling in spherical coordinates: equation formulation, numerical implementation, and application”, *Journal of Hydraulic Research*, vol. 32, núm. 1, pp. 3–24, ene. 1994, doi: 10.1080/00221689409498786.

- [22] P. Sedgwick, "Pearson's correlation coefficient", *BMJ*, vol. 345, núm. jul04 1, pp. e4483–e4483, jul. 2012, doi: 10.1136/bmj.e4483.
- [23] C. Croux y C. Dehon, "Influence functions of the Spearman and Kendall correlation measures", *Stat Methods Appt*, vol. 19, núm. 4, pp. 497–515, nov. 2010, doi: 10.1007/s10260-010-0142-z.
- [24] Y. LeCun, Y. Bengio, y G. Hinton, "Deep learning", *Nature*, vol. 521, núm. 7553, pp. 436–444, may 2015, doi: 10.1038/nature14539.
- [25] G. Litjens *et al.*, "A survey on deep learning in medical image analysis", *Med Image Anal*, vol. 42, pp. 60–88, dic. 2017, doi: 10.1016/j.media.2017.07.005.
- [26] M. Sokolova y G. Lapalme, "A systematic analysis of performance measures for classification tasks", *Inf Process Manag*, vol. 45, núm. 4, pp. 427–437, jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
- [27] B. Mahesh, "Machine Learning Algorithms - A Review", *International Journal of Science and Research (IJSR)*, vol. 9, núm. 1, pp. 381–386, ene. 2020, doi: 10.21275/art20203995.
- [28] P. Liashchynskyi y P. Liashchynskyi, "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS", dic. 2019.
- [29] H. Wu, J. Jia, H. Wang, Y. Dou, C. Duan, y Q. Deng, "Imitating Arbitrary Talking Style for Realistic Audio-Driven Talking Face Synthesis", en *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*, Association for Computing Machinery, Inc, oct. 2021, pp. 1478–1486. doi: 10.1145/3474085.3475280.