

2025
Detección de mala postura mediante
visión artificial utilizando redes
neuronales convolucionales

Perla Pérez Carrasco



Universidad Autónoma de
Querétaro

Facultad de Ingeniería

**Detección de mala postura mediante
visión artificial utilizando redes
neuronales convolucionales**

Tesis

Que como parte de los requisitos para obtener el
Grado de

Maestro en

Ciencias en Inteligencia Artificial

Presenta

Ing. Perla Pérez Carrasco

Dirigido por:
Dr. Jesús Carlos Pedraza Ortega

Co-Director:
MCIA Osmar Antonio Espinosa Bernal

Querétaro, Qro. a 29 de septiembre de 2025

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciatario no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciatario.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



**Universidad Autónoma de
Querétaro**
Facultad de Ingeniería
**Maestría en Ciencias en Inteligencia
Artificial**

**Detección de mala postura mediante visión artificial
utilizando redes neuronales convolucionales**

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta

Ing. Perla Pérez Carrasco

Dirigido por

Dr. Jesús Carlos Pedraza Ortega

Dr. Jesús Carlos Pedraza Ortega
Presidente

MCIA Osmar Antonio Espinosa Bernal
Secretario

Dr. Saúl Tovar Arriaga
Vocal

Dr. Leonardo Barriga Rodríguez
Suplente

MCIA Edgar Rodrigo López Silva
Suplente

Centro Universitario, Querétaro, Qro. México
Fecha de aprobación por el Consejo Universitario (Septiembre 2025)

Dedicatorias

Dedico este trabajo a mis padres y hermanos, por su apoyo incondicional y confianza en mí durante todo el proceso.

A mis profesores, quienes me brindaron las herramientas y los conocimientos necesarios para alcanzar esta meta.

Y a todos aquellos que, de una forma u otra, hicieron posible la realización de este proyecto.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a mi asesor el Dr. Jesús Carlos Pedraza Ortega y a mi co-asesor MCIA Osmar Antonio Espinosa Bernal por su guía, conocimiento y paciencia a lo largo de esta investigación que fueron claves para el desarrollo de este proyecto.

Agradezco también a los miembros del sínodo por sus valiosas observaciones y apoyo durante el proceso de revisión.

Agradezco el apoyo brindado por la Universidad Autónoma de Querétaro, que proporcionaron el acceso a materiales, infraestructura y orientación técnica.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el financiamiento brindado para llevar a cabo este trabajo.

Finalmente, agradezco a cada persona que, desde su posición, contribuyó a enriquecer esta experiencia académica y personal.

A mi familia por su comprensión y respaldo permanente.

Abreviaturas y siglas

IA Inteligencia Artificial

CNN Red Neuronal Convolucional (Convolutional Neural Network)

ResNet Red Neuronal Residual

VGG Visual Geometry Group

GEI Imágenes de energía de la marcha (Gait Energy Image)

IM2D Coordenadas 2D basadas en imagen

HC3D Coordenadas 3D centradas en el ser humano

LBP Local Binary Patterns

GLCM Gray-Level Co-occurrence Matrix

CCA Análisis de correlación canónica (Canonical Correlation Analysis)

LDA Análisis discriminante lineal (Linear Discriminant Analysis)

CCR Tasa de clasificación correcta

PCA Análisis de Componentes Principales (Principal Component Analysis)

COM Centro de Masa (Center of Mass)

CVS Valores separado por comas (comma separated values)

ReLU Rectified Linear Unit

RMSprop Root Mean Square Propagation

Adam Adaptive Moment Estimation

SI Simetría

Resumen

La locomoción humana es un rasgo físico y conductual único para cada persona, rasgo que se considera como un patrón biométrico debido a la singularidad con la que la marcha manifiesta el estilo de caminar de una persona. Debido a la estructura musculoesquelética única, permite estudios del patrón de marcha o de la mala postura basado en imágenes o en esqueletos. Puesto que, entre el al 20 % de los adultos sufren lumbalgia entre los 30 y 50 años y del 70 al 80 % pertenecen a una población menor de 45 años, lo que da pauta para considerar estos índices como causas para una posible mala postura al caminar o al estar de pie. La extracción precisa de partes del cuerpo y ángulos de las articulaciones en imágenes reales es una tarea tediosa, ya que, el movimiento humano no rígido abarca una amplia gama de posibles transformaciones y además la falta de métricas para determinar una mala postura mediante la pose muestran un reto computacional y de interpretación de datos. Analizando el ciclo de marcha mediante el uso de técnicas de procesamiento digital de imágenes y Deep Learning para detectar anomalías en la forma de caminar, implementando técnicas de extracción de características. Este estudio utiliza el dataset de OU-ISIR MVLPI que este compuesto por imágenes y coordenadas de la imagen para crear el esqueleto de la persona. La metodología se centra en analizar la extracción de características de las imágenes y de los esqueletos según las métricas del tipo de datos. Las métricas obtenidas de los esqueletos se destacan por sus valores para analizar la mala postura a través de un análisis estadístico en el que se evalúan los datos de anomalías según la simetría de las articulaciones y del tipo de ángulo. Donde se obtiene un 73.2 % de datos con buena postura y un 26.8 % presentan alguna anomalía de datos que da lugar a una mala postura.

Palabras clave: Extracción de características, Redes Neuronales Convolucionales, Ciclo de marcha, Postura.

Abstract

Human locomotion is a physical and behavioral trait unique to each person and is considered a biometric pattern due to the uniqueness with which gait manifests a person's walking style. Due to its unique musculoskeletal structure, it allows for studies of gait patterns or poor posture based on images or skeletons. Given that 20 % of adults suffer from low back pain between 30 and 50 years of age and 70 to 80 % belong to a population younger than 45 years of age, it can be deduced that these indices may be indicative of poor posture when walking or standing. Extracting body parts and joint angles from real images accurately is a tedious task. This is because non-rigid human motion encompasses a wide range of possible transformations and the lack of metrics for determining poor posture through pose presents a computational and data interpretation challenge. The gait cycle is analyzed using digital image processing and deep learning techniques. These techniques detect gait anomalies by implementing feature extraction techniques. This study uses the OU-ISIR MVLP dataset, which is composed of images and image coordinates, to create the person's skeleton. The methodology focuses on analyzing the feature extraction from images and skeletons. It does this according to the metrics of the data type. The metrics obtained from the skeletons are notable for their ability to analyze poor posture through statistical analysis, evaluating anomaly data according to joint symmetry and angle type. The presence of good posture is indicated in 73.2 % of cases, while 26.8 % show some data anomaly resulting in poor posture.

Keywords: Feature extraction, Convolutional Neural Networks, Gait cycle, Posture.

Índice

Índice de cuadros	9
Índice de figuras	10
1 Introducción	2
2 Antecedentes	4
2.1 Anatomía del sistema locomotor	4
2.2 Planos anatómicos	5
2.3 Biomecánica	5
2.4 Ciclo de la marcha	6
2.5 Fase de apoyo (carga de peso)	7
2.6 Fase oscilante	7
2.7 Anomalías de la postura	8
2.8 Biometría de la marcha	9
2.9 Estimación de la pose	10
2.9.1 Estimadores de pose	10
2.10 Imágenes de la marcha	11
2.10.1 Imágenes de energía de la marcha (GEI)	12
2.11 Datasets	13
2.12 Extracción de características	13
2.12.1 Características espaciales	15
2.12.2 Características temporales	15
2.13 Inteligencia Artificial	16
2.13.1 Redes Neuronales Convolucionales	17
2.13.2 Visión artificial	22
2.13.3 Métricas	23
2.13.4 Técnicas para extracción de características	24
3 Hipótesis	26
4 Objetivos	27
4.1 Objetivo general	27
4.2 Objetivo específicos	27
5 Metodología	28
5.1 Dataset	28
5.1.1 Adquisición de imágenes	28
5.2 Acondicionamiento del dataset	29
5.3 Extracción de características	31
5.3.1 Métodos manuales	32
5.3.2 Método basado en texturas	37

5.4	Redes Neuronales Convolucionales	39
5.4.1	VGG16	40
5.4.2	Inception V3	42
5.4.3	ResNet50	44
5.5	Estudio de ablación	46
5.6	Estructuración de datos de coordenadas	48
5.6.1	Verificación de datos	48
5.6.2	Identificador de los individuos	49
5.6.3	Análisis de simetría	50
5.7	Análisis estadístico	57
5.7.1	Métricas	58
6	Resultados y discusión	61
6.1	Resultados	61
6.1.1	Extracción de características	61
6.1.2	Clasificación	67
6.1.3	Detección de anomalías de datos para la detección de una mala postura	70
6.2	Discusión	74
7	Conclusiones	78
7.1	Limitaciones	78
7.2	Trabajo a futuro	79
Referencias bibliográficas		80
A	Anexos	87
A.1	Figuras complementarias	87
A.2	Documentos	90

Índice de cuadros

1	Data sets más usados para el análisis de la marcha.	14
2	Hiperparámetros utilizados por arquitectura.	47
3	Hiperparámetros utilizados por arquitectura de ResNet.	47
4	Codificación correspondiente a cada clase.	61
5	Métricas finales de la exactitud, pérdida y extracción de características de los modelos de clasificación empleados.	64
6	Reporte de clasificación de los modelos VGG16, ResNet50 e InceptionV3.	69
7	Cantidad de valores atípicos (outliers) por ángulo.	73

Índice de figuras

1	Sistema musculoesquelético. Obtenida de [1].	4
2	Planos anatómicos tomando como referencia el punto central del cuerpo conocido como centro de masa. Adaptada al español de [2].	6
3	Fases del ciclo de marcha. Adaptada al español de [3].	8
4	Ejemplos de imágenes de entrada para el reconocimiento y extracción de características de la marcha.	12
5	Secuencia de imágenes normalizadas e imagen GEI. Obtenida de [4]. . .	13
6	Esquema de aprendizaje supervisado, el cual genera un modelo que utiliza un conjunto de entrenamiento para predecir una variable objetivo, es decir, el conjunto de entrenamiento esta etiquetado. Recuperado de [5].	17
7	Esquema de aprendizaje no supervisado, consiste en entrenar un modelo que utiliza datos sin etiquetas para realizar un análisis descriptivo que puede ser asociaciones, detección de anomalías o agrupación. Imagen recuperada de [5].	17
8	Representación de la entrada tridimensional de una CNN. Imagen propia.	18
9	Arquitectura CNN sencilla compuesta por cinco capas, en la que se observa sus respectivos componentes. Adaptada al español de [6].	19
10	Arquitectura general del modelo VGG-16. Imagen propia.	20
11	Arquitectura general de InceptionV3. Imagen adaptada al español de [7].	21
12	Arquitectura general de ResNet50. Imagen propia.	22
13	Etapas de la visión artificial de acuerdo con Alonso y Jara [8]. Imagen mejorada de [8].	22
14	Matriz de confusión. Imagen propia.	23
15	Diagrama de flujo de la metodología general.	28
16	Posición de las cámaras para la captura de la base de datos.	29
17	Imágenes de la base de datos en diferentes puntos de vista (o ángulos).	30
18	Fotogramas para crear un ciclo de marcha con las imágenes de energía de la marcha (GEI) a partir del dataset de [9].	30
19	Modelo del esqueleto humano con 24 articulaciones. Imagen recuperada de [10].	31
20	Secuencia de los ángulos que conforman el ciclo de marcha con su imagen GEI y esqueleto respectivamente. Imagen propia.	31
21	Tipos de reescalamiento. Imagen propia.	33
22	Tipos de interpolaciones.Imagen propia.	33
23	Binarización en diferentes umbrales. Imagen propia.	34
24	Segmentación por segmentos y por partes.	35
25	Método de segmentación Chan-Vese. Imagen propia.	36
26	Detección de bordes en umbrales específicos. Imagen propia.	37
27	Matriz de coocurrencia de niveles de grises (GLCM). Imagen propia. . .	38
28	Histogramas. Imagen propia.	39
29	Diagrama de capas convolucionales de la arquitectura VGG16 modificado.	42

30	Diagrama de las capas de la arquitectura Inception V3.	44
31	Bases de datos generadas en CSV. Imágenes de autoría propia.	49
32	Metodología general para la construcción de bases de datos.	51
33	Metodología general del análisis estadístico.	58
34	Distribución por grupo de edad y sexo. Imagen de creación propia . . .	59
35	Frecuencia absoluta y relativa de la distribución de edad y sexo. Imágenes de autoría propia	59
36	Imagen de prueba y arquitectura para la extracción de características de la capa 5 de VGG16.	62
37	Extracción de características de los patrones de textura y combinaciones de características de la capa block5_conv1 de VGG16.	63
38	Gráficas de PCA de los modelos entrenados.	65
39	Gráficas de LDA de los modelos entrenados.	66
40	Matrices de confusión de los modelos de clasificación.	68
41	Evolución de la pérdida y precisión de los modelos de VGG16, ResNet50 e Inceptionv3 respectivamente.	70
42	Mapa de calor de desviación estándar.	71
43	Correlación de Pearson	72
44	Distribución de los valores atípicos.	73
45	Comparación de valores atípicos respecto al tipo de ángulo y el tipo de simetría.	74
46	Mapa de calor que compara las métricas de desviación estándar, varianza y correlación de Pearson entre dos variables: D H-BS (la distancia entre la cabeza y la base de la columna) y D legs (la distancia entre las piernas) de dos datasets diferentes. El primer dataset es de datos globales y el segundo es de solo una vista o fotograma de un individuo que es en el ángulo de 180°.	75
47	Matriz de confusión de la clasificación global de simetría para las clases de Buena y Anomalía.	77
48	Distribución global de la clasificación Buena vs Anomalía	77
A.1	Arquitectura ResNet50	87
A.2	Distribución de los tipos de simetría.	88
A.3	Estimación de la pose con estimadores de COCO y Mediapipe. Imágenes de creación propia.	89
A.4	Constancia de manejo de la lengua inglés.	90
A.5	Constancia de comprensión de textos en inglés.	91
A.6	Artículo presentado en el Decimosexto Congreso Internacional de Cómputo en Optimización y software (CICOS 2024).	92
A.7	Carta de aceptación para la publicación del artículo presentado en CICOS 2024.	93
A.8	Retribución social: clase a niños de primaria en el Programa Adopte un Talento (PAUTA).	94

1. Introducción

La marcha humana es uno de los rasgos físicos individuales y conductuales únicos de cada individuo, dando pauta a considerarse como un patrón biométrico debido a su singularidad debido que la marcha manifiesta el estilo de caminar de cada persona reconociendo su significado como autenticación de un individuo [11-14]. El estudio de la marcha se remonta desde la antigua Grecia y Roma, por Aristóteles al registrar las primeras observaciones de la caminata de los humanos y animales en su trabajo “On the Gait of Animals”. Posterior Borell, considerado como el padre de la biomecánica, en su trabajo póstumo mediante matemáticas relacionó las máquinas en los animales. Fue hasta Etienne-Jules Marey que desarrolló cámaras las cuales podían tomar múltiples fotos por segundo expuestas en el mismo negativo, conocida como cronofotografía [15].

A lo largo de los años, se ha estudiado el ciclo de marcha en ramas de la biomecánica, medicina, ingeniería y visión por computadora debido a sus patrones únicos con los que cuenta cada objeto de estudio, es decir, los humanos. Los objetos de estudio en diferentes investigaciones depende del enfoque y en la especificación del sector de estudio debido a que en la población mundial existen diversas enfermedades que llegan a afectar la forma de caminar de una persona o en su postura como menciona Zhang [16], que hace referencia a que el movimiento biomecánico del ser humano no solo se basa en la parte conductual o física del individuo sino también de las consecuencias biomecánicas que llegan asistir significativamente en las características de la marcha incluyendo las condiciones del ambiente. De igual forma, se puede estudiar la marcha debido a las limitaciones motrices de una persona con desorden musculoesquelético, enfermedades cardiovasculares o algún tipo de derrame [16].

El estilo de caminar de una persona se describe mediante el ciclo de marcha, que consiste en un movimiento cíclico que comprende la coordinación de segmentos oscilantes [17], sin embargo, la marcha no es perfectamente periódica a pesar de su naturaleza cíclica dando origen a un problema para el análisis de la marcha continua en el tiempo [17, 18]. Generalmente, la marcha es perpendicular al eje óptico del mecanismo de captura como es el caso de las cámaras de seguridad. Por lo tanto, el ángulo de visión transversal es el punto de vista más preferible para obtener información relacionada a la marcha [19]. Además, se llega a considerar directamente a la marcha como un conjunto para extraer información temporal debido a la secuencia de siluetas que contiene un período de la marcha.

La anatomía del cuerpo humano este compuesto por diversos sistemas y órganos que a la vez crean conexiones y comunicaciones para poder realizar una acción como es el levantar un pie para dar inicio al ciclo de marcha y por ende a la acción de caminar. De acuerdo con hipótesis propuestas por neurocientíficos, proponen un concepto que se basa en que las actividades complejas musculares necesarias para coordinar varias articulaciones y músculos son generadas mediante la combinación flexible de un grupo pequeño de módulos básicos donde cada uno de ellos activa un conjunto de músculos con patrones espacio-temporales específicos [20]. A este concepto le denominan sinergias musculares. Dando lugar a este concepto, las sinergias de movimientos que involucran el

ciclo de marcha y por ende la simple acción de coordinar los movimientos para levantar un pie o el braceo al momento de caminar.

Aguilar Moreno y Taboada [21] señalan que los músculos a lo largo de la columna vertebral siguen dos cadenas musculares cinéticas, una anterior con predominio flexor y otra posterior con predominio extensor. La estabilidad postural se deriva de la armonía entre estas cadenas, que, mediante contracciones y relajaciones, mantienen el equilibrio general y, por ende, la posición final de la cabeza. También destacan que los problemas posturales suelen originarse en la infancia debido a la adopción de posturas incorrectas, las cuales, si no se corrigen a tiempo, pueden provocar no solo deformidades estéticas, sino también trastornos en la actividad de órganos internos y funciones como la respiración, la deglución, la circulación y la locomoción. Reguera [22] identifica diversos factores que aumentan el riesgo de dolor de espalda, como el envejecimiento, el sedentarismo, la obesidad, enfermedades hereditarias (como discopatías), enfermedades adquiridas (como el cáncer y la artritis), posiciones inadecuadas y el tabaquismo. Estos factores también contribuyen al riesgo de lesiones y mala postura, puesto que, la espalda desempeña un papel crucial en el soporte al caminar y en el mantenimiento del equilibrio. Además, señalan que el dolor lumbar representa una proporción significativa de los dolores de espalda, afectando al 70 a 80 % de la población adulta mundial en algún momento de sus vidas. Esta región experimenta una carga considerable y se destaca como la causa más común de discapacidad en personas menores de 45 años, generando un impacto económico significativo debido a su alta prevalencia. Además, constituye la segunda razón principal de ausentismo laboral después de las enfermedades respiratorias. Aproximadamente el 15 al 20 % de los adultos sufren de lumbalgia, siendo en su mayoría inespecífica y afectando a todas las edades, aunque es más prevalente entre los 30 y 50 años, aumentando con la edad.

Sin embargo, una de las desventajas existentes en los métodos de reconocimiento de la marcha es que utilizan plantillas de la marcha, las cuales es difícil de preservar la información temporal o la secuencia de la misma. Perdiendo así la flexibilidad del reconocimiento de la marcha y por ende las características de la secuencia. Adicional a esto, el no reconocer la mayor cantidad de información y de características de la marcha de una imagen, la precisión de la estimación de la pose se ve expuesta para obtener buenas estimaciones dando lugar a encontrar anomalías en los datos para poder realizar una buena clasificación y detección de datos.

Debido a estos problemas que enfrenta el estudio de la marcha y su aplicación en inteligencia artificial, la importancia de este proyecto de tesis radica en analizar el ciclo de marcha mediante el uso de técnicas de procesamiento digital de imágenes y Deep Learning para detectar anomalías en la forma de caminar. Esto permite un estudio detenido en el ciclo de marcha, ya que, la información de entrada es considerablemente extensa al tratarse de fotogramas de videos que contienen cientos de fotos para el estudio de la caminata de una persona. Esto presenta un desafío al intentar clasificar ciertas características, como la longitud de la zancada, la velocidad, la ropa que lleva el individuo, entre otras.

2. Antecedentes

2.1. Anatomía del sistema locomotor

El sistema locomotor, también conocido como sistema musculoesquelético, está compuesto principalmente por el sistema muscular y el sistema esquelético, los cuales colaboran de manera conjunta para facilitar el movimiento corporal, proporcionar soporte, dar forma y mantener la estabilidad. Además, cuenta con funciones homeostáticas como es almacenar minerales en el caso del hueso y en la parte muscular almacena carbohidratos en forma de glicógeno [23]. El sistema muscular, como sugiere su nombre, está constituido por músculos de tipo esquelético, que presentan fibras musculares, haces o fascículos, y son de naturaleza voluntaria y estriada. Se estima que existen alrededor de 800 músculos en el sistema muscular humano [23]. Los músculos esqueléticos desempeñan un papel crucial al unirse a las articulaciones para generar movimiento; actúan como conexiones entre los huesos y permiten la movilidad. Además, los tendones se encargan de unir los músculos a los huesos. En contraste, el sistema esquelético está formado por huesos, que pueden ser clasificados como largos, cortos, planos o irregulares, y su función principal es la protección de los órganos internos y el soporte estructural, como en el caso de la caja torácica. En la Figura 1, se muestra la diferencia visual anatómica entre el sistema esquelético y el muscular, en la cual los músculos están señalados de color rojiza y el esqueleto de color blanco, haciendo hincapié que el músculo recubre al hueso y este último a los órganos.



Figura 1: Sistema musculoesquelético. Obtenida de [1].

2.2. Planos anatómicos

Para lograr una localización más precisa de órganos o miembros del cuerpo humano utilizando terminología médica, es esencial comprender los planos anatómicos y sus respectivas posiciones. De acuerdo con Saladin [23], la posición anatómica se define como una postura en la cual la persona se mantiene erguida con los pies planos sobre el suelo y cercanos entre sí, los brazos a los lados y las palmas dirigidas hacia adelante [19], tal como se ilustra en la Figura 1. Esta posición proporciona un marco de referencia fundamental.

La orientación de los planos anatómicos se determina según la región corporal en consideración, siendo la referencia para el corte o plano el aspecto fundamental. El primero implica un corte real con el propósito de visualizar la anatomía interna, mientras que el segundo hace referencia a una superficie plana imaginaria que atraviesa el cuerpo. Se identifican tres planos anatómicos principales: sagital, frontal y transversal, como se ilustra en la Figura 2.

- Plano sagital: Aquel que toma como referencia la línea media vertical del cuerpo u órgano, formando dos secciones: la derecha y la izquierda. Cabe destacar que el plano que divide de manera precisa a la mitad se le llama plano medio sagital, puesto que, el corte del plano pasa precisamente por la línea media. En cambio, si el corte o plano pasa de manera paralela a la línea media, se le llama plano parasagital.
- Plano frontal o coronal: Plano que divide en anterior y posterior, también conocidos como adelante y atrás. Esto se debe a que el corte pasa por la parte coronal, es decir, la cabeza.
- Plano transversal o horizontal: Como indica su nombre, divide el cuerpo u órgano en dos secciones: la parte superior e inferior.

2.3. Biomecánica

La biomecánica es una palabra compuesta por el prefijo bios y la raíz de la palabra mecánica [2], dos ciencias que estudian fundamentos diferentes donde la primera se encarga del fundamento biológico, la anatomía para tener vías de señalización y hacer relaciones de los ligamentos y tendones con la parte física, en cambio la parte mecánica se centra en la física como se mencionó y en la parte matemática. Al unir ambos conceptos por separado podemos definir a la biomecánica como una ciencia que se encarga de estudiar las fuerzas y los efectos aplicados en el cuerpo humano [24], es decir, estudia el comportamiento biológico del cuerpo humano mediante el estudio de la estática, cinética, dinámica y cinemática.

Esta ciencia permite mostrar una mejor descripción generalizada de los movimientos de las articulaciones del cuerpo y por ende tener una principal referencia anatómica de

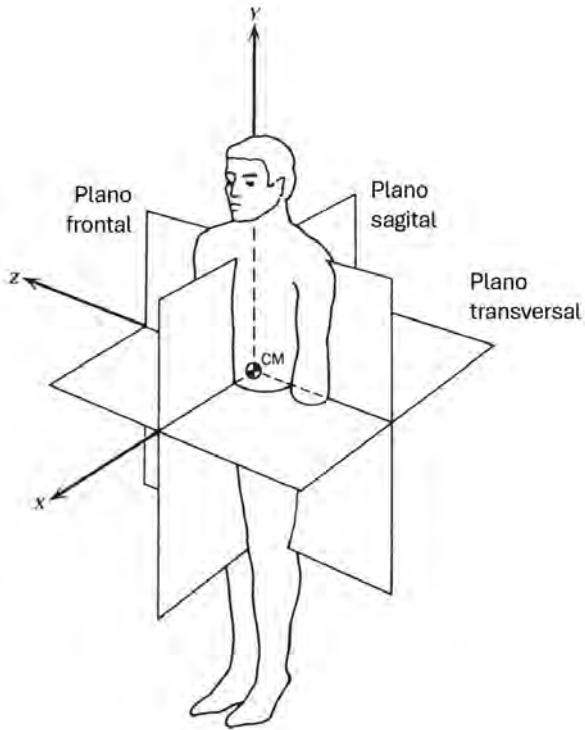


Figura 2: Planos anatómicos tomando como referencia el punto central del cuerpo conocido como centro de masa. Adaptada al español de [2].

los ejes y planos anatómicos como también entender algunas propiedades de los tejidos al realizar algún movimiento, las tensiones y relajaciones de los músculos, la cantidad de fuerza aplicada en una flexión, entre otros. Toda evaluación cuantitativa del movimiento humano debe contener dos pre-fases: la descripción y la medición, en dado caso de tener un diagnóstico más preciso se necesitaría de un análisis biomecánico [19].

2.4. Ciclo de la marcha

El ciclo de marcha o patrón de marcha es un proceso locomotor repetitivo donde el cuerpo humano en posición erecta se mueve hacia adelante siendo su peso soportado por sus miembros inferiores, los cuales se van alternando en un intervalo de tiempo. Este movimiento se caracteriza por el contacto que hay entre el suelo o piso donde camina con al menos uno de sus pies [25]. La marcha se compone por sucesiones de apoyos dobles y unilaterales, obteniendo como consecuencia que el ciclo de marcha se disponga de una sucesión de dos fases: la fase de apoyo y la fase oscilante que se dan en el extremo de miembro inferior derecho.

De acuerdo con Gras [25] y Di Gregorio [26], el ciclo de la marcha se compone de cuatro fases o momentos, la cual depende totalmente de cómo se empieza a estudiar según el momento, aunque normalmente se empieza a estudiar el inicio del ciclo con el contacto del talón derecho, es decir, si se analiza primero con el talón del pie del

miembro inferior atrasado o con el despliegue de la pierna de dicho miembro(derecho) atrasado. A continuación, se describen las cuatro fases del ciclo de marcha tomando como referencia la Figura 3, en la cual se detallan las dos fases principales de la marcha (fase de apoyo y fase oscilante), al igual que los porcentajes de avance en el ciclo de marcha total.

2.5. Fase de apoyo (carga de peso)

La fase de apoyo, se puede referir como carga de peso de acuerdo con Henning [3], debido a que la fase empieza con el contacto inicial del pie con el suelo hasta el despegue de la pierna de apoyo como se ilustra en la Figura 3.

1. **Primer tiempo de doble apoyo:** Corresponde a un periodo de recepción del 0 al 12 % del ciclo y ocurre cuando se despliega en una secuencia específica que comprende un tiempo instantáneo, marcado por el ataque del talón, correspondiente al inicio del ciclo de la marcha, seguido de dos tiempos consecutivos que involucran el descenso del pie derecho hasta el suelo y la verticalización subsiguiente de la pierna [27].
2. **Primer apoyo unilateral:** Abarca del 12 al 40 % del ciclo. Periodo en el que implica el apoyo de la extremidad inferior derecha y la oscilación de la extremidad inferior izquierda en la marcha, con una duración que abarca el 35 % del ciclo total. Durante esta fase, la extremidad inferior derecha desempeña una triple función: sostener el peso corporal, mantener el equilibrio en los tres planos espaciales y avanzar. Inicialmente, la extremidad está inclinada hacia abajo y adelante, luego se vuelve vertical y posteriormente se inclina hacia abajo y hacia atrás. Mientras tanto, la extremidad inferior izquierda experimenta un movimiento oscilante, primero hacia atrás e inclinado hacia esa misma dirección, y luego hacia adelante e inclinado hacia adelante. Esta etapa se puede dividir en cuatro momentos distintos: antes de alcanzar la vertical, después de alcanzar la vertical, un instante coincidente con la vertical y, finalmente, el despegue del pie de apoyo, marcando el cierre del primer periodo de apoyo unilateral [26, 27].

2.6. Fase oscilante

Esta fase, también se puede referir como avance de la extremidad en el espacio. La cual empieza con el despegue de la extremidad hasta el período de máxima flexión de la rodilla hasta el periodo justo antes del contacto inicial [3] como observa en la Figura 3.

1. **Segundo doble apoyo o apoyo posterior de impulso:** Abarca desde el 40 % hasta el 75 % del ciclo total. Durante este periodo, la extremidad inferior derecha se coloca en una posición posterior, ejecutando un apoyo posterior de impulso. La articulación tibiotarsiana (o talocrural, es la unión entre los huesos astrágalo, tibia y peroné) derecha experimenta una extensión gradual, resultando en una

flexión plantar de 15°, mientras que la rodilla permanece ligeramente flexionada, permitiendo que la pierna se incline hacia arriba y hacia adelante para canalizar la energía hacia el avance del cuerpo. La cadera se mantiene extendida. Simultáneamente, la extremidad inferior izquierda realiza un apoyo de recepción y frenado similar al descrito para la extremidad derecha [27].

2. **Segundo apoyo unilateral:** Fase oscilante, sucede cuando la extremidad inferior derecha oscila desde una posición posterior a una anterior, siguiendo las etapas detalladas previamente. En este contexto, la extremidad inferior izquierda, que ahora soporta el peso, repite las mismas etapas que la extremidad inferior derecha [27].

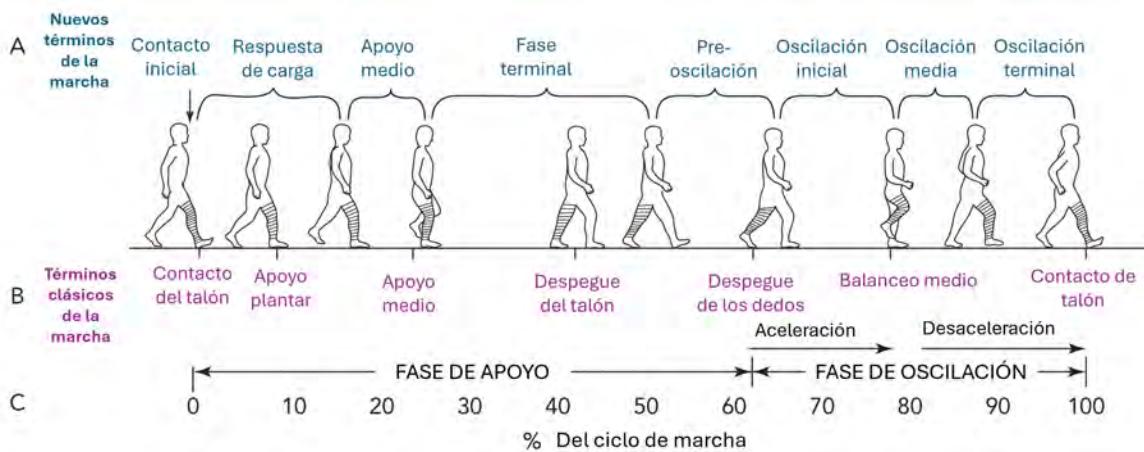


Figura 3: Fases del ciclo de marcha. Adaptada al español de [3].

2.7. Anomalías de la postura

La postura correcta o posición anatómica se caracteriza por la posición de pie, con los pies juntos, los brazos extendidos a lo largo del cuerpo y las palmas hacia adelante. La anormalidad en la marcha se define como un trastorno que afecta alguno de los movimientos necesarios para caminar, involucrando factores como el cerebro, los huesos y los músculos. Diversos problemas articulares, como la artritis, deformidades óseas y afecciones circulatorias, como enfermedades vasculares, pueden influir en la forma de caminar [28].

La alteración en la marcha no solo puede afectar la calidad de vida, sino que también aumenta el riesgo de caídas o lesiones. Desde el punto de vista anatómico, ciertos músculos desempeñan un papel crucial en el mantenimiento de la postura y la prevención de movimientos no deseados. Entre ellos se encuentran el músculo abdominal oblicuo externo, que estabiliza la columna vertebral al levantar objetos pesados [23], y los músculos transversospinales, ubicados en la espalda, que contribuyen significativamente al mantenimiento de la postura y al control del movimiento vertebral cuando los

erectores de la columna actúan sobre la espina dorsal [23]. Dado el proceso evolutivo, las rodillas soportan una carga considerable, ya que enderezarlas al estar de pie permite mantener una postura erguida con mínimo esfuerzo muscular.

Magee [29] enumeran diversas anomalías de la marcha relacionadas con condiciones físicas, tales como:

- Marcha festinante, se caracteriza por una postura rígida y encorvada con la cabeza inclinada hacia adelante [29].
- Marcha espástica, implica caminar rígidamente arrastrando los pies debido a contracciones musculares prolongadas [29].
- Marcha en tijeras donde las piernas se flexionan a la altura de las caderas y las rodillas, golpeándose o cruzándose en un movimiento similar a tijeras [29].
- Marcha en estepaje, con el pie que cuelga y los dedos apuntando hacia abajo [29].
- Marcha de pato, se asemeja a caminar como un pato [29].
- Marcha atáxica o de base amplia, consiste en movimientos irregulares y zigzagueantes [29].
- Marcha arrastrada, en la cual los pies no se levantan normalmente y se deslizan por el suelo, común en casos de enfermedad de Parkinson [29].

2.8. Biometría de la marcha

La marcha es una técnica biométrica que se basa en patrones únicos al caminar haciéndola un identificador para cada persona, puesto que, refleja las características físicas y de comportamiento que permiten definir y diferenciar de forma única a un individuo [30]. De acuerdo con Murray en uno de sus estudios biomédicos en 1967 [30,31] sugirió que un conjunto de factores individuales como son los físicos contribuye a la marcha como: la altura, el peso, la longitud de las piernas y las proporciones de las articulaciones, entre otros. Ocasionalmente que formen un patrón de marcha único para cada individuo [30]. Tomando en cuenta este factor físico singular, la marcha cuenta con cinco propiedades únicas en su captura según Mason [30] que se describen a continuación:

- La captura de la imagen o video puede ser a larga distancia, a diferencia de otras biometrías que deben de estar cerca del recolector como son las huellas digitales.
- Las cámaras de video o cámara pueden tener configuraciones de baja resolución, es decir, tecnologías de bajo costo.
- La instrumentación es simple, ya que las más comunes son: cámaras, acelerómetro en un celular, sensor de suelo o un radar para recopilar la marcha humana.

- El reconocimiento se puede hacer sin la cooperación de la persona.
- Las características de las marchas son difíciles de imitar, ya que utiliza siluetas y actividades humanas.

Las limitaciones de la marcha están sometidas a que ha dificultado sus incorporaciones para fines prácticos desde una perspectiva técnica. La biometría de la marcha puede verse oscurecida por los cambios en la ropa, mochilas y tipo de calzado [30], etc. Ocasionando occlusiones en la captura de este tipo de biometría. Adicionalmente, está sujeta a la condición de la persona objetivo u observada y factores como lesiones pueden afectar negativamente su rendimiento. Para abordar estas limitaciones, se han diversificado los métodos para examinar las características y el comportamiento de la marcha.

2.9. Estimación de la pose

La estimación de la postura humana infiere en los puntos de referencia o puntos clave del cuerpo humano como son los puntos de las uniones de las articulaciones, un ejemplo es la ubicación de los codos o rodillas o partes específicas del cuerpo humano como la cabeza, los hombros, los brazos y las piernas [32].

Por otra parte la estimación de la postura se puede clasificar según el número de coordenadas a ubicar como es en 2D (X, Y) o 3D(X, Y, Z). De acuerdo con Yun et al. [33], la estimación busca predecir y rastrear la ubicación de una persona mediante la postura de las partes y las articulaciones del cuerpo humano en imágenes o videos. Por otra parte, la estimación de la postura humana en 2D, es utilizada para estimar la posición 2D o la ubicación espacial de puntos clave del cuerpo humano.

La estimación de la pose funciona encontrando puntos clave de una persona, que en este caso son las articulaciones. Existen dos tipos:

- Multipose: Se utiliza para detectar poses de varios objetos.
- Monopose: Se utiliza para estimar las poses de un solo objeto en una escena.

2.9.1. Estimadores de pose

Los estimadores de pose actuales son generales y gran parte de estos estimadores se encuentran entrenados en base de datos de la pose, los cuales pueden no ser adecuados para bases de datos de marcha específicas y causar errores de estimación.

Algunos estimadores de pose más comunes se mencionan a continuación:

1. OpenPose: Propuesto por Zhe Cao et al. [34], el cual se caracteriza por detectar las partes del cuerpo o los puntos clave de la imagen, posterior, asigna los puntos clave apropiados para formar pares. Además, es ideal para lograr una alta precisión en la detección de puntos clave del cuerpo, pies, manos y rostro.

2. AlphaPose (Estimación regional de poses multipersonales): Es un método descendente, es decir, ejecuta un detector de personas y estima las articulaciones dentro de cuadros delimitadores, lo cual genera muchos errores de localización e imprecisión durante la predicción. Este tipo de estimador es aplicable para detectar poses individuales o multipersonales en imágenes o videos. Siendo capaz de identificar puntos clave de la cara, el cuerpo, las manos y los pies, resultando crucial para analizar en detalle el comportamiento humano en comportamientos complejos [35].
3. HRNet (Red de alta resolución): Es una red neuronal para la estimación de la postura, la cual es utilizada en el procesamiento de imágenes para encontrar los puntos clave de una persona en una imagen a partir de mapas de calor manteniendo representaciones de alta resolución a lo largo de todo el proceso [36].
4. DeepPose: Estimador de pose humana que aprovecha las redes neuronales profundas (DNN), la cual captura todas las articulaciones al capturar una capa de agrupamiento, una capa de convolución y una capa completamente conectadas [37]. La DNN captura el contexto completo de cada articulación del cuerpo y cada regresor de articulaciones utiliza la imagen completa como señal. Además, Toshet et al. [37] proponen una cascada de predictores de pose en DNN, la cual permite aumentar la precisión de la localización de una articulación utilizando sub-imágenes de mayor resolución.

2.10. Imágenes de la marcha

La adquisición de las imágenes de la silueta para la marcha dependen principalmente del tipo de información que se quiera recuperar como es de forma numérica o de forma visual como son las imágenes y vídeos.

En la Figura 4, se muestran algunos de los ejemplos de los tipos de entrada de la imagen de las bases de datos mencionadas para la extracción de características y del reconocimiento de la marcha como es la imagen RGB (es decir, a color), imágenes binarias que solo se encuentran en dos canales que son blanco y negro, imágenes de energía de la marcha (GEI) que se describe en el siguiente apartado, imágenes con preprocessamiento aplicado como es la segmentación semántica, la cual realiza una segmentación por partes de la imagen asignando una clase a cada uno de los píxeles que se encuentren en dicha clase. De igual forma, se encuentran, las imágenes sintéticas que se pueden obtener mediante softwares específicos como es Blender, entre otros. Sin dejar de lado las entradas que son números o coordenadas como lo son las mallas de la base de datos OUMVLP-Mesh, las localizaciones de las articulaciones en los esqueletos en 3D y 2D.

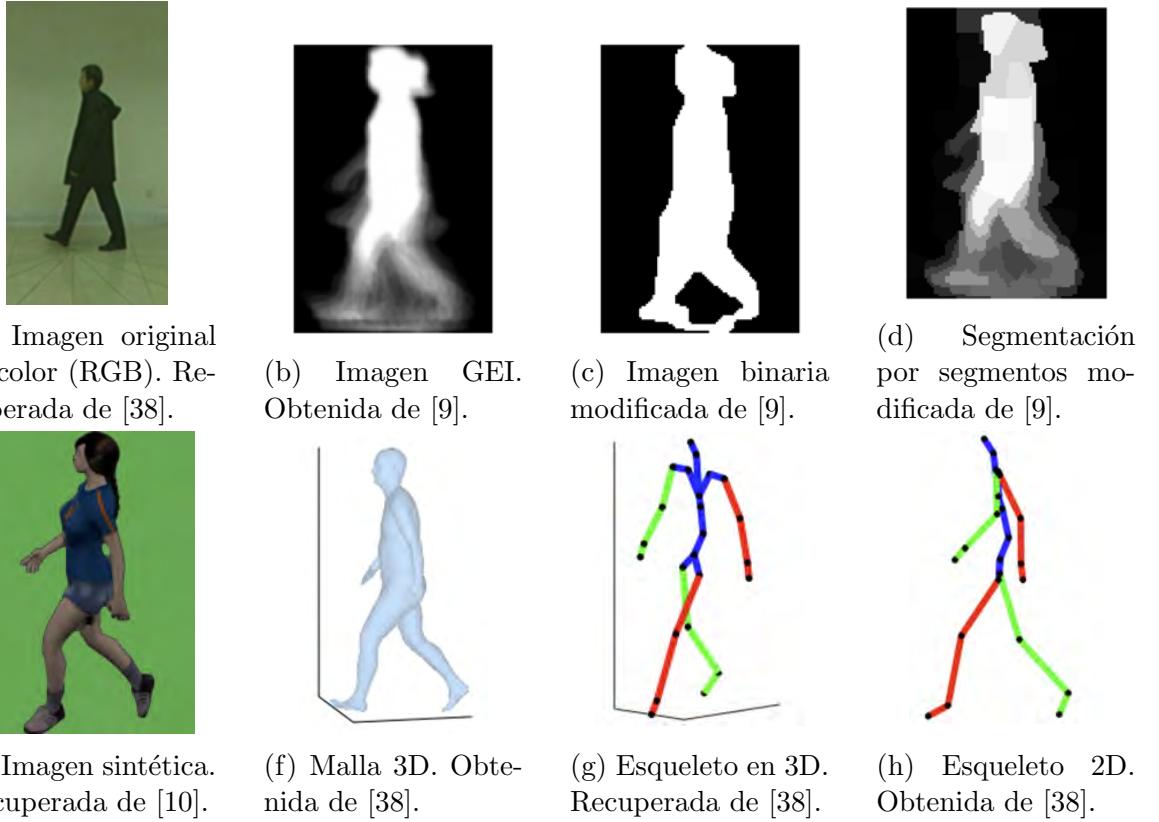


Figura 4: Ejemplos de imágenes de entrada para el reconocimiento y extracción de características de la marcha.

2.10.1. Imágenes de energía de la marcha (GEI)

Las imágenes de energía de la marcha de acuerdo con Hongye y Zhuoya [4], son un promedio de todas las imágenes en un período de marcha que retiene información de las secuencias periódicas del contorno de la silueta de la imagen original, la transformación de fase, la información de frecuencia, y reduce la cantidad de período de marcha.

Consiste en normalizar la secuencia de imágenes como se observa en la Figura 5. Esto con el fin de hacer las imágenes más robustas a variaciones en la adquisición como es la iluminación, distancia, etc. La imagen GEI es considerada como un método tradicional y uno de los tipos de imágenes más utilizadas en la extracción de características, que es una representación espacio-temporal de la marcha que permite extraer las características de la marcha de la imagen de la silueta determinando el esqueleto a partir de los segmentos corporales [4, 39–44]. Dada una secuencia de imágenes binarias del período de la marcha $f_i(x, y)$, la GEI se calcula según [4, 45] cómo se observa en la ecuación 1.

$$G(x, y) = \frac{1}{N} \sum_{i=1}^N f_i(x, y) \quad (1)$$

Donde N es la longitud del período completo de la marcha, i la secuencia de la imagen, (x, y) son las coordenadas de píxeles de la imagen bidimensional. En el caso de la imagen sintetizada GEI de la Figura 5, se representa como una difuminación en escala de grises, es decir, el rango va de 0 a 255 [4].



Figura 5: Secuencia de imágenes normalizadas e imagen GEI. Obtenida de [4].

2.11. Datasets

Algunas bases de datos de la marcha con mayor renombre y de las más utilizadas son CASIA-B/C, SOTON, OU-ISIR- OUMVLP-Mesh, USF HumanID, TUM GAID, entre otras [14, 46]. Alvarez y Sahonero [47], Chao et al. [12] y Huang et al. [48], hacen referencia a CASIA-B, al ser reconocida como uno de los conjuntos de datos más extensos para evaluar técnicas de reconocimiento de la marcha. Por otra parte, OUMVLP-Mesh, propuesta por Li et al. [10] menciona que posee dos conjuntos de coordenadas distintas: 1) coordenadas 2D basadas en imagen (IM2D), que coinciden con las poses en OUMVLP-Pose; y 2) coordenadas 3D centradas en el ser humano (HC3D), que recuperan la información faltante de las imágenes 2D de entrada y se alinean con una coordenada común resistente al ángulo de visión [10].

En la Tabla 1, se describen a mayor profundidad las bases de datos más utilizadas para el análisis de la marcha.

2.12. Extracción de características

La extracción de características consiste en capturar eficazmente las características de la marcha. Asimismo, deben de ser robustas a las condiciones operativas y deben de brindar una buena capacidad de discriminación entre individuos [49]. Esta extracción de características es un proceso crucial, ya que, se han desarrollado varios métodos, incluidos descriptores espaciales y temporales o descriptores de bajo nivel basados en información anatómica extraída de la imagen [41]. Múnera [41] menciona algunos descriptores inspirados en las proporciones del cuerpo humano como es el conteo de cabezas para la relación de la silueta o altura, el área relativa, el perímetro relativo y el HOG (Histograma de Gradientes Orientados) promedio como características de entrada para para modelos del cuerpo. Algunos enfoques extraen parámetros de movimiento como la longitud de la zancada, el tiempo de ciclo de marcha, la velocidad y los ángulos de las articulaciones a partir de imágenes de la postura 2D [44].

Dataset	Núm. datos	Captura	Condiciones	Núm. de variaciones	Ref.
Soton Multimodal	400	Caminata, cara y oído	-	3	[9]
CASIA-B	124	Caminata	Caminata normal, con un abrigo y con una mochila	11	[9]
CMU MoBo	25	Caminata	Caminata: lenta, rápida, inclinada y con una pelota	6	[9]
AVA	20	Caminata	Caminata en caminos curvos y rectos	6	[9]
WOSG	155	Caminata	8 variaciones de vista	8	[9]
KY 4D	42	Caminata	-	-	[9]
OU-ISIR C	200	Caminata	Inclinación en la vista de la cámara	25	[9]
OU-ISIR LP	4016	Caminata	-	4	[9]

Tabla 1: Data sets más usados para el análisis de la marcha.

De acuerdo con Ascenso [50], la extracción de siluetas mediante visión artificial ofrece varios algoritmos, estos se pueden dividir en tres categorías: sustracción de fondo, segmentación semántica y segmentación multivista. Cada categoría tiene conjuntos de datos específicos que se utilizan para probar los algoritmos y las métricas que se utilizan para evaluar su rendimiento, es decir, los algoritmos que pertenecen a diferentes categorías se prueban en diferentes métricas. Debido a que, cada categoría suele estar dedicada a una tarea específica para la que se han desarrollado.

Se han propuesto varios métodos para este tipo de extracción. Ye Bo [51] introdujo un algoritmo basado en un modelo gaussiano para extraer siluetas completas y cerradas eliminando el ruido y las sombras. Chen y Zhang [52] desarrollaron un enfoque de segmentación conjunta (en inglés cosegmentation), la cual segmenta simultáneamente regiones comunes en múltiples imágenes, aprovechando la similitud entre fotogramas en secuencias de marcha para mejorar la precisión de la extracción. En la cual solo los puntos que tienen alto contraste con el fondo se utilizan como puntos del núcleo de primer plano, los puntos en el fondo se utilizan como núcleo de fondo de puntos. Este tipo de método, de acuerdo con los autores tiene un mejor rendimiento en comparación con la silueta humana tradicional de métodos de extracción. Ng et al. [53], presentaron un enfoque multietapa donde incluían operaciones morfológicas, segmentación anatómica y transformada de Hough para extraer ángulos de articulación y dimensiones corpora-

les. El objetivo de estos métodos es superar problemas comunes como el desorden del fondo y las similitudes de color entre el primer plano y el fondo, mejorando en última instancia el rendimiento de los sistemas de reconocimiento de la marcha.

La información espacial derivada de imágenes como las imágenes de energía de la marcha (GEI), funciona bien en condiciones de visión no extremas. La información temporal complementa esto al proporcionar información adicional sobre los cambios en la marcha, mejorando el reconocimiento en ángulos de visión extremos u occlusiones donde la información espacial es limitada [38]. Sin embargo, el movimiento humano normalmente se representa mediante la información espacial y temporal, ocasionando que la complejidad de la extracción de las características en el reconocimiento de la marcha sea influenciada por diversos factores, como las variaciones de los puntos de vista, la ropa, las condiciones de transporte y las occlusiones [4, 39–44, 50].

2.12.1. Características espaciales

De acuerdo con Hongye [4], la división de extracción de características se divide en información espacial e información temporal [39–44] que permiten capturar con precisión los patrones únicos de la marcha.

Las características espaciales describen los aspectos estáticos o estructurales de la marcha centrándose en las características estáticas y dinámicas de la forma del cuerpo, la silueta, la posición de las extremidades y las relaciones espaciales entre las partes del cuerpo en una imagen [42, 54]. Algunas técnicas incluyen redes neuronales convolucionales, extracción de características basadas en partes o se centran en regiones corporales específicas. Algunos ejemplos son la forma del torso, el ángulo de los brazos al balancearse o la longitud de las piernas, entre otros [39–44]. Por otra parte, algunos autores dividen en siete o cinco áreas [43] para un análisis de partes del cuerpo donde se extraen las características de cada región respectivamente y posterior se ensamblan en las características de la imagen de la marcha para su identificación [4]. Otros autores como Lishani [40] extraen las características de textura de Haralick localmente en tres regiones de la imagen. Finalmente, las características espaciales como su nombre indica se basan en el espacio y se enfocan principalmente en la configuración física del cuerpo durante la marcha. Algunas características espaciales clave incluyen las posiciones angulares de las articulaciones como la cadera, rodilla y tobillo que proporcionan una representación detallada de la geometría de la marcha [55].

2.12.2. Características temporales

Las características temporales como su nombre indica, se basan en el tiempo y la secuencia de los movimientos. Este tipo de características capturan los cambios dinámicos de la marcha a lo largo del tiempo como es la secuencia y el ritmo de los pasos, los ángulos de las articulaciones y los patrones de movimiento entre fotogramas, entre otros [42, 56].

Por otra parte, las mediciones temporales incluyen la cadencia, los tiempos de apoyo

y balanceo, la duración del doble apoyo, la variabilidad del tiempo de zancada y la velocidad de la marcha que en algunas ocasiones es normalizada en función de la altura [54]. Debido a que las partes del cuerpo poseen patrones de movimiento diversos, se requiere de modelos temporales que consideren representaciones a múltiples escalas. Algunos trabajos se enfocan en características temporales como es la captura de movimiento dinámico donde Wan et al. [57], capturan características temporales como información clave de sincronización de la marcha mediante el análisis de los movimientos articulares a lo largo del tiempo, permitiendo la ponderación adaptativa de diferentes movimientos. Xiaohu Huang et al. [42], proponen una red de aprendizaje de características temporales sensibles al contexto (CSTL) para el reconocimiento de la marcha, en el que integran características temporales multiescalares de acuerdo con la información contextual a lo largo de la dimensión temporal, permitiendo la comunicación de información entre tres escalas diferentes, es decir, a nivel de fotograma, a corto plazo y a largo plazo. Chen et al. [56], se centran extraer relaciones entre siluetas de marcha adyacentes a nivel de píxel y bloque, potenciando la extracción de características temporales de las secuencias de la marcha.

2.13. Inteligencia Artificial

El concepto de la inteligencia artificial fue cuñado en 1956 por John McCarthy, definiéndolo como “la ciencia y la ingeniería de máquinas inteligentes”. Según Ruiz [58], la inteligencia artificial se describe como la combinación de habilidades mediante un conjunto de técnicas informáticas para abordar temas o encontrar soluciones específicas mediante el uso de máquinas.

Dentro de este campo, se destacan dos subáreas conocidas como aprendizaje automático (Machine Learning) y aprendizaje profundo (Deep Learning). El primero, definido por Arthur L. Samuel, se refiere a un campo de estudio que permite que las computadoras aprendan sin ser programadas previamente. Así, el aprendizaje automático se define como el conjunto de algoritmos capaces de aprender a partir de datos específicos, extrayendo patrones para facilitar la toma de decisiones futuras [58]. Por otro lado, el aprendizaje profundo, subcampo del Machine Learning, se centra en modelos predictivos para audio, video, texto e imágenes, utilizando redes neuronales que automatizan la fase de extracción de características y se enfocan en una mayor cantidad de datos [5].

Es esencial señalar que tanto el Machine Learning como el Deep Learning utilizan dos clasificaciones principales de datos y algoritmos: supervisado y no supervisado. Esto hace referencia a si los datos de entrada están etiquetados o no, es decir, si las variables tienen valores reales para la variable objetivo.

En la Figura 6, se presenta el esquema seguido para el aprendizaje supervisado con Machine Learning, donde se emplea un conjunto de entrenamiento para predecir una variable objetivo. En contraste, en el aprendizaje no supervisado, el entrenamiento se basa en datos sin etiquetas para realizar análisis descriptivos, como la detección de anomalías, agrupación o asociaciones, como se muestra en la Figura 7.

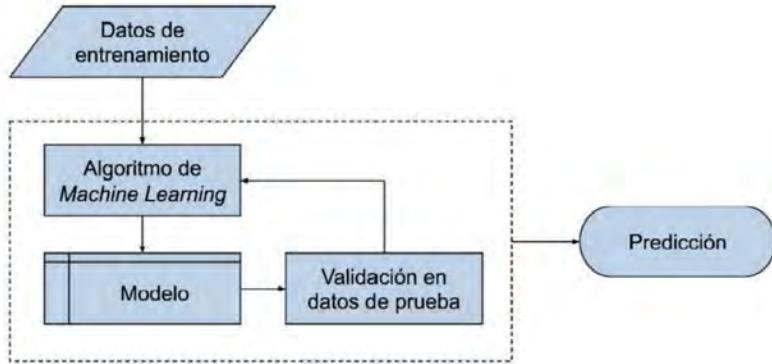


Figura 6: Esquema de aprendizaje supervisado, el cual genera un modelo que utiliza un conjunto de entrenamiento para predecir una variable objetivo, es decir, el conjunto de entrenamiento esta etiquetado. Recuperado de [5].

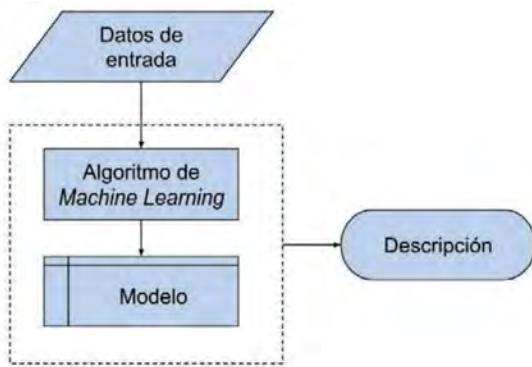


Figura 7: Esquema de aprendizaje no supervisado, consiste en entrenar un modelo que utiliza datos sin etiquetas para realizar un análisis descriptivo que puede ser asociaciones, detección de anomalías o agrupación. Imagen recuperada de [5].

2.13.1. Redes Neuronales Convolucionales

De acuerdo con Albawi et al. [59], la Red Neuronal Convolucional (CNN) recibe su nombre debido a la operación matemática lineal entre matrices conocida como convolución. La convolución se refiere al proceso de deslizar una ventana o filtro sobre la matriz de píxeles de entrada, multiplicando los valores de la ventana por los pesos correspondientes y sumándolos para generar un valor único en una nueva matriz [59]. La entrada o imagen de entrada en la CNN se compone de bloques de construcción llamados píxeles. Los píxeles están organizados secuencialmente en una disposición similar a una matriz donde el brillo y el matiz de cada píxel están especificados por su valor de píxel que puede ir de 0 a 255 [60]. En la Figura 8, se observa una representación gráfica de una CNN, donde cada dimensión corresponde a un canal de color RGB si consideramos una imagen.

Una red neuronal convolucional (CNN) está conformada típicamente por cuatro

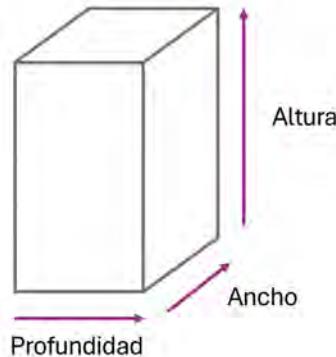


Figura 8: Representación de la entrada tridimensional de una CNN. Imagen propia.

capas: la capa convolucional, la capa de agrupamiento o pooling en inglés, la función de activación y la capa totalmente conectada (fully connected) [60] como se muestra en la Figura 9. La capa convolucional, esencial en la CNN, se compone de múltiples núcleos de convolución (filtros) y funciones de activación (capas no lineales). Además, su función principal es extraer información profunda de las características de los datos de entrada mediante un proceso de filtrado multicapa. El tamaño del núcleo de convolución puede especificarse como cualquier valor menor que el tamaño de la imagen de entrada, pero su profundidad es la misma que la de los datos de entrada [61]. Por otra parte, la capa de no linealidad sigue a la convolución, permitiendo que la salida generada sea cambiada o terminada, es decir, la función de activación determina si se activa o no una neurona en respuesta a una entrada determinada generando la salida correspondiente [60]. Algunas funciones de activación mayormente ocupadas en las CNN son: ReLu donde todos los valores de entrada se convierten en rango positivo [60] y sigmoide, función que permite valores de salida entre 0 y 1 [60] por mencionar algunas.

La capa de agrupamiento o capa de submuestreo se encarga de reducir la dimensionalidad de los mapas de características, es decir, reducir la resolución [6,60]. Generalmente ubicada entre capas convolucionales sucesivas, realiza la selección de características después de la extracción de características en la capa convolucional. Su objetivo es conservar las características fundamentales de los datos de entrada mientras reduce la complejidad de la información de las características, disminuyendo parámetros y dimensiones, y mejorando así la capacidad de generalización del modelo [61].

Finalmente, la función de la capa completamente conectada radica en integrar las características obtenidas de las capas superiores, generando un mapa de características como un único valor. Esta capa elimina la información de posición y mejora la robustez del modelo. Cada capa de la capa completamente conectada consta de numerosas neuronas, siendo esta la capa con más parámetros en la CNN, representando aproximadamente el 80 % de todos los parámetros de la red [61].

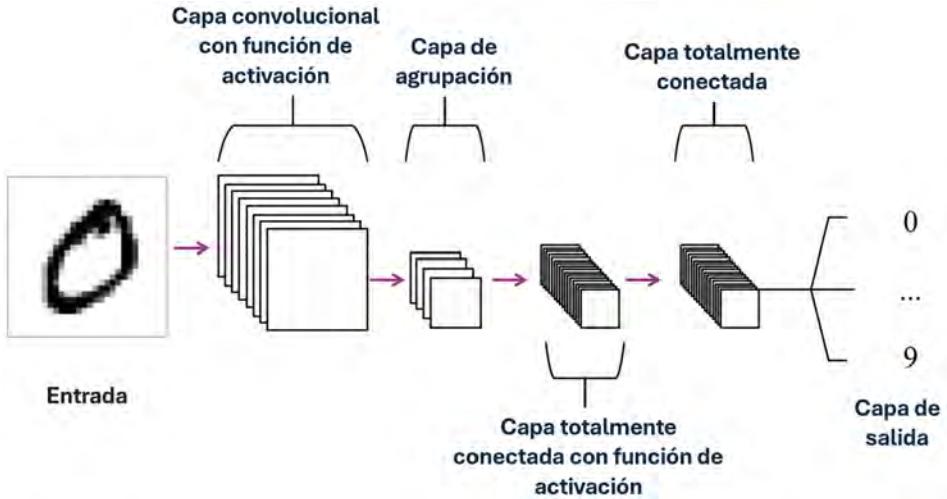


Figura 9: Arquitectura CNN sencilla compuesta por cinco capas, en la que se observa sus respectivos componentes. Adaptada al español de [6].

2.1 VGG-16

La arquitectura VGG16 por sus siglas en inglés de Visual Geometry Group Network o en español Red de Grupos de Geometría Visual [62, 63] fue inventada por Simonyan et al. [64]. Esta red neuronal convolucional, se destaca por su profundidad y simplicidad. Se caracteriza por su arquitectura profunda que consta de 16 capas que incluye 13 capas convolucionales, 3 capas completamente conectadas y 5 capas de agrupamiento (max-pooling), permitiendo que con esta profundidad capture características complejas [62, 65]. Además, está diseñada específicamente para tareas de reconocimiento y clasificación de imágenes.

La estructura de una VGG16 se conforma por una capa de entrada que recibe los datos de la imagen de entrada, la cual es redimensionada a una dimensión fija, mayormente de 224×224 píxeles. Posterior, se encuentran las 13 capas convolucionales, cada capa extrae características de la imagen de entrada una serie de filtros. Cada filtro se recorre sobre los datos de entrada, implementando multiplicaciones y sumas de elemento por elemento para generar un mapa de características de bajo nivel para que la red aprenda patrones más complejos con menos parámetros según el tamaño del filtro (comúnmente 3×3) [63]. Las capas convolucionales son seguidas por una función de activación, es decir, la capa de activación. Esta capa aplica una función de activación tipo ReLu (Unidad Lineal Rectificada o en inglés Rectified Linear Unit) mayormente. Introduciendo la no linealidad al modelo donde el modelo aprende patrones complejos [63]. Para el caso de las 5 capas de agrupamiento (pooling), reducen las dimensiones espaciales de los mapas de características con el fin de disminuir la complejidad computacional y controlar el sobreajuste [63]. Adicionalmente el max-pooling o el valor máximo de los mapas de características se encarga de retener el valor máximo dentro de cada mapa. Finalmente, las últimas tres capas completamente conectadas, combinan las características extraídas

de las capas anteriores para realizar la predicción y en la última capa se aplica una función de activación softmax para generar las probabilidades de las clases [63]. En la Figura 10 se muestra un diagrama de los bloques de las capas que componen a VGG16.

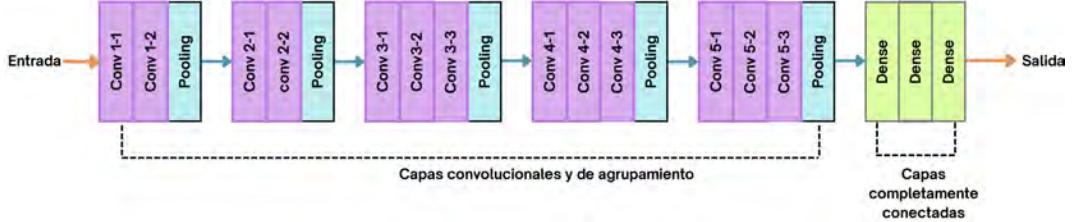


Figura 10: Arquitectura general del modelo VGG-16. Imagen propia.

2.2 InceptionV3

InceptionV3 es la tercera versión de la CNN Inception de Google [65]. Es una arquitectura de red neuronal convolucional profunda diseñada para clasificación siendo eficiente y precisa en imágenes, con 48 capas [66]. Se conforma de múltiples capas y módulos especializados que le permiten extraer características complejas de las imágenes sin aumentar excesivamente el costo computacional [66]. Además, divide la convolución en convoluciones más pequeñas, reduciendo el número de parámetros.

Su estructura general se compone de un bloque de entrada, módulos Inception, reducción de tamaño, clasificador auxiliar y un bloque final que están integradas por capas convolucionales, promedio de agrupamiento (Avg Pool), MaxPool, Concat, DropOut, capas totalmente conectadas (Fully Connected) y Softmax [7] como se observa en la Figura 11. El bloque de entrada se compone de la entrada de imágenes de $229 \times 229 \times 3$, varias capas convolucionales de filtros de 3×3 y 1×1 para extraer características básicas. Posterior, los bloques o módulos de Inception, combinan múltiples filtros convolucionales de diferentes tamaños para capturar características multiescala [67]. Estos módulos son únicos puesto que, utilizan convoluciones paralelas con diferentes tamaños de filtro: 1×1 , 3×3 o 5×5 para capturar información a múltiples escalas y posterior concatenar los resultados [66, 68]. Las capas concatenadas fusionan los mapas de características de diferentes filtros [66] y las capas de agrupación, reducen las dimensiones espaciales e incluyen una agrupación media global de la clasificación final [66]. Esta última sustituye las capas totalmente conectadas tradicionales antes de la salida, reduciendo el sobreajuste y el tamaño del modelo [66]. La inserción de un clasificador auxiliar en etapas intermedias facilita regularizar el entrenamiento, mejorando la convergencia y la generalización. Finalmente, la capa final, se compone de capas densas o totalmente conectadas que realizan la clasificación final mediante la función de activación Softmax.

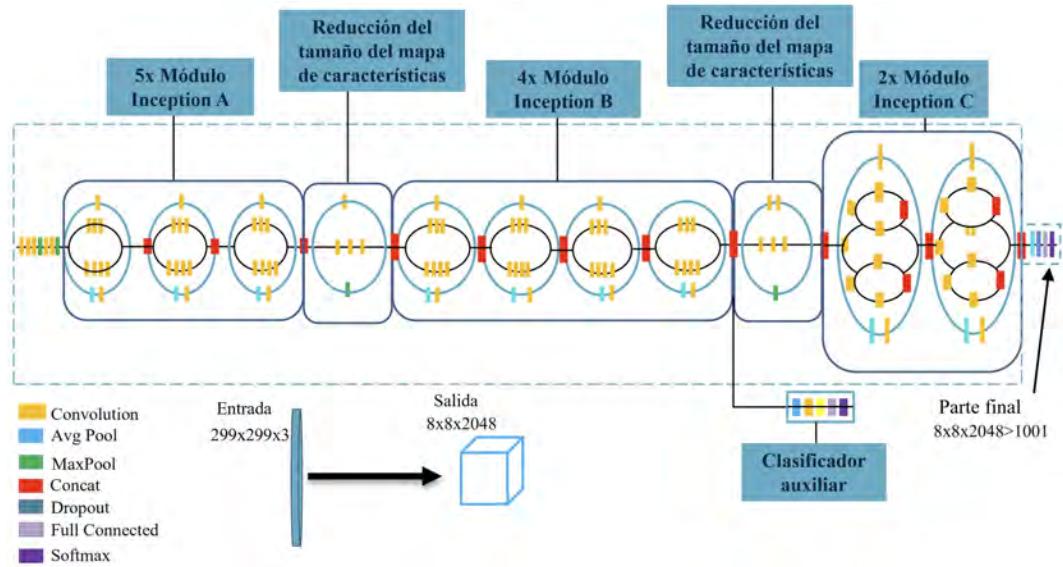


Figura 11: Arquitectura general de InceptionV3. Imagen adaptada al español de [7].

2.3 ResNet50

ResNet es uno de los pioneros de la normalización por lotes, ya que, permite diseñar CNN más profundas que pueden ser de hasta 152 capas sin comprometer la capacidad de generalización del modelo, cómo también permite extraer características [65]. ResNet50 como su nombre indica se compone de 50 capas y se basa en bloques residuales con conexiones de atajo (skip connections) que permiten entrenar redes profundas y disminuir el problema de gradiente desvanecido.

La estructura general de ResNet50, se componen de una capa de inicial de convolución y pooling que capturan las características iniciales y reducen la dimensión respectivamente. Los bloques residuales donde cada uno tiene una estructura tipo cuello de botella o en inglés bottleneck que consiste en tres capas convolucionales en serie: 1×1 convolución para reducir la dimensión del canal, es decir, comprimirlo, 3×3 para procesar las características espaciales y 1×1 para restaurar la dimensión del canal [69]. Adicional a esto, las conexiones residuales o saltos son rutas de acceso directo que omiten una o más capas, permitiendo que la red aprenda las funciones residuales y sea más profunda [69]. Cabe destacar que se compone de varios bloques organizados en 4 etapas, obteniendo un total de 16 bloques residuales que permiten a la red detallar simples hasta estructuras complejas. La capa final se compone de una capa de average pooling global para reducir las características espaciales a un vector y una capa completamente conectada (fully connected (FC)) que mapea el vector de características al espacio de clases para una predicción final [69]. La Figura 12 muestra su arquitectura general.

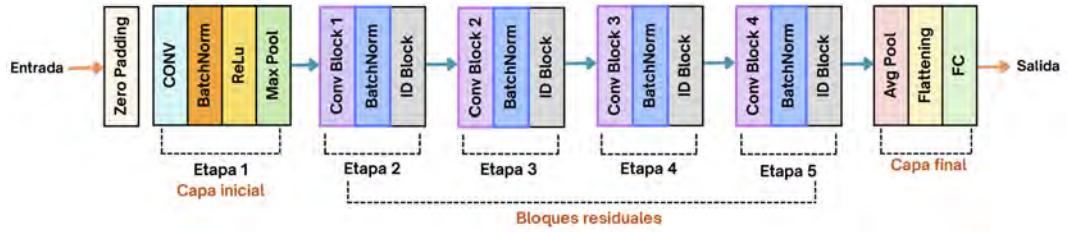


Figura 12: Arquitectura general de ResNet50. Imagen propia.

2.13.2. Visión artificial

Alonso y Jara [8] definen la visión por computadora, también conocida como visión artificial, como una rama de la Inteligencia Artificial clásica. Esta disciplina involucra diversas fases cruciales destinadas a conferir a la computadora la capacidad de percibir y comprender una imagen, con el objetivo de emular el proceso realizado por los seres humanos. Aunque la arquitectura y las etapas de un sistema de visión varían según la aplicación específica, algunas fases son comunes en la mayoría de las aplicaciones, tal como se ilustra en la Figura 13.

La visión artificial puede llegar a tener diversas aplicaciones, se destaca el reconocimiento de objetos y patrones. Se pueden distinguir dos tipos de reconocimiento: la detección de objetos cuando se busca un objeto específico en una imagen, y el reconocimiento de instancias cuando se pretende identificar diversas instancias de una categoría genérica. El último tipo implica la clasificación de objetos en distintas clases, como reconocer diferentes marcas y modelos de vehículos, asignándolos a la categoría de „autos” basándose en características compartidas. El reconocimiento de objetos se realiza mediante clasificadores, herramientas que asignan clases a objetos según características numéricas como intensidad, color, esquinas o textura. Estas características, denominadas patrones, se representan como vectores en un espacio de características. Los clasificadores pueden reconocer patrones, pero no objetos directamente, ya que operan en base a las características numéricas proporcionadas [8].

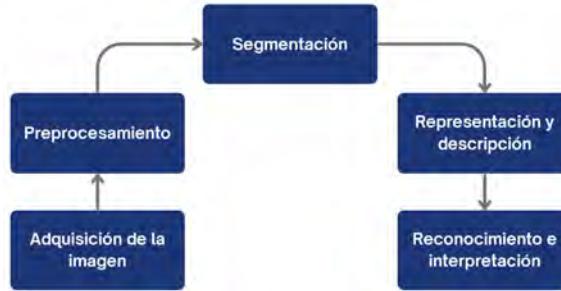


Figura 13: Etapas de la visión artificial de acuerdo con Alonso y Jara [8]. Imagen mejorada de [8].

2.13.3. Métricas

Para la evaluación del rendimiento del modelo se pueden emplear diferentes métricas que permitan evaluar la efectividad del modelo en predecir los resultados correctos en relación con los conceptos de la matriz de confusión [63, 70]. La matriz de confusión muestra cómo se desempeñó el modelo al clasificar las diferentes clases: las clases predichas por el modelo y las clases reales o verdaderas. Además, permite visualizar cómo se distribuyen las predicciones del modelo comparando los valores reales (las filas) con los valores predichos (las columnas) como se observa en la Figura 14 donde la diagonal de la matriz corresponde a las predicciones correctas y los valores fuera son los errores o que fueron predicciones incorrectas.

		VALORES PREDICCIÓN	
		Predicho Positivo	Predicho Negativo
VALORES REALES	Real Positivo	Verdadero Positivo (VP)	Falso Negativo (FN)
	Real Negativo	Falso Positivo (FP)	Verdadero Negativo (VN)

Figura 14: Matriz de confusión. Imagen propia.

La matriz de confusión cuenta con cuatro componentes principalmente:

- **Verdadero Positivo (VP):** Representa los casos donde tanto los valores de clase reales como los predichos son positivos, es decir, el número de veces que el modelo predice positivo y lo es [63].
- **Verdadero Negativo (VN):** Representa los casos donde tanto los valores reales como los predichos son negativos, [63].
- **Falso positivo (FP):** Cuando el valor real es negativo y el valor predicho es positivo, es decir, el número de veces que las predicciones fueron positivas pero son negativas [63].
- **Falso Negativo (FN):** Cuando el valor de clase real es positivo, y el valor de clase predicho es negativo [63].

Se pueden derivar métricas a partir del rendimiento de la matriz de confusión como es:

- **Exactitud:** También conocido por Accuracy en inglés, es la proporción de todas las predicciones que fueron correctas, es útil cuando los conjuntos de datos tienen falsos positivos y falsos negativos simétricos [63]. Se representa matemáticamente por la ecuación 2, la cual fue adaptada al español de [63, 70].

- **Precisión** : Corresponde al número de predicciones verdaderas y positivas dividido por el número total de resultados positivos devueltos por el clasificador [70]. Es decir, la relación de valores que fueron realmente positivos de todos los valores predichos como positivos [63]. Está relacionada con la tasa de falsos positivos [63] y se representa matemáticamente por la ecuación 3 adaptada al español de [63,70].
- **Sensibilidad o Recall**: Es la relación de observaciones positivas correctamente predichas frente al total de observaciones en la clase real, es decir, cuantos fueron identificados correctamente de todos los positivos reales [63]. En la ecuación 4 se representa su expresión matemática que fue adaptada de [63,70].
- **F1 Score o F-score**: Es la media armónica de la precisión y sensibilidad donde un valor de 1 indica que es el mejor de los casos y 0 en el peor de los casos [70]. Es mayormente útil cuando los datasets están desbalanceados o cuando los falsos positivos y falsos negativos tienen diferentes costos [63] y se representa matemáticamente por la ecuación 5 recuperada de [63,70].
- **CCR**: La tasa de clasificación correcta o por sus siglas en inglés CCR (Correct Classification Rate), es otra forma de referirse a la exactitud como se observa en la ecuación 6.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2)$$

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (3)$$

$$\text{Recall} = \frac{VP}{VP + FN} \quad (4)$$

$$\text{F1 Score} = \frac{2(\text{Recall} \times \text{Precisión})}{\text{Recall} + \text{Precisión}} \quad (5)$$

$$\text{CCR} = \frac{VP + VN}{VP + FP + VN + FN} \quad (6)$$

2.13.4. Técnicas para extracción de características

Métodos para evaluar la calidad de la extracción de características se centran en cuánta información relevante y discriminativa han conservado las nuevas características en relación con la separación de clases o según el objetivo de estudio.

- **Análisis de Componentes Principales (PCA)**: Reduce la dimensionalidad de los datos manteniendo la máxima varianza posible sin tomar en cuenta las diferencias entre clases. Es decir, mide cuánta variabilidad de los datos originales retienen las nuevas características, transformando las variables originales en un

nuevo conjunto de variables no correlacionadas llamadas componentes principales, centrándose en minimizar el error cuadrático medio entre los vectores de datos originales y los de la dimensión reducida [71].

- **Análisis de Discriminante Lineal (LDA):** Reduce la dimensionalidad maximizando la separación entre clases, con el objetivo de maximizar la dispersión entre clases y minimizar la dispersión dentro de la clase [71]. Además, LDA se enfoca principalmente en reducir la variación de datos dentro de las clases, al tiempo que aumenta la separación entre ellas [72].
- **Análisis de Correlación Canónica (CCA):** CCA (Análisis de Correlación Canónica): Es un método que examina las correlaciones entre dos grupos de variables [73], maximizando la correlación entre dos conjuntos de variables a través de combinaciones lineales. Su evaluación se basa en los valores de correlación canónica obtenidos.

3. Hipótesis

La extracción de características del ciclo de marcha mediante visión artificial y uso de Deep Learning permiten ayudar a la detección de datos de anomalías de la postura al caminar, validando las características mediante el análisis de correlación canónica (CCA), el análisis discriminante lineal (LDA) y la tasa de clasificación correcta (CCR).

4. Objetivos

4.1. Objetivo general

Analizar el ciclo de marcha mediante el uso de técnicas de procesamiento digital de imágenes y Deep Learning para detectar anomalías en la forma de caminar.

4.2. Objetivo específicos

1. Aplicar técnicas de redes neuronales convolucionales (CNN) y técnicas de preprocesamiento para extraer características de la silueta de la persona.
2. Estudiar las posturas a partir del preprocesamiento de imágenes y redes neuronales convolucionales para detectar si hay o no anomalía en la postura.
3. Validar los resultados mediante el análisis de correlación canónica (CCA), el análisis discriminante lineal (LDA) y la tasa de clasificación correcta (CCR).

5. Metodología

En este apartado se describirá a detalle las etapas y pasos empleados para la extracción de características de las imágenes de energía de la marcha, modelos de clasificación y análisis de datos para la detección de anomalías para clasificar si una persona tiene una posible buena o mala postura. El diagrama de flujo de la Figura 15 se reflejan los pasos generales.

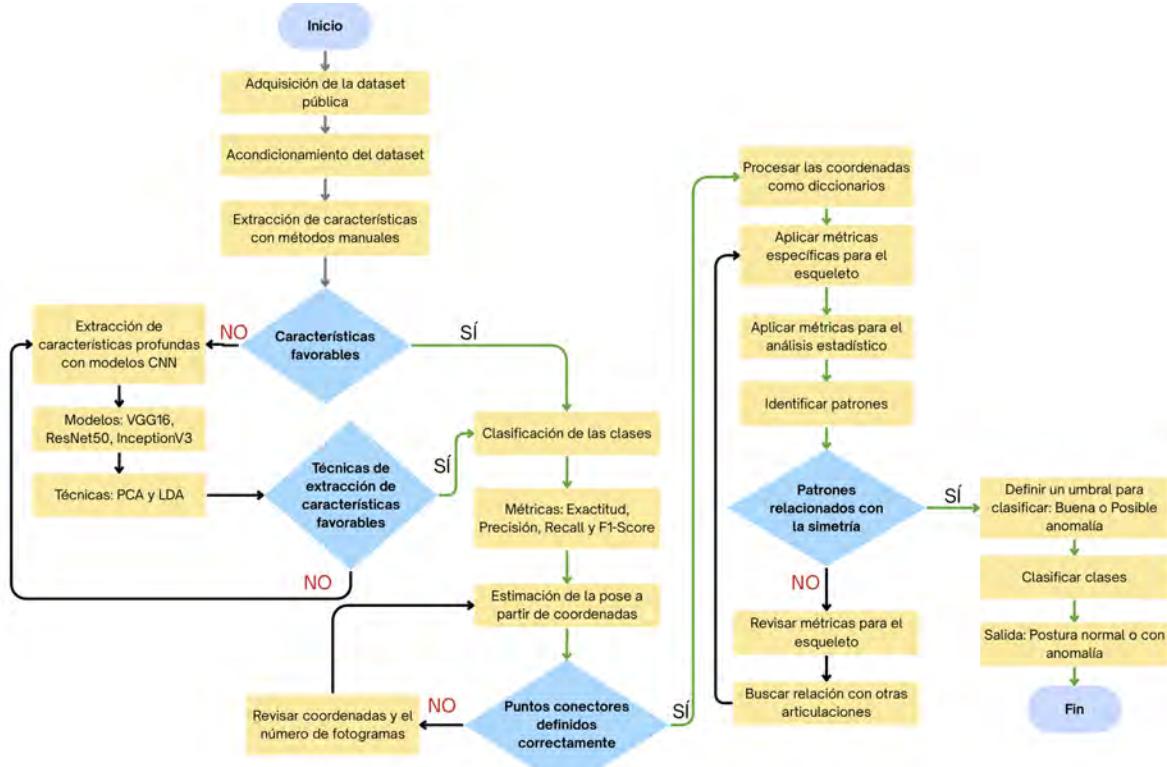


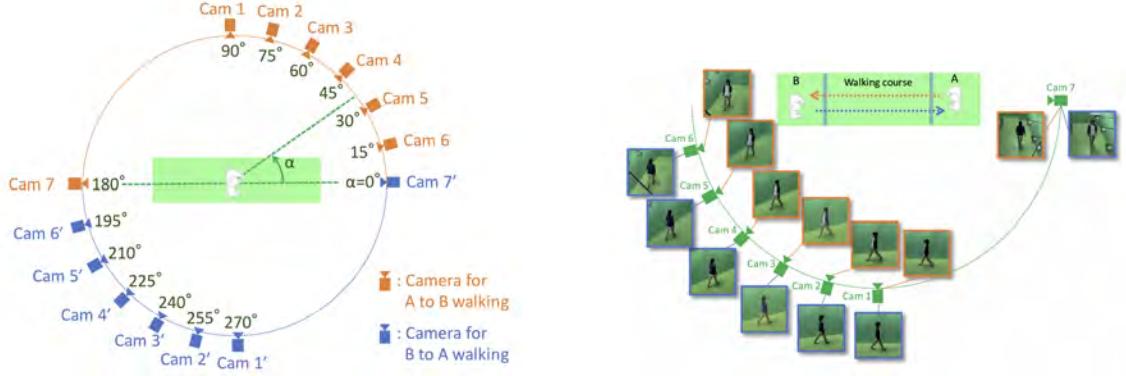
Figura 15: Diagrama de flujo de la metodología general.

5.1. Dataset

5.1.1. Adquisición de imágenes

La base de datos requerida para este trabajo de tesis fue OU-ISIR MVLPI, proporcionada por la Universidad de Osaka. La cual cuenta con 10,307 sujetos con 14 vistas, es decir, 14 ángulos que van de 0° - 90° y 180° - 270° . La base de datos se destaca por capturar imágenes de 1280×980 píxeles a 25 fps (fotogramas por segundo) por siete cámaras colocadas en intervalos de ángulo de acimut de 15° a lo largo de un círculo cuyo centro coincide con el centro del recorrido de la caminata como se muestra en la Figura 16a.

Esta base de datos contiene variaciones intrasujeto de la marcha para todas las vistas de la cámara debido a dos secuencias de la marcha capturadas para cada cámara como se muestra en la Figura 16b donde la vista de la cámara 1 a 7 corresponden a las cámaras señaladas en color naranja, es decir, cuando el individuo camina de ida (A-B) y la vista de 1' a 7' a las cámaras de color azul cuando camina de regreso (B-A).



(a) Posición de las cámaras acorde a los ángulos propuestos por Takemura et al. [9]. Imagen obtenida de [9].

(b) Captura de la base de datos OU-ISIR MVLP en dos secuencias diferentes. Imagen recuperada de [9].

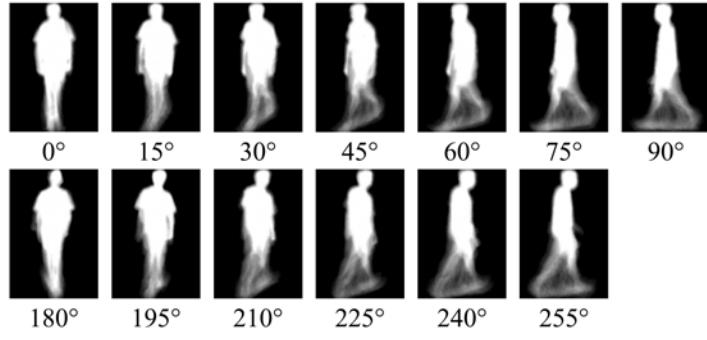
Figura 16: Posición de las cámaras para la captura de la base de datos.

La base de datos se conformaba de dos tipos de información:

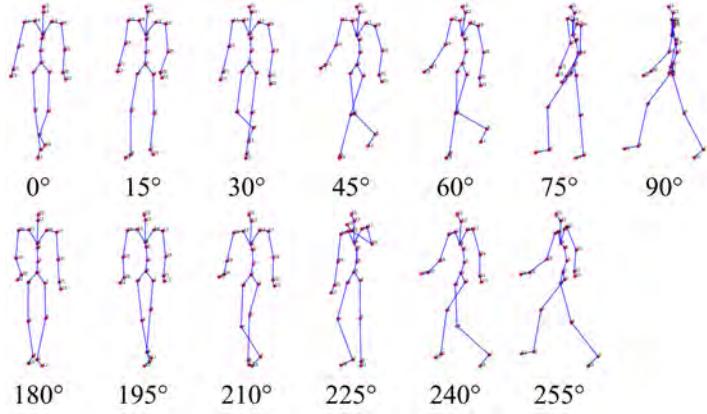
1. **Imágenes de energía de la marcha (GEI)**: Son imágenes de energía acumulativa normalizada en el tiempo de la caminata en un ciclo completo de marcha, es decir, una secuencia de silueta durante un período de marcha, mezclando las partes estáticas y dinámicas del individuo al caminar. Estas imágenes cuentan con una dimensión de 88×128 píxeles. Ver Figura 17a
2. **Coordenadas 2D basadas en imágenes (IM2D)**: Coordenadas basadas a partir de las imágenes OU-ISIR MVLP, las cuales cambian de acuerdo con el ángulo de visión de la imagen cómo se muestra en la Figura 17b.

5.2. Acondicionamiento del dataset

Para este trabajo de tesis, la base de datos mencionada con anterioridad fue modificada en el número de individuos o imágenes que se ocuparon para el entrenamiento de los modelos de redes neuronales convolucionales. De la dataset OU-ISIR MVLP que eran imágenes de energía de la marcha (GEI) con dimensiones de 88×128 píxeles con extensión PNG, se escogieron 4000 sujetos con diversidad de género y edad para cada ángulo teniendo como requisito de oclusión las imágenes que tuvieran algún indicio de oclusión de la ropa como son abrigos o mochila en la espalda. Se enfocó en una sola



(a) Imágenes de energía de la marcha del dataset OU-ISIR MVLP [9].



(b) Imágenes de los esqueletos a partir de las coordenadas IM2D. Imagen propia.

Figura 17: Imágenes de la base de datos en diferentes puntos de vista (o ángulos).

vista de la caminata, la vista de 1 a 7 que comprende a los ángulos 15° , 30° , 45° , 60° , 75° , 90° y 180° como se muestra en Figura 18. Creando un ciclo de marcha.

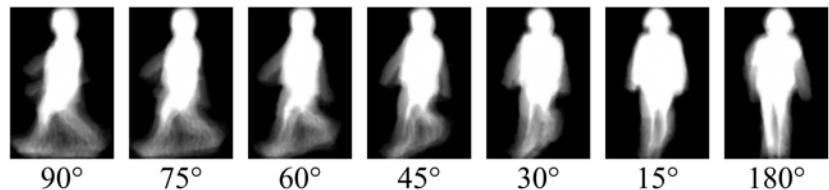


Figura 18: Fotogramas para crear un ciclo de marcha con las imágenes de energía de la marcha (GEI) a partir del dataset de [9].

La segunda parte del dataset, fueron las coordenadas 2D basadas en imágenes (IM2D) que consisten en una serie de archivos JSON por cada fotograma, es decir, cada ángulo de cada individuo. El archivo JSON esta conformado por un diccionario de coordenadas donde cada coordenada indica un punto de interés para la formación del

modelo del esqueleto que se muestra en la Figura 19. Este esqueleto esta conformado por 24 articulaciones o puntos de interés que son: Pelvis, Lhip, Rhip, Spine1, Lknee, Rknee, Spine2, Lankle, Rankle, Spine3, Lfoot, Rfoot, Neck, Lcollar, Rcollar, Head, Lshoulder, Rshoulder, Lelbow, Relbow, Lwrist, Rwrst, Lhand y Rhand. Donde cada punto interés es una articulación del esqueleto. De acuerdo con el esqueleto base de la Figura 19, se creó un diccionario siguiendo las uniones de los puntos de interés para crear los gráficos de los esqueletos correspondientes como se observa en la Figura 20.

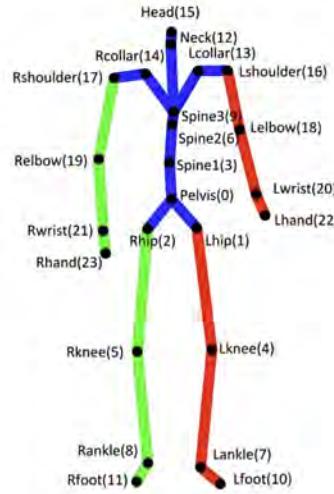


Figura 19: Modelo del esqueleto humano con 24 articulaciones. Imagen recuperada de [10].

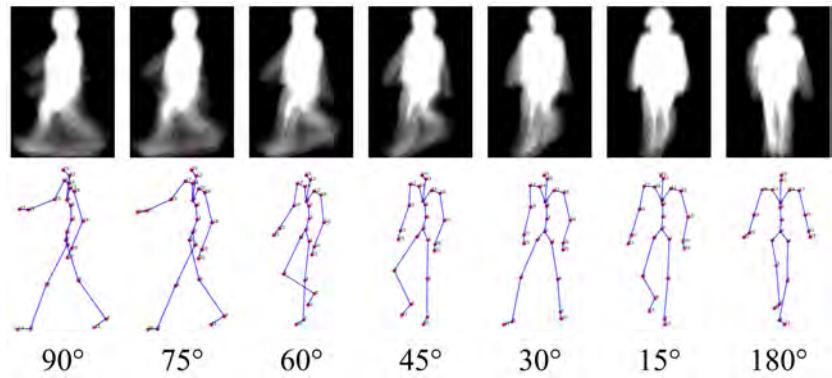


Figura 20: Secuencia de los ángulos que conforman el ciclo de marcha con su imagen GEI y esqueleto respectivamente. Imagen propia.

5.3. Extracción de características

La etapa de extracción de características fue realizada en el software Python. Es esta etapa se probaron los métodos manuales que comprende la redimensión con diferentes

tipos de interpolación, binarización, procesos de segmentación, detección de contornos y extracción de máscaras binarias en diferentes rangos. Además de estos, se empleo la extracción de características basadas en texturas y con modelos de clasificación con Redes Neuronales Convolucionales.

Cada uno de los métodos empleados para la extracción de características se realizó usando todas las imágenes de entrenamiento que comprende un total de 4000 individuos por cada ángulo (7 ángulos), obteniendo así un total de 28,000 imágenes.

5.3.1. Métodos manuales

Los métodos manuales comprenden métodos de preprocesamiento que permiten extraer características de la silueta del individuo en las imágenes GEI, las cuales tienen una propiedad peculiar al tratarse de una secuencia de varios fotogramas en una sola imagen, dando origen a una imagen en escala de grises con algunas sombras. Debido a esto es que la extracción de forma manual es indispensable para comprender la complejidad de la imagen y de la información de la silueta.

5.1 Redimensión con diferentes tipos de interpolación

La imagen original cuenta con una dimensión de 88×128 píxeles, imagen con una escala muy pequeña. Debido a esto es que se emplearon las siguientes configuraciones:

- Tipos de reescalamiento:
 - Reescalamiento por factor: Se redimensiona utilizando factores de escalamiento en porcentaje en el ancho y alto de la imagen.
 - Reescalamiento a dimensiones específicas: Se redimensiona la imagen a un tamaño específico en píxeles.
- Tipos de interpolación:
 - Vecino más cercano
 - Bilineal
 - Bicúbica
 - Mediante la relación de área de píxeles
 - Lanczos sobre el vencinario 8×8

5.2 Segmentación

Los métodos de segmentación permiten separar o dividir en regiones de interés una imagen, con el fin de aislar las partes de la imagen a estudiar. La mayoría de las técnicas de este procesamiento se llevan a cabo al cambiar las intensidades de los píxeles.

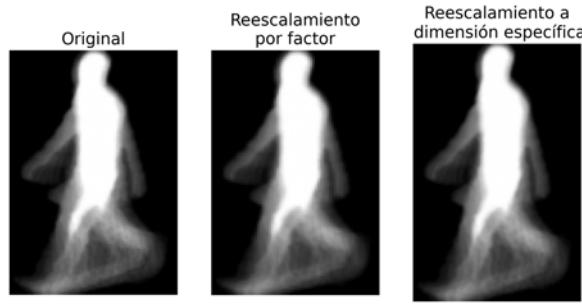


Figura 21: Tipos de reescalamiento. Imagen propia.

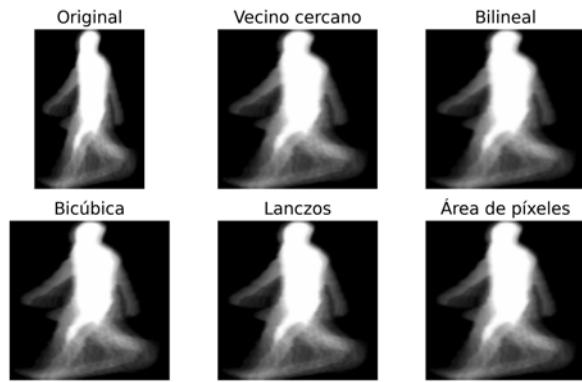


Figura 22: Tipos de interpolaciones.Imagen propia.

5.3.1 Binarización

Binarización aplicada en diferentes umbrales con el fin de cubrir la mayor parte de la silueta, principalmente el área de las piernas y en los brazos en el momento del balanceo de los brazos. La binarización consistió en los siguientes pasos:

1. Cargar la imagen GEI con sus dimensiones iniciales de 128×88 píxeles.
2. Redimensionar la imagen a 264×384 píxeles.
3. Aplicar la función `threshold` de OpenCV para aplicar un umbral binario a la imagen de entrada, la cual segmenta en dos partes: los píxeles que se encuentran dentro del umbral establecido (se establecen con el valor máximo de 255 que da el color blanco) y los que se encuentran fuera del rango (0, color negro).
4. Definir un umbral que se encuentre entre 0 a 255.
5. Visualizar la imagen binarizada en un umbral establecido.

La Figura 23 muestra 6 ejemplos de segmentación mediante binarización con diferentes umbrales.

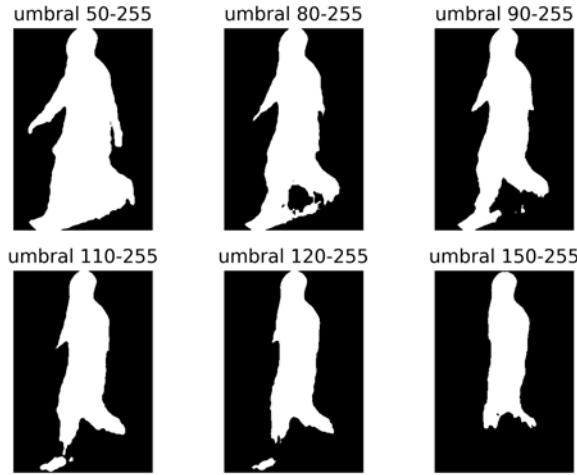


Figura 23: Binarización en diferentes umbrales. Imagen propia.

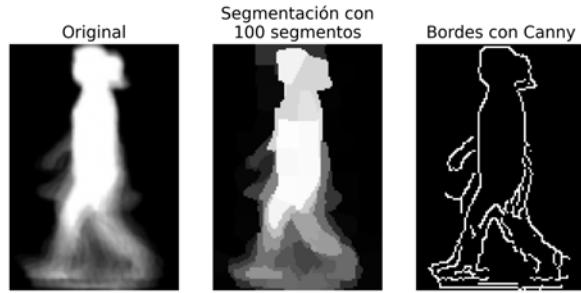
5.3.2 Segmentación por segmentos

Consiste en separar por segmentos la imagen de interés a partir de umbrales específicos que abarcan dicha área como se observa en la Figura 24b o una segmentación por agrupación de píxeles similares en regiones coherentes como se muestra en la Figura 24a. A continuación se describen los pasos generales para ambos casos:

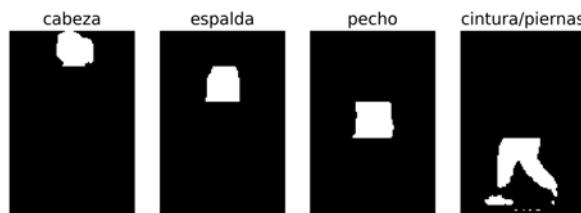
- Cargar la imagen GEI con sus dimensiones iniciales de 128×88 píxeles.
- Convertir a escala de grises.
- **Segmentación por partes**
 - Aplicar un umbral binario de 127 a 255.
 - Definir las proporciones de las partes a segmentar para el caso de segmentación por partes del cuerpo.
 - Cortar las imágenes en partes de acuerdo con las proporciones definidas.
- **Segmentación por agrupación de píxeles**
 - Aplicar SLIC (Simple Linear Iterative Clustering) con 100 segmentos, es decir, 100 regiones coherentes.

5.3.3 Segmentación Chan-Vese

Segmenta los objetos sin límites definidos donde se basa en conjuntos de niveles que se desarrollan de forma iterativa para minimizar la energía que se define mediante valores ponderados correspondientes a la suma de las diferencias de intensidad con respecto al valor medio fuera de la región segmentada. Además depende de la longitud



(a) Preprocesamiento de la silueta con segmentación por segmentos con 100 iteraciones y detección de contornos con el método de Canny.
Imagen propia.



(b) Segmentación en cuatro partes de la silueta.
Imagen propia.

Figura 24: Segmentación por segmentos y por partes.

del límite de la región segmentada y solo se implementa en imágenes en escala de grises. A continuación, se en listan los pasos de esta segmentación:

1. Cargar la imagen GEI con sus dimensiones iniciales de 128×88 píxeles.
2. Convertir imagen a escala de grises.
3. Convertir la imagen a flotante.
4. Aplicar función de chan_vese de skimage donde se parametriza los valores del peso que controla la suavidad del contorno segmentado (μ), los pesos de control para la fidelidad dentro y fuera del contorno segmentado (λ_1 y λ_2 respectivamente), tolerancia para el criterio de paro, el número máximo de iteraciones que se realizará para afinar la segmentación y el tamaño del lote de paso para la actualización iterativa.
5. Graficar los resultados, ver Figura 25.

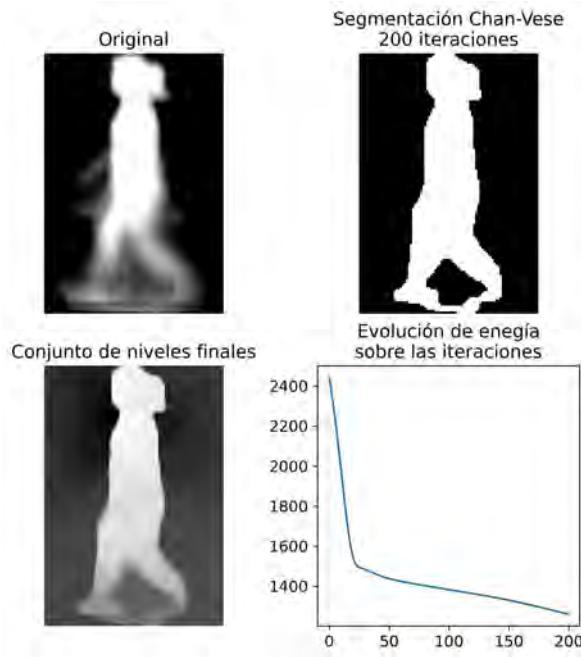


Figura 25: Método de segmentación Chan-Vese. Imagen propia.

5.3 Detección de bordes

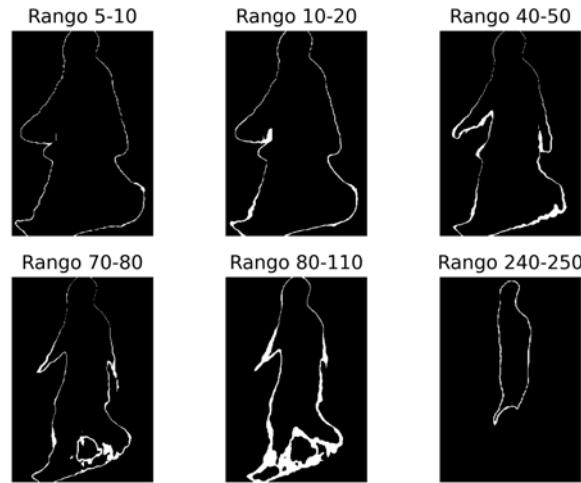
5.3.1 Extracción de los contornos mediante máscaras binarias

Se definen máscaras binarias bajo rangos específicos que van de 0 a 255 como se muestra en la Figura 26a.

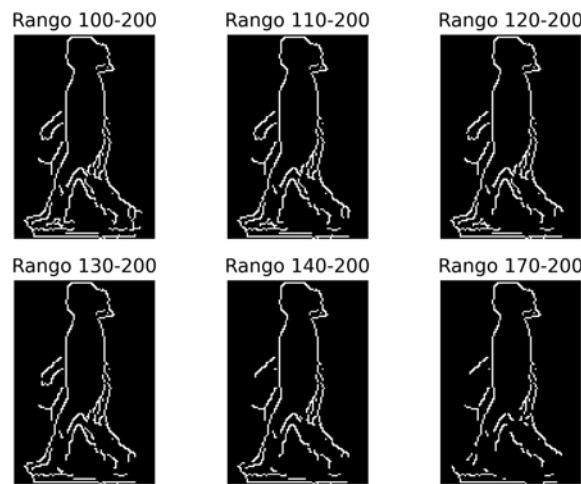
5.3.2 Método de Canny

Función de OpenCV con el mismo nombre, en el cual se establecen los umbrales para la detección de los contornos de la silueta tal como se observa en la Figura 26b. Para encontrar los contornos de las imágenes de entrada GEI, se siguieron los siguientes paso generales:

1. Cargar imagen GEI.
2. Convertir a escala de grises.
3. Aplicar el detector de bordes Canny con `cv2.Canny` definiendo los parámetros de los umbrales inferior y superior. En el caso de la máscara se toman los mismos parámetros cambiando la función por `cv2.inRange`.



(a) Máscaras binarias en diferentes rangos.



(b) Detección de bordes con Canny.

Figura 26: Detección de bordes en umbrales específicos. Imagen propia.

5.3.2. Método basado en texturas

5.1 Matriz de coocurrencia de niveles de grises

También conocido como Grey level co-occurrence matrix features por sus siglas en inglés (GLCM), es un método que examina la relación espacial de los píxeles donde el principal objetivo de estudio son las texturas a partir de las intensidades de los píxeles en niveles de grises de una imagen. En los siguientes puntos se mencionan los pasos para este método.

1. Cargar imagen GEI.
2. Convertir a escala de grises.

3. Obtener el vector con los valores promedio de las 13 características de Haralick.
4. Seleccionar los puntos a estudiar que correspondan al fondo de la imagen y a la silueta de la imagen.
5. Aplicar la matriz GLCM con la función graycomatrix.
6. Graficar cada uno de los puntos seleccionados del fondo y la silueta para visualizar las texturas cercanas como se muestra en la Figura 27.

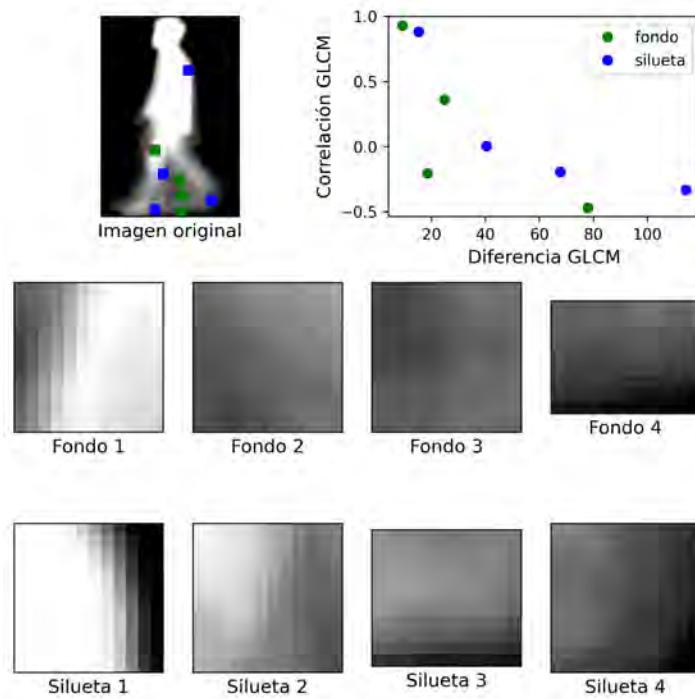
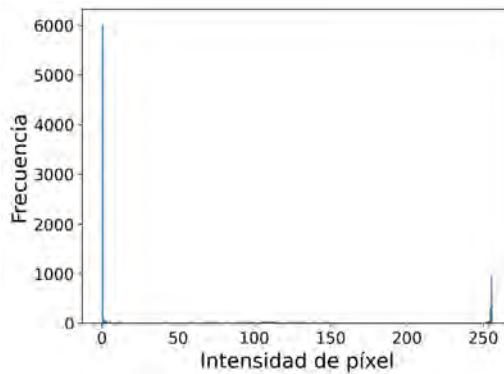


Figura 27: Matriz de coocurrencia de niveles de grises (GLCM). Imagen propia.

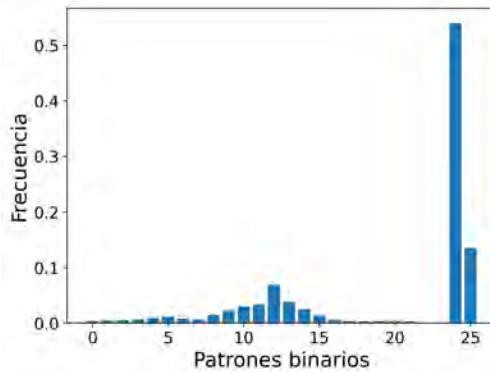
5.2 Patrón binario local (LBP)

Es un operador que describe las características de textura local de las imágenes. El cual compara cada píxel de una imagen con los píxeles vecinos para codificar la información de textura local en patrones binarios [74]. Se calcula un histograma sobre todos los píxeles, que captura la frecuencia de los diferentes patrones de LBP como en la Figura 28b donde cada barra del histograma indica la frecuencia de patrones binarios en la imagen. En este histograma, se observa que hay un patrón dominante en 25, que indica que la mayoría de los píxeles de la imagen tiene este patrón y que a la vez este patrón dominante corresponde a áreas homogéneas de la imagen, es decir, regiones con poca variación de intensidad como es el fondo negro de la imagen GEI.

Por otra parte, los patrones de menor frecuencia se encuentran entre el rango de 5 y 20, un rango intermedio. Estas frecuencias corresponden a regiones con cierta textura o variación en los valores de intensidad que pertenecen a la silueta de la persona. Al compararlo con un histograma general como el que se muestra en la Figura 28a, se observa que solo marca ambos extremos indicando que hay una mayor cantidad de píxeles con intensidad igual a cero, es decir, el fondo negro de la imagen, y aproximadamente 1000 píxeles correspondiente a una intensidad de 255, que indica un color blanco y a la vez se centra en la parte central de la silueta. De forma general, el histograma LBP de una imagen GEI nos indica que son imágenes poco detalladas al tener una textura mayormente uniforme en la silueta permitiendo ver mejor los píxeles con texturas grises a si se observa en un histograma general.



(a) Histograma de una imagen GEI.



(b) Histograma de patrón binario local (LBP).

Figura 28: Histogramas. Imagen propia.

5.4. Redes Neuronales Convolucionales

Se extrajeron características de los patrones de textura de tres modelos de Redes Neuronales Convolucionales: VGG16, ResNet50 e InceptionV3. Arquitecturas especializadas en procesamiento de imágenes que extraen características mediante sus capas

convolucionales.

5.4.1. VGG16

5.1 Material

Los recursos de hardware y software empleados para desarrollar, entrenar y testear el modelo, se detalla en los siguientes puntos. Así mismo, la implementación del software fue trabajado en lenguaje de programación de Python con las especificaciones mencionadas de las librerías utilizadas.

- Computadora:

- Windows 11
- Procesador Intel(R) Core(TM) Ultra 7
- Sistema operativo de 64 bits

- Colab Pro:

- T4 RAM amplia
- GPU del backend de Google Compute Engine en Python 3
- RAM de sistema: 51 GB
- RAM de GPU: 15 GB

- Software:

- Python 3.11.13

- Librerías:

- TensorFlow 2.19.0
- Keras 3.10.0
- scikit-learn 1.6.1
- NumPy 2.0.2
- Matplotlib 3.10.0
- Seaborn 0.13.2

5.2 Datos

Para el entrenamiento y prueba del modelo VGG16, se emplearon dos conjuntos de datos:

- Conjunto de entrenamiento: Este conjunto de imágenes se compone de 4000 imágenes GEI para cada uno de los 7 ángulos. Obteniendo así un total de 28,000 imágenes que pertenecen a 7 clases (7 ángulos).
- Conjunto de validación: Conjunto de 7022 imágenes GEI perteneciente a 1000 aproximadamente para cada uno de los ángulos.

Cada uno de los conjuntos de datos para entrenamiento y validación se especificaron de la siguiente forma:

- Tamaño de requerido de las imágenes: 224×224 píxeles
- Tamaño de lote: 32
- Tipo de etiqueta: Categórica (codificada como one-hot)

5.3 Arquitectura

Arquitectura de modelos de Redes Neuronales Convolucionales especializado en procesamiento de imágenes. La VGG16 se caracteriza por su estructura homogénea compuesta por 16 capas: 13 capas convolucionales y 3 componentes conectados. Se cargó el modelo VGG16 con TensorFlow y Keras con los pesos de la base de datos de ImageNet especificando una tamaño de las imágenes de 224×224 píxeles con 3 canales. Una vez cargado el modelo base de VGG16, se congelaron las capas del modelo base para que no se entrenen inicialmente y se añadieron 3 capas adicionales para la tarea de clasificación de 7 clases. Las capas añadidas fueron:

- Capa de pooling global (GlobalAveragePooling2D) que hace un agrupamiento global sobre el mapa de características de la salida de la última capa convolucional, es decir, convierte el mapa de características en un vector
- Capa densa que toma la salida del GlobalAveragePooling, transformando el vector a uno de 1024 elementos y se aplica una función de activación ReLu a la salida de cada neurona
- Se añade una capa para regularizar la red y prevenir posibles sobre ajustes con un Dropout de 0.5. Desactivando aleatoriamente un 50 % de neuronas en la capa anterior de cada iteración.

En la Figura 29 se visualiza las capas del modelo modificado para la clasificación de imágenes GEI. Posterior, se compiló el modelo con un optimizador de tipo Adam, una tasa de aprendizaje de 1e-4, una pérdida de tipo categorical_crossentropy y como métrica principal la exactitud (accuracy) durante 30 épocas.

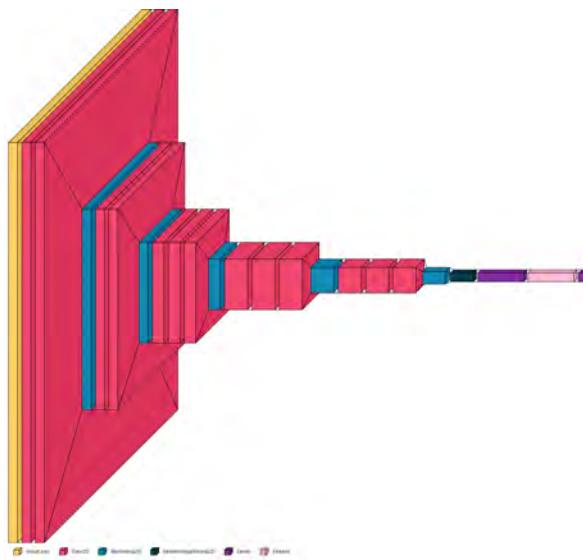


Figura 29: Diagrama de capas convolucionales de la arquitectura VGG16 modificado.

5.4.2. Inception V3

5.1 Material

Los recursos computacionales empleados para entrenar y testear la arquitectura de Inception V3, se describen en los puntos siguientes. Adicional a esto, se trabajo en un entorno de Python con las versiones especificadas de las librerías.

- Computadora:
 - Windows 11
 - Procesador Intel(R) Core(TM) Ultra 7
 - Sistema operativo de 64 bits
- Colab Pro:
 - T4 RAM amplia
 - GPU del backend de Google Compute Engine en Python 3
 - RAM de sistema: 51 GB
 - RAM de GPU: 15 GB
- Software:
 - Python 3.11.13
- Librerías:
 - TensorFlow 2.19.0

- Keras 3.10.0
- scikit-learn 1.6.1
- NumPy 2.0.2
- Matplotlib 3.10.0
- Seaborn 0.13.2

5.2 Datos

Al igual que la arquitectura de VGG16 en InceptionV3, se emplearon los mismos dos conjuntos de datos: entrenamiento y validación. A diferencia de las especificaciones que se definieron de la siguiente manera:

- Tamaño de requerido de las imágenes: 299×299 píxeles
- Tamaño de lote: 32
- Tipo de etiqueta: Categórica (codificada como one-hot)
- Normalización de los píxeles de un rango de 0 a 1
- Transformaciones para el aumento de datos: transformaciones de corte (shearing) y zoom aleatorio con un 2 % de los datos cada uno y un volteo horizontal aleatorio de las imágenes.

Debido a que se realizó un aumento de datos se obtuvo un total de 32,040 imágenes de entrenamiento perteneciente a 7 clases y 7,022 imágenes en el conjunto de validación.

5.3 Arquitectura

La arquitectura InceptionV3, combina múltiples operaciones convolucionales en paralelo para capturar características en diferentes escalas. Este modelo cuenta con 48 capas, en la Figura 30 se observa su diagrama. Para la clasificación empleada y la extracción de características, se definió un modelo de InceptionV3 preentrenado con los pesos de ImageNet, añadiendo una parte superior personalizada de las siguientes capas:

- Se toma la salida del modelo InceptionV3 base preentrenado, que contiene representaciones de características de alto nivel de las imágenes de entrada.
- Se añade un capa GlobalAveragePooling2D para reducir las dimensiones espaciales de los mapas de características. Preparando los datos para la siguiente capa densa.
- Capa densa de 1024 unidades con activación de tipo ReLu, para permitir que el modelo aprenda patrones complejos y tener una representación de las características extraídas según el número de clases para la clasificación.



Figura 30: Diagrama de las capas de la arquitectura Inception V3.

La compilación inicial del modelo fue parametrizada de la siguiente forma: Optimizador tipo RMSprop (Root Mean Square Propagation), una pérdida de tipo categorical_crossentropy y cómo métrica principal la exactitud (accuracy) durante 10 épocas iniciales. Después se descongelaron algunas capas del modelo base para ajustar de forma fina, parametrizando de nuevo el modelo con un optimizador tipo SGD, con una tasa de aprendizaje de 0.0001, un momentum de 0.9, conservando el tipo de pérdida y métrica. Este compilador modificado se entreno durante otras 10 épocas.

5.4.3. ResNet50

5.1 Material

Los recursos computacionales que soportan la implementación, el entrenamiento y la validación del modelo de ResNet50 se describen a continuación.

- Computadora:
 - Windows 11
 - Procesador Intel(R) Core(TM) Ultra 7
 - Sistema operativo de 64 bits
- Colab Pro:
 - T4 RAM amplia
 - GPU del backend de Google Compute Engine en Python 3
 - RAM de sistema: 51 GB
 - RAM de GPU: 15 GB
- Software:
 - Python 3.11.13
- Librerías:
 - PyTorch 2.6.0+cu124
 - scikit-learn 1.6.1
 - NumPy 2.0.2
 - Matplotlib 3.10.0
 - Seaborn 0.13.2

5.2 Datos

Se utilizaron los mismos conjuntos de datos de entrenamiento y validación que las arquitecturas anteriores. Sin embargo, para la arquitectura ResNet50 se empleo mediante las librerías de PyTorch, por lo que cambiaron las configuraciones de los conjuntos de datos, los cuales se describen a continuación:

- Redimensiona las imágenes a un tamaño de 224×224 píxeles
- Tamaño de lote: 32
- Tipo de etiqueta: Categórica (codificada como one-hot)
- Normalización de los píxeles de un rango de 0 a 1
- Transformaciones para el conjunto de entrenamiento: Aplica un volteo horizontal aleatorio a las imágenes, convierte la imagen a un formato de imagen a un tensor PyTorch y normaliza los valores de los píxeles del tensor para cada canal de color, que en este caso se manejan 3 canales (RGB) con los valores de media y desviación estándar del modelo preentrenado con el conjunto de ImageNet que son: mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225].
- Transformaciones para el conjunto de validación: Convierte la imagen a un tensor de PyTorch y normaliza las imágenes de la misma manera que el conjunto de entrenamiento.

5.3 Arquitectura

La arquitectura residual ResNet50, es una arquitectura que emplea un diseño de cuello de botella en cada bloque residual. En el que cada bloque residual, utiliza secuencias convolucionales de diferentes tamaños para reducir y luego restaurar la dimensionalidad. Para este tipo de clasificación con imágenes GEI, la ResNet50 empleada tuvo la siguiente forma:

- Capa de entrada
 - Capa inicial: convolución 7×7 , BatchNorm, ReLU y max pooling
- Bloque bottleneck: Son bloques residuales
 - Primera capa convolucional de tamaño 1×1 para la reducción de dimensionalidad con un BatchNorm2d.
 - Segunda capa de un kernel de 3×3 con un stride y BatchNorm2d.
 - Tercera capa convolucional con un tamaño de kernel de 1×1 para restaurar la dimensionalidad que contiene un BatchNorm2d, stride y downsample que permite cambiar el tamaño.

- Capa de salida
- Capa final: Global Average Pooling y capa Fully Connected

En la Figura 12 del capítulo de Antecedentes se presenta el diagrama de la arquitectura general de ResNet50. Debido a que se modificaron algunas capas de la arquitectura base de ResNet50, la Figura A.1 de Anexos se desglosan las capas a mayor detalle.

5.5. Estudio de ablación

Se realizó un estudio de ablación cambiando las características de los hiperparámetros de las arquitecturas para analizar y comprender la importancia y contribución de los diferentes componentes de la red al rendimiento global del sistema. El estudio de ablación se centró en modificar de manera controlada las funciones de activación y los hiperparámetros como el tamaño de lote de las imágenes de entrenamiento y validación, el número de capas convolucionales de la red, tipo de optimizador, tasa de aprendizaje, decaimiento del peso y patience.

Este estudio de ablación se enfocó principalmente en la evaluación del rendimiento de diferentes arquitecturas una vez comparada las arquitecturas base propuestas: VGG16, ResNet50 e Inceptionv3, las cuales en la Tabla 2, se detalla los hiperparámetros aplicados. En esta tabla, se denota que, al ser diferentes arquitecturas de CNN, el número de capas convolucionales cambia. Para la VGG16 es de 16 capas como indica su nombre, 50 y 48 capas para la ResNet50 e InceptionV3 respectivamente. La tasa de aprendizaje fue significativamente diferente variando de 0.001 a 0.0001, como también el tipo de activación fue el mismo para los tres modelos. Para este último influyó en que parte de las capas se quería la clasificación, la activación con Softmax es utilizada mayormente en la capa de salida de una red cuando el objetivo es una clasificación multclases, puesto que, convierte las salidas en probabilidades que suman 1, facilitando la interpretación de la predicción como fue aplicada para la VGG16. Sin embargo, la activación con la función de rampa o ReLu es más específica para introducir no la linealidad al emplearla en las capas ocultas de las redes. Adicional a esto, ayudan a superar los problemas de desvanecimiento del gradiente durante el entrenamiento.

Por otra parte, la elección de los optimizadores se basó en el cómo calculan y ajustan las tasas de aprendizaje adaptativas para los parámetros establecidos. El RMSprop ajusta la tasa de aprendizaje adaptativo para cada. Esto permite que la tasa de aprendizaje se mantenga estable a lo largo del entrenamiento, en especial cuando es una parte de una red preentrenada. Además, al entrenar las capas congeladas, el comportamiento de la pérdida cambia respecto al descongelamiento de las capas. A diferencia del optimizador Adam, el cual ajusta la tasa de aprendizaje de cada parámetro individualmente. Incluyendo una corrección de sesgo para estas estimaciones, al mejorar la estabilidad y la convergencia en las primeras etapas de entrenamiento evitando oscilaciones bruscas. Debido a esto, es que este optimizador se caracteriza por converger más rápido y con un mejor rendimiento. Finalmente, el tamaño del lote de entrenamiento y validación se consideró de la misma magnitud de 32 para no dar pauta a índices de sobreajuste.

Hiperparámetros	Arquitectura		
	VGG16	ResNet50	Inceptionv3
Número de capas convolucionales	16	50	48
Tasa de aprendizaje	0.0001	0.001	0.0001
Tipo de activación	ReLU	ReLU	ReLU
Tamaño del lote de entrenamiento	32	32	32
Tamaño del lote de validación	32	32	32
Optimizador	Adam	Adam	RMSprop

Tabla 2: Hiperparámetros utilizados por arquitectura.

A partir de los hiperparámetros propuestos en la Tabla 2 de acuerdo con mi estudio de ablación, se observó que los resultados fueron satisfactorios conforme al tipo de imagen de entrada, la arquitectura que tenía un mejor comportamiento y los resultados más deseables era la ResNet50. Debido a esto, se realizó otro estudio de ablación en el que solo se enfocó en estudiar a diferentes configuraciones de ResNet, es decir, una variación entre sus capas convolucionales, la tasa de aprendizaje y los tamaños de lote para entrenamiento y validación. En la Tabla 3, se observan las variaciones del modelo ResNet. En la que se puede notar que los lotes de entrenamiento y validación van variando, puesto que, al tener una brecha mayor entre ambos conjuntos y el número de capas convolucionales incrementaban, el sobre ajuste era menor y alcanzaba un validación cercana a 1.

Hiperparámetros	Arquitectura					
	ResNet-34	ResNet-34	ResNet-50	ResNet-50	ResNet-101	ResNet-152
Número de capas	34	34	50	50	101	152
Tasa de aprendizaje	0.001	0.001	0.001	0.001	0.001 con LRsccheduler	0.001 con LRsccheduler
Patience	–	–	–	–	3	5
Decaimiento del peso	–	1e-4	–	1e-4	1e-4	1e-4
Tipo de activación	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU
Lote de entrenamiento	32	64	32	32	32	32
Lote de validación	32	128	32	128	128	128
Optimizador	Adam	Adam	Adam	Adam	Adam	Adam

Tabla 3: Hiperparámetros utilizados por arquitectura de ResNet.

5.6. Estructuración de datos de coordenadas

Se estructuraron los datos de las coordenadas de cada uno de los individuos para tener una mejor manipulación y crear sub bases de datos.

5.1 Material

Los recursos computacionales que soportan la implementación de la estructura de datos y el análisis tanto de la simetría como el estadístico, se describen a continuación.

- Computadora:
 - Windows 11
 - Procesador Intel(R) Core(TM) Ultra 7
 - Sistema operativo de 64 bits
- Software:
 - Python 3.12.2
- Librerías:
 - JSON 2.0.9
 - NumPy 1.26.4
 - Pandas 2.2.1
 - Matplotlib 3.8.3
 - Seaborn 0.13.2
 - scikit-learn 1.4.1.post1
 - SciPy 1.12.0

5.6.1. Verificación de datos

Se hizo una verificación de individuos que contaran con todo los ángulos de estudio tanto para las imágenes como para las coordenadas 2D. Obteniendo así un total de 4000 individuos como estudio de prueba y 1000 para validación, en el caso de la clasificación. Por parte de detección de anomalías de datos mediante las coordenadas. Se realizó una segunda validación de datos para crear una nueva base de datos con la información de los individuos como es la edad, género y una verificación true/false para identificar que se tenían todos los ángulos a estudiar en cada número de identificador (ID). A continuación se en listan los pasos seguidos:

1. Se verificó que cada uno de los IDs o individuos tuvieran la información de edad, género y si se contaba con la información en los ángulos 15° , 30° , 45° , 60° , 75° , 90° y 180° . Esto con el fin de crear un DataFrame con los datos correspondiente por cada uno de los IDs. La validación se realizó con un etiquetado de True y False.

2. Una vez realizado la primera verificación de datos, se prosiguió a filtrar los etiquetados con la validación de que todas las columnas fueran True, en caso contrario se eliminaban los IDs. Finalmente, como última comprobación se revisó que todos los individuos tuvieran la columna con la edad, de no ser así se eliminaba.
3. La validación final se guardó en un archivo en formato CVS (valores separados por comas) nombrada como validacion_completa. En la Figura se refleja el archivo creado de acuerdo a las especificaciones planteadas.
4. Una vez generado el archivo de CVS de validación de ID, se contaron 4000 individuos como datos de análisis. Teniendo en cuenta que coincidieran con las imágenes GEI ocupadas en el apartado de clasificación.

5.6.2. Identificador de los individuos

El archivo generado con la verificación de los IDs permitió crear una nueva de base de datos llamada identificador de los individuos, en el que tomaba los IDs válidos para extraer la información de las carpetas, creando así un DataFrame con las columnas de ID, género y edad tal como se muestra en la Figura 31b.

	A	B	C	D	E	F	G	H	I	J
1	ID	gender	age	15	30	45	60	75	90	180
2	2	F	22	True						
3	3	F	26	True						
4	4	M	14	True						
5	5	M	13	True						
6	7	M	13	True						
7	8	M	8	True						
8	9	M	13	True						
9	10	M	13	True						
10	11	M	18	True						
11	12	M	18	True						
12	13	M	19	True						
13	15	M	18	True						
14	16	M	19	True						
15	17	M	19	True						

(a) Archivo CSV de la verificación de cada ID.

	A	B	C
1	ID	gender	age
2	00002	F	22
3	00003	F	26
4	00004	M	14
5	00005	M	13
6	00007	M	13
7	00008	M	8
8	00009	M	13
9	00010	M	13
10	00011	M	18
11	00012	M	18
12	00013	M	19
13	00015	M	18
14	00016	M	19
15	00017	M	19

(b) Archivo CSV de la información de cada ID.

Figura 31: Bases de datos generadas en CSV. Imágenes de autoría propia.

5.6.3. Análisis de simetría

El análisis de simetría se dividió en tres bases de datos diferentes. Tal como se observa en el esquema de la Figura 32, se divide en tres procesamientos de datos:

- **Análisis de simetría para articulaciones específicas de manera individual:** Consiste en analizar de manera individual la simetría de las coordenadas del hombro, cadera, rodilla, tobillo y pie. Considerando que el fotograma está en un plano sagital, permite obtener medidas de su altura, equilibrio y distancias entre las dos piernas y de la cabeza a la base de la columna vertebral. Debido a que sólo se escoge un fotograma en este plano, se considera que es una medida estable y por ende es la medida inicial a la que comparar.
- **Análisis por fotograma (frame) de cada ángulo:** Analiza por cada fotograma del individuo en sus diferentes ángulos, creando un archivo por ángulo y un resumen general de simetría. En esta base se guardan las medidas de simetría para el hombro, cadera, rodilla, tobillo y pie. Además de la distancia entre las dos piernas. En la parte del resumen se obtiene un promedio y una desviación estándar por cada tipo de simetría, es decir, la simetría de las articulaciones a estudiar.
- **Análisis de medidas globales:** Esta base de datos se enfoca en la simetría general del individuo, es decir, considerando solo las articulaciones correspondientes del lado derecho e izquierdo, con el fin de analizar la asimetría en ambos lados. Debido a que guarda valores como la simetría de la marcha y su normalización, así como el promedio de la distancia entre la cabeza y la base de la columna toma en cuenta todos los fotogramas de los individuos.

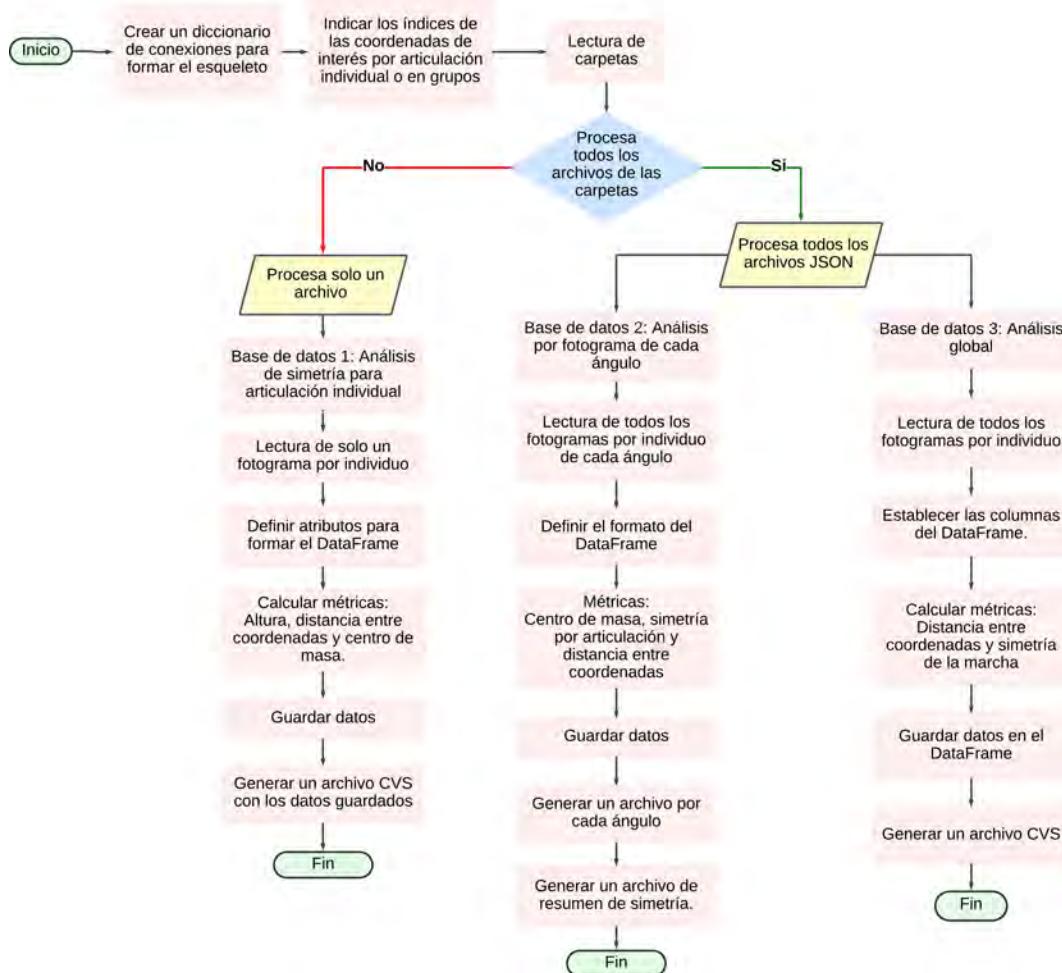


Figura 32: Metodología general para la construcción de bases de datos.

Estas bases de análisis siguieron los siguientes pasos:

1. Importación de librerías
2. Crear un diccionario de conexiones para el esqueleto de acuerdo con la Figura 19.
3. Indicar los índices de las coordenadas de interés, es decir, las articulaciones a estudiar: cabeza (head), pie derecho (Rfoot), pie izquierdo (Lfoot), base espinal (spine_base), hombro derecho (Rshoulder), hombro izquierdo (Lshoulder), cadera derecha (Rhip), cadera izquierda (Lhip), rodilla derecha (Rknee), rodilla izquierda (Lknee), tobillo derecho (Rankle) y tobillo izquierdo (Lankle).
4. Crear una lista con los índices de las articulaciones para el lado derecho e izquierdo: right_idx, left_idx.

5. Crear dos listas para indicar los segmentos que corresponden a las piernas derecha e izquierda respectivamente: Rleg_idx y Lleg_idx que comprende a la cadera, rodilla, tobillo y pie.
6. Importación de carpetas en las que se encuentran todas las carpetas de las coordenadas y definir los ángulos con los nombres de las carpetas que se van a analizar: “015_01”, “030_01”, “045_01”, “060_01”, “075_01”, “090_01”, “180_01”.

7. Análisis de simetría para articulación individual:

- a. Se leen solo las carpetas “180_01”, el cual hace referencia a un plano anatómico: plano sagital.
- b. Se definen los atributos del DataFrame a crear: ID (identificador del individuo), Height, D H-SB (distancia de la coordenada cabeza a las base espinal), Balance (equilibrio), SI shoulder (simetría del hombro), SI hip (simetría de la cadera), SI knee (simetría de la rodilla), SI ankle (simetría de la rodilla), SI foot (simetría del pie) y D legs (distancia entre las piernas). Los cuales van a medir el esqueleto del individuo en su forma más estable que es en un plano estático y además los dos pies se encuentran sobre el piso, obteniendo así información de su equilibrio, altura, simetría en un estado estable y distancia entre sus hombros para ver si hay alguna desviación.
- c. Leer los archivos JSON para cada uno de los IDs mediante un for. Consistió en leer solamente el primer archivo en el cual el individuo a estudiar cumplía con la regla de que los dos pies se encontraran en el piso y la posición de su postura fuera estable. Además, es el primer fotograma, para ellos se siguieron los siguientes pasos:
 - I. Inicializar listas que guardan las métricas para cada carpeta ID
 - II. Extraer las coordenadas de las articulaciones de interés: head, Rfoot, Lfoot, spine_base, Rshoulder, Lshoulder, Rhip, Lhip, Rknee, Lknee, Rankle y Lankle.
 - III. Calcular métricas: altura, distancia entre la cabeza y la base de la columna, centro de masa (COM) mediante el promedio de todas las coordenadas, equilibrio, distancia entre los dos pies y la distancia entre las dos piernas usando las coordenadas de Rleg_idx y Lleg_idx; y la simetría para 5 articulaciones específicas: hombros, cadera, rodillas, tobillo y pie.
 - IV. Guardar en el DataFrame.
 - V. Guardar resultados en un archivo Excel con extensión CVS.

8. Análisis por fotograma de cada ángulo:

- a. Se requieren de todas las carpetas de los ángulos con todos los fotogramas que tenga la carpeta por individuo de acuerdo con su ángulo. Aproximadamente hay de 25 a 30 fotogramas por ángulo. Es decir, se tienen: ≈25

fotogramas \times 7 ángulos \times 4000 individuos, obteniendo así un total 700,000 datos aproximadamente.

- b. Se crea un diccionario para guardar los datos por ángulo.
- c. Se empieza iteración por cada carpeta de ID.
 - I. Reinicia listas por subcarpetas por ID y por ángulo, tomando solo las articulaciones correspondientes a hombro, cadera, rodilla, tobillo y pie.
 - II. Extracción de las coordenadas de las articulaciones.
 - III. Calcular métricas: Centro de masa, simetría por articulación y la distancia entre las piernas.
 - IV. Guardar los valores por fotograma en las listas.
 - V. Se crea un resumen por ID y ángulo. En el cual se van guardando el promedio y la desviación estándar para cada tipo de simetría que corresponde a las articulaciones mencionadas.
- d. Al terminar la iteración de cada ángulo, convierte los datos a DataFrame y se guardan en un archivo con extensión CVS.
- e. Se crea un DataFrame para el resumen de simetría y se guarda en un archivo CVS.

9. Análisis global:

- a. Se establecen las columnas del DataFrame para las medidas globales, que involucra todos los fotogramas que tenga el ID (individuo), es decir: n fotogramas \times 7 ángulos. Las columnas establecidas son: ID, Gait_Symmetry (simetría de la marcha), CC_symmetry_norm (simetría de la marcha normalizada), Avg_H_BS (promedio de la distancia de la cabeza a la base espinal de la columna vertebral), Avg_Euclidean_Distance (H_BS) (promedio de la distancia de la cabeza a la base espinal pero con la distancia euclíadiana).
- b. Se empieza a iterar sobre carpetas por ID con un ciclo for.
 - I. Se establecen las listas iniciales para almacenar los valores de las métricas y un contador: x_r_list (guarda los valores en x de las coordenadas del lado derecho), x_l_list (para coordenadas del lado izquierdo), y_distance_list (guarda la distancia medida entre la coordenada de la cabeza y la base espinal), euclidean_distance_list (distancia euclíadiana entre la cabeza y la base de la columna), frames_list (guarda el número de fotogramas por cada individuo).
 - II. Extracción de información de las coordenadas de acuerdo con los índices de interés, para este análisis se utilizaron las articulaciones (coordenadas) del lado derecho, lado izquierdo, cabeza (índice 3) y base de la columna (índice 3).
 - III. Calcular métricas: distancia entre la cabeza y la base espinal con método tradicional y con distancia euclíadiana.

- IV. Guardar los datos en las listas correspondientes.
- V. Se comprueba que las listas `x_r.list` y `x_l.list` tengan un formato de array y se encuentren en las carpetas. Posterior, se aplica un aplanamiento de vectores para obtener el centro de las coordenadas según el lado derecho e izquierdo.
- VI. Se normaliza los vectores obtenidos del aplanamiento y emplea la fórmula de simetría de la marcha.
- VII. Se obtiene la simetría global normalizada.
- VIII. Se guarda en el DataFrame.
- c. Guardar resultados en un archivo CVS.

5.1 Cálculo de métricas

El cálculo de las métricas para las bases descritas fue enfocado en la parte del esqueleto, teniendo como entrada los datos de las posiciones de las coordenadas. Las fórmulas explicadas fueron recuperadas del artículo titulado “Extracting gait and balance pattern features from skeleton data to diagnose attention deficit/hyperactivity disorder in children” [70]. En el cual hace referencia a un estudio de niños con TDAH (trastorno por déficit de atención e hiperactividad) en el que extrajeron características del esqueleto automáticamente mediante fórmulas que se destacan por evaluar la función motora de un niño en ese estudio.

5.6.1 Altura

Como se define en [70], la ecuación 7 expresa la altura medida desde la cabeza hasta el pie más bajo tomando en cuenta sólo las coordenadas en y . Debido a que la altura es medida hasta los pies, se encontró el valor mínimo entre los dos elementos: Rfoot(pie derecho) y Lfoot(pies izquierdo), este punto fue sustituido en y_{Pie} . Para y_{Cabeza} se sustituye por la coordenada en y de head (cabeza).

$$H = |y_{Cabeza} - y_{Pie}| \quad (7)$$

5.6.2 Distancia euclíadiana

La distancia euclíadiana, es la distancia medida entre dos puntos. Además, es una de las propiedades estáticas al emplearla en el análisis del esqueleto, ya que, se obtiene en función de la distancia entre articulaciones. La ecuación 8 fue adaptada de [70].

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (8)$$

Donde d_{ij} es la distancia euclíadiana de dos articulaciones i y j , calculada en dos dimensiones X e Y .

5.6.3 Equilibrio

Tal como señala Rohani et al. en [70], el equilibrio es una medición de pérdida de equilibrio durante un movimiento, esto quiere decir, que perder el equilibrio afecta de forma negativa la evaluación del rendimiento. Los autores en [70], miden la distancia entre el centro de masa del cuerpo (barycenter) y el centro de una zona rectangular formada por los pies del individuo en plano tridimensional. Debido a que es una ecuación adaptada de Rohani en [70] donde se cambia el plano de proyección a x e y , se tomaron en cuenta tres aspectos:

- Se verifica que en el fotograma, ambos pies estén tocando el piso. Debido a esta verificación sólo requirieron los primeros fotogramas de cada individuo del ángulo de 180° , que es cómo si fuese un plano sagital.
- Se toman las coordenadas de los marcadores de los pies: Rfoot y Lfoot.
- Se obtiene el centro de masa para obtener las coordenadas del centro del cuerpo (x_b, y_b).

El centro de masa (CoM por sus siglas en inglés) es definido por McGinnis como un punto de un cuerpo en el que se puede suponer que toda la masa está concentrada [2]. Este centro de masa es usado como referencia para evaluar que tanto esta equilibrando el cuerpo con respecto a sus lados izquierdo y derecho. Permitiendo calcular las coordenadas del centro del cuerpo mediante CoM, el cual se puede calcular biomecánicamente con el promedio de todas las articulaciones (joints). Esto consiste en promediar las 24 articulaciones del esqueleto como se mostró en la Figura 19, aplicado para las coordenadas en x e y , obteniendo así la siguiente ecuación:

$$X_{COM} = X_B = \frac{1}{n} \sum_{j=1}^n x_j \quad (9)$$

Donde X_{COM} corresponde al promedio de todas las articulaciones en x , n es el número de articulaciones y j representa el índice. La misma configuración es aplicada para las coordenadas en y . Teniendo como resultado un vector $[x, y]$ del centro de masa del cuerpo de la forma: $X_B = X_{COM}$ e $Y_B = Y_{COM}$.

De acuerdo con lo planteado en la ecuación 9, la expresión matemática en la ecuación 10, se denota la varianza de la distancia del centro del soporte entre ambos pies: $(x_{derecha}, y_{derecha})$ para Rfoot y $(x_{izquierda}, y_{izquierda})$ para Lfoot al centro de masa (x_b, y_b) .

$$Equilibrio = \text{var} \left(\sqrt{\left(\frac{x_{derecha} - x_{izquierda}}{2} - x_b \right)^2 + \left(\frac{y_{derecha} - y_{izquierda}}{2} - y_b \right)^2} \right) \quad (10)$$

5.6.4 Simetría

La simetría mide la simetría corporal durante un movimiento, evaluando que el cuerpo esté alineado de forma equilibrada entre su lado derecho e izquierdo [70]. La simetría puede ser medida para cada parte del cuerpo o en áreas de interés, en dónde se compara la posición del centro de masa con las partes del lado derecho e izquierdo. La ecuación 11 y 12 fueron adaptadas de [70], cada una de las ecuaciones mencionadas miden la simetría por el tipo de eje.

$$SI_{xi} = \frac{(X_B - X_{Li}) - (X_B - X_{Ri})}{X_{Ri} - X_{Li}} \quad (11)$$

$$SI_{yi} = \frac{(Y_B - Y_{Li}) - (Y_B - Y_{Ri})}{Y_{Ri} - Y_{Li}} \quad (12)$$

Donde x_b es la coordenada del centro de masa en la dimensión x , x_{Li} corresponde a la coordenada del i-ésimo punto del lado izquierdo de una de las articulaciones, por ejemplo, la rodilla izquierda y x_{Ri} es la coordenada del lado derecho de la misma articulación, por ejemplo, la rodilla derecha. Esto mismo aplica para las coordenadas de la dimensión y .

Por otra parte, se puede obtener una sola medida de la simetría mediante el promedio de SI_{xi} y SI_{yi} como se manifiesta en la ecuación 13.

$$SI = \frac{SI_{xi} + SI_{yi}}{2} \quad (13)$$

5.6.5 Simetría de la marcha

Según lo documentado por Rohani et al. [70], la simetría de la marcha es mayormente asumida como la función idéntica de la locomoción entre los lados derecho e izquierdo del cuerpo humano y su cambio, es decir, la asimetría de la marcha, pueden ser hallados al examinar la diferencia significativa entre ambos lados, un ejemplo claro, es las extremidades inferiores. Es decir, mide que tan similar son los movimientos de las piernas derecha e izquierda durante la locomoción. La ecuación 14 representa la simetría de la marcha mediante un producto punto entre los datos del lado derecho e izquierdo, donde n es el número de todos los fotogramas, $x_{r(n)}^i$ e $y_{l(n)}^i$ son los valores de los parámetros de las extremidades derecha e izquierda respectivamente en el fotograma n .

$$CC_{\text{simetría}} = \sum_{n=1}^N x_{r(n)}^i \cdot x_{l(n)}^i \quad (14)$$

Como se mencionó anteriormente, la simetría de la marcha puede sufrir desviaciones entre extremidades: la asimetría. Término que se asocia a la divergencia del lado derecho e izquierdo del cuerpo [75]. Debido a esto, se aplicó una normalización de datos para la simetría de la marcha. La normalización por norma euclídea también conocida como normalización L2, la cual se centra en escalar un vector de tal manera que su magnitud

sea 1 [76]. Esto lo realiza a través de dividir cada componente del vector por su norma euclíadiana, que es la raíz cuadrada de la suma de los cuadrados de los componentes del vector, encontrando la distancia más corta desde el vector hasta el vector cero [76], [77]. En la ecuación 15, obtenida de [77], se plasma la norma euclíadiana representada como $\|x\|_2$ donde x_1, x_2, \dots, x_n son los componentes del vector x .

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (15)$$

Donde \vec{v}_{norm} es el vector concatenado de las características para cada uno de los lados: derecho e izquierdo, $\|\vec{v}\|_2$ es el vector normalizado en ese lado, siguiendo la expresión de la ecuación 15. La ecuación 16, transforma cada vector a un vector unitario en la misma dirección, pero de una longitud o magnitud de 1 como Khosravy et al. [77].

$$\vec{v}_{\text{norm}} = \frac{\vec{v}}{\|\vec{v}\|_2} = \frac{\vec{v}}{\sqrt{\sum_{i=1}^n v_i^2}} \quad (16)$$

5.7. Análisis estadístico

Para el análisis estadístico se realizó una vez generado todas las bases de datos mencionadas en el apartado de análisis de la simetría. El cual permitió aplicar un conjunto de técnicas y métodos para organizar, describir, interpretar y analizar los datos de las métricas del esqueleto para extraer información significativa que pudiera ayudar a hacer la clasificación entre individuos con una mala o buena postura tomando en cuenta las mediciones medidas por fotogramas, medidas globales de simetría de la marcha por individuo y medidas globales por ángulo en cada individuo. Al igual hacer un censo de la población al medir los rangos de edad y el género para entender la estructura demográfica de la población a estudiar. En el esquema de la Figura 33, se detallan los pasos generales, tomando en cuenta que el análisis se basa en variables cuantitativas.

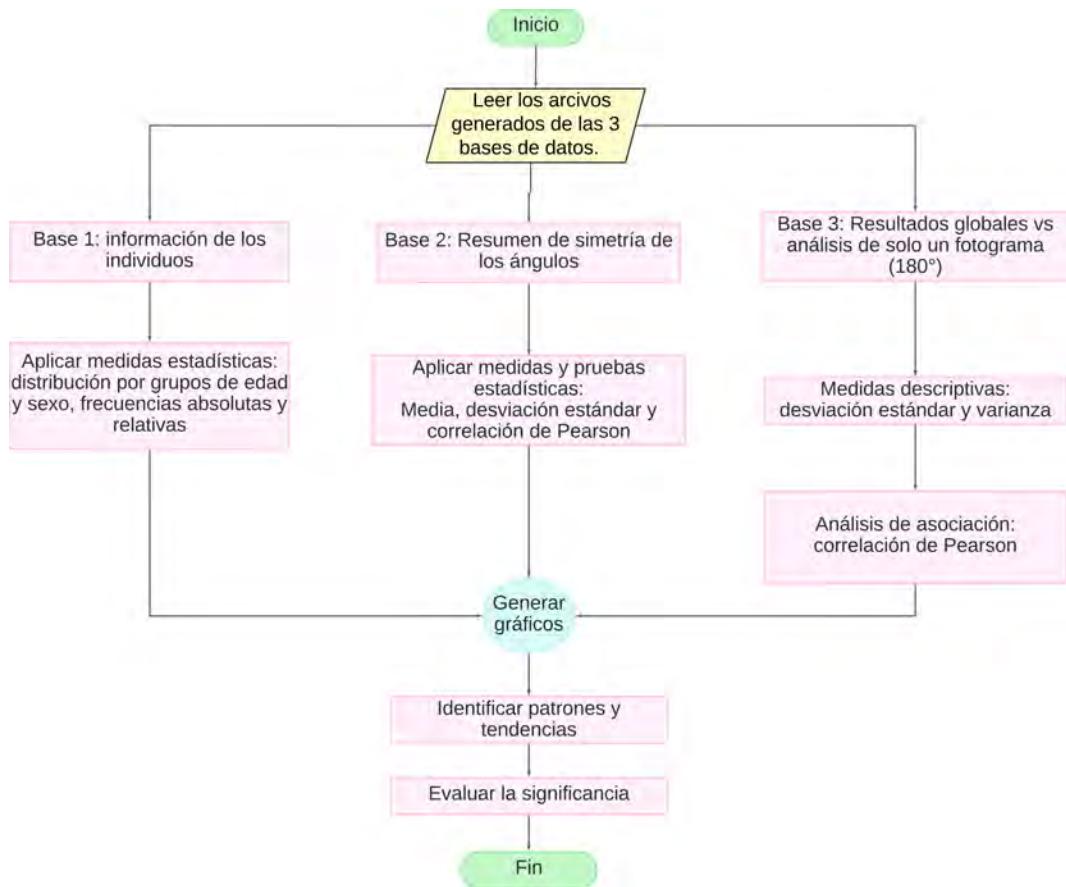


Figura 33: Metodología general del análisis estadístico.

5.7.1. Métricas

Las métricas descritas a continuación fueron empleadas de acuerdo con la base de datos, tal como se señala en el diagrama de la Figura 33.

5.1 Indicadores demográficos

- Distribución por grupos de edad y sexo de la población de estudio: El cual clasifica la población en grupos, separando por género, permitiendo analizar la composición poblacional como se observa en la Figura 34.
- Frecuencias absolutas y relativas: Calcula el número total de individuos en cada grupo de edad y sexo que son las frecuencias absolutas, las cuales se representan como porcentaje o proporciones respecto al total de la población como lo ilustra la Figura 35.

En la Figura 34 se observa la distribución de la población de 4000 individuos para analizar sus esqueletos. Se observa que, de los rangos establecidos por edad, el género

que predomina más es el sexo masculino, tomando en cuenta que las edades van de 0 a más de 76 años. La frecuencia absoluta para el grupo de 0-18 años es 673 para hombres y 536 para mujeres, siendo este grupo la predominancia del sexo masculino, sin embargo, el sexo femenino predomina en los rangos de 19-30 años y de 31-45 años con un índice de 600 y 610 respectivamente.

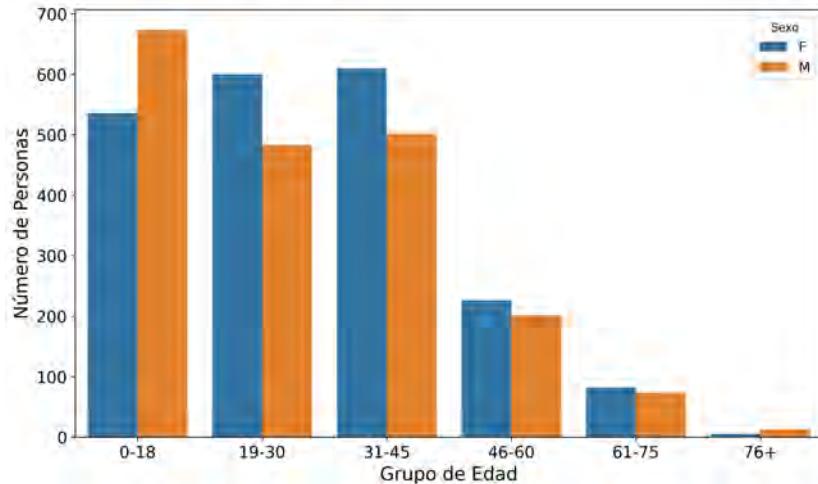


Figura 34: Distribución por grupo de edad y sexo. Imagen de creación propia

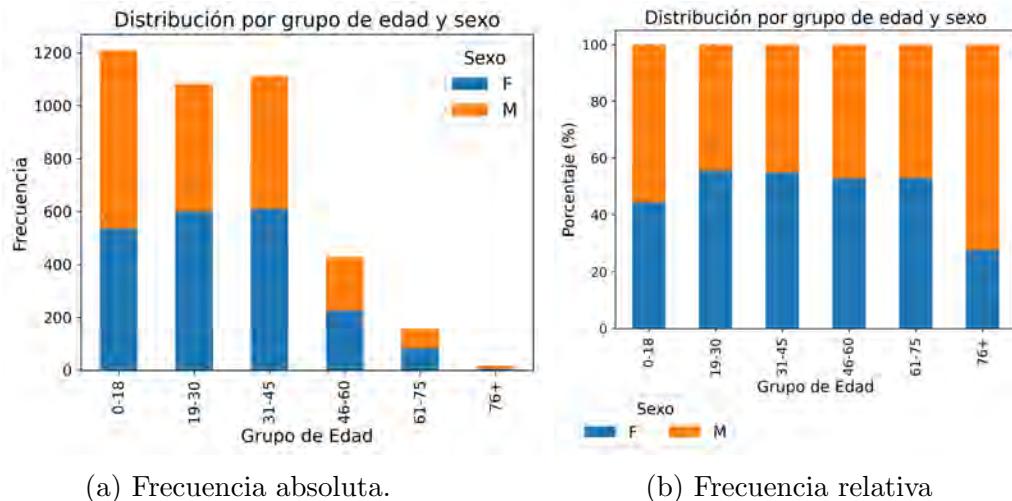


Figura 35: Frecuencia absoluta y relativa de la distribución de edad y sexo. Imágenes de autoría propia

5.2 Media

Es el promedio de un conjunto de mediciones. La media de una población se denota como μ , en el cual un conjunto de n mediciones es igual a la suma de las mediciones

medidas entre n . En la ecuación 17 de acuerdo con Mendenhall et al. [78], se denota su expresión matemática donde x_i son las n variables a sumar.

$$\mu = \frac{\sum(x_i)}{n} \quad (17)$$

5.3 Varianza

La varianza de una población se denota como σ^2 . Es el promedio de los cuadrados de las desviaciones de las mediciones alrededor de su media μ , donde N es el número de mediciones en la población y esta dada por la siguiente fórmula obtenida de [78].

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} \quad (18)$$

5.4 Desviación estándar

La desviación estándar es definida como una medida de variación o medida de dispersión de todos los valores respecto a la media [78]. Una desviación estándar s suelen ser valores positivos y sólo es igual a cero cuando todos los valores de los datos son del mismo número. Adicional a esto, un valor grande de s indicaría una mayor cantidad de variación. La fórmula requerida para el tipo de datos fue la desviación estándar de una población representada como σ . De acuerdo con Mendenhall et al. [78], la desviación estándar es la raíz cuadrada positiva de la varianza como se denota en la ecuación 19.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (19)$$

Donde σ es la desviación estándar poblacional, N el tamaño de la población, x_i los valores de la población y μ la media.

5.5 Correlación de Pearson

La correlación de Pearson es un método estadístico que mide la similitud o correlación entre dos variables de dos conjuntos de datos, calculando valores de -1 a $+1$. Un índice alto de correlación indica una similitud alta, mientras que un valor bajo cercano a cero es un indicio de ausencia de correlación. En la ecuación 20 se observa que se calcula entre dos variables con atributos pareados, que consiste en la suma del producto de sus diferencias con las medias de las variables y divide la suma por el producto de las diferencias al cuadrado con las medias de las variables de acuerdo con Berman [79].

$$\text{Correlación de Pearson} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \cdot \sqrt{\sum(y_i - \bar{y})^2}} \quad (20)$$

6. Resultados y discusión

6.1. Resultados

Los modelos de clasificación VGG16, ResNet50 e InceptionV3, son arquitecturas que también permiten extraer características de las imágenes. Para la primera parte que consistió en extraer las características de las imágenes tipo GEI, se observó que al tratarse de una imagen en escala de grises y con un barrido de varias secuencias en una misma imagen, al aplicar métodos de preprocesamiento de imágenes como las que se mencionan en la metodología, las técnicas manuales no son suficientes para extraer completamente la silueta de la imagen. La zona de interés en la que se puede estudiar una detección de anomalías de una buena o mala postura según la pose de la persona basándose en solamente la imagen. Sin embargo, al no obtener resultados satisfactorios propiamente para el tipo de imágenes GEI, se optó por aplicar otro método. El cual son las redes neuronales convolucionales, arquitecturas que fueron mencionadas anteriormente en el capítulo 6.2.

La extracción de características con las arquitecturas de VGG16, ResNet50 e InceptionV3, permitieron analizar la imagen mediante filtros o mapas de características la imagen de entrada cómo también evaluar la efectividad de dicha extracción y la clasificación. En la Tabla 4, se presenta las clases correspondientes a su codificación usadas para las gráficas y reportes obtenidos.

Codificación	Clase (ángulo)
0	15°
1	30°
2	45°
3	60°
4	75°
5	90°
6	180°

Tabla 4: Codificación correspondiente a cada clase.

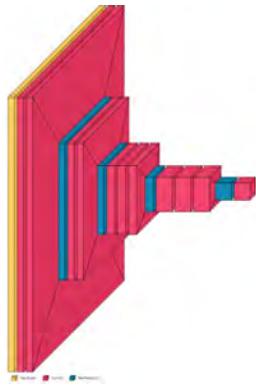
6.1.1. Extracción de características

A partir de tres arquitecturas distintas, se trajeron las características al entrenar todas las capas de los modelos estudiados. Sin embargo, se realizó una prueba al extraer filtros mediante una capa convolucional específica de la VGG16. Para este caso, se trajeron los filtros en la capa convolucional 5, es decir, los embeddings generada con: 14×14 dimensiones espaciales del mapa de características y 512 filtros que captura un tipo específico de característica visual.

La extracción de los filtros se empleó en los tres modelos estudiados. En la Figura 36a, se ilustra la arquitectura de un modelo de VGG16 hasta la c5 (capa convolucional 5: block5_conv1), una vez generado se selecciona una imagen de entrada, tal como se

muestra en la Figura 36b y a partir de la imagen de entrada se visualiza el mapa de características de dicha imagen. La cual corresponde a diferentes regiones de la imagen GEI. Los filtros generados se observan de la siguiente forma en la Figura 37a donde los píxeles son señalados en distintas texturas, abarcando tanto el fondo como parte de la silueta de la imagen GEI. En la Figura 37b, se superponen los mapas de características de la Figura 37a en la imagen GEI (Figura 36b), obteniendo en colores vibrantes las regiones específicas para cada filtro. En el cual es notorio que no lograr abarcar la suficiente información de la silueta de la imagen, sino también del fondo.

Los mapas de características de ResNet50 e InceptionV3 se indican en la Figura ?? y ?? respectivamente en el apartado de Anexos, donde se evidencia la extracción en diferentes capas. Para ResNet50 se ocupó la layer4 y para InceptionV3 mixed10. En ambos mapas de características es notorio que por filtros abarca parte de la forma de la silueta y del fondo, en unos más suavizado como es la capa de InceptionV3 en comparación con la ResNet50 que se enfoca en la silueta de la parte superior.



(a) Capas del modelo VGG16 de la extracción de características.

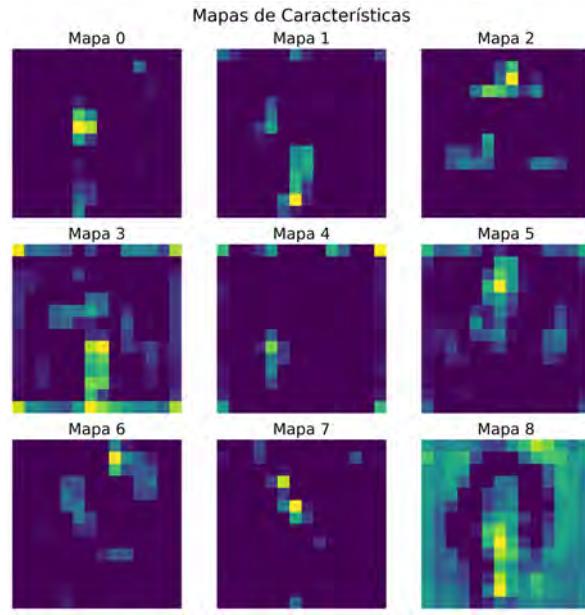


(b) Imagen de prueba en un ángulo de 15°.

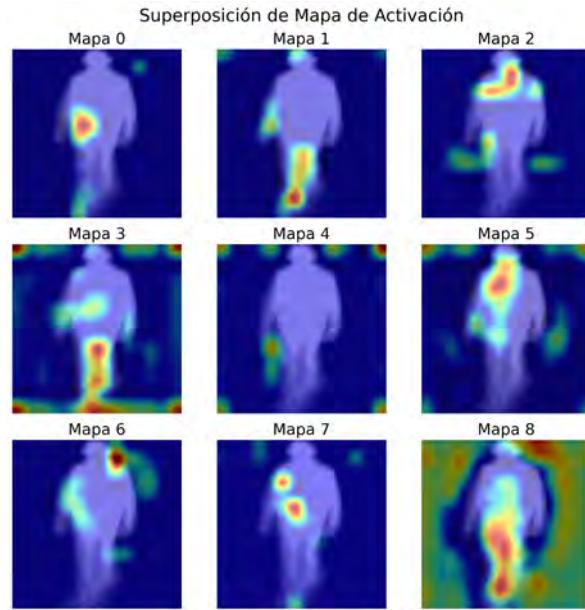
Figura 36: Imagen de prueba y arquitectura para la extracción de características de la capa 5 de VGG16.

Para el caso de la extracción de características de las tres arquitecturas entrenadas: VGG16, ResNet50 e InceptionV3. Se entrenaron todas las capas de convolución hasta la clasificación de las 7 clases que corresponden a los 7 ángulos que conforman un ciclo de marcha. Sin embargo, se seleccionaron capas específicas para la extracción, estas fueron: bolck5_conv1 para VGG16, layer4 en ResNet50 y mixed10 para InceptionV3. Estas características fueron medidas mediante las técnicas de PCA, LDA, CCA y CCR. Las cuales permiten evaluar la capacidad de los modelos de extraer características discriminativas y significativas de las imágenes como se ilustra en la Tabla 5.

El modelo que obtuvo un mayor índice de PCA es ResNet50 con 0.8429, que indica que captura casi en su totalidad la variabilidad de las imágenes originales y que a su vez mantienen la estructuración de los datos y las características extraídas son informativas en comparación con InceptionV3 que solo llega a capturar una tercera parte de los



(a) Mapa de características en diferentes regiones de la imagen.



(b) Mapa de características superpuesto en la imagen analizada.

Figura 37: Extracción de características de los patrones de textura y combinaciones de características de la capa block5_conv1 de VGG16.

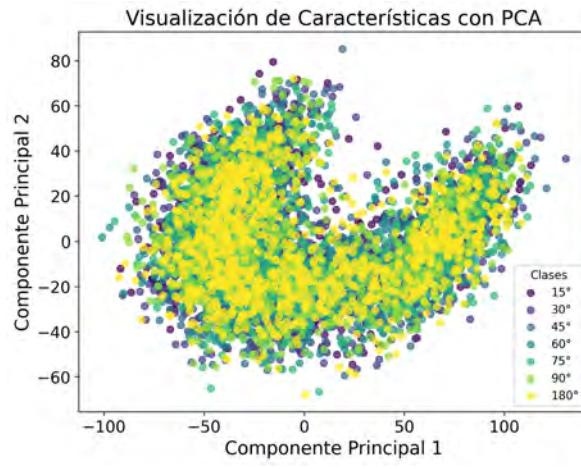
datos con un valor de 0.2544 y VGG16 con 0.4018. Las gráficas de PCA de los tres modelos de la Figura 38 muestran que tanto ResNet50 como Inceptionv3 tienen una

Modelo	Épocas	Exactitud	Pérdida	PCA	LDA	CCA	CCR
VGG16	30	0.9762	0.0661	0.4018	0.3322	0.2853	0.1462
ResNet50	50	0.9798	0.0893	0.8429	0.9990	0.9996	0.9990
Inceptionv3	20	0.9695	0.0833	0.2544	0.9977	0.9901	1.0

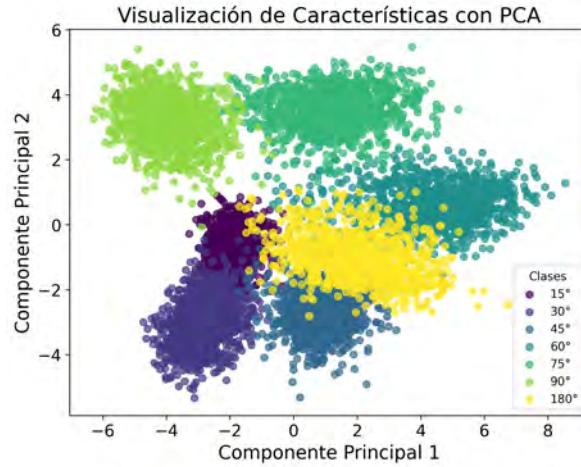
Tabla 5: Métricas finales de la exactitud, pérdida y extracción de características de los modelos de clasificación empleados.

reducción significativa en la dimensionalidad. El comportamiento de esta última tiene una proyección hacia abajo en donde las clases de los ángulos son diferenciadas (Figura 38c), sin embargo, no presenta una linealidad esperada y esto se debe principalmente a que los datos de las imágenes cuentan con estructuras más complejas al tratarse de sus texturas y que la mayoría de sus píxeles tienen intensidades parecidas, ocasionando que se superpongan las clases de 75°, 90° y 60° mayormente, puesto que, el ángulo de captura son muy parecidos y la posición de la silueta en estas tres clases presentan una variación mínima. En la parte media de la curva se superponen la clase de 30° con 45° debido a la poca variación del ángulo de captura como se aprecia en la ???. La Figura 38b expone que la proyección de PCA presenta agrupamiento para cada una de las clases, mostrando que captura similitudes de las características de los componentes principales 1 y 2. Adicional a esto, las clases de 15°, 30°, 45°, 60° y 180° llegan a superponerse con una fluctuación muy baja, a diferencia de la clase de 75° y 90° que presentan una separación marcada. Sin embargo, esto no sucede con la proyección de VGG16 mostrado en la Figura 38a, al tener una curvatura ligeramente inclinada hacia abajo sin relación lineal y sin separación de clases.

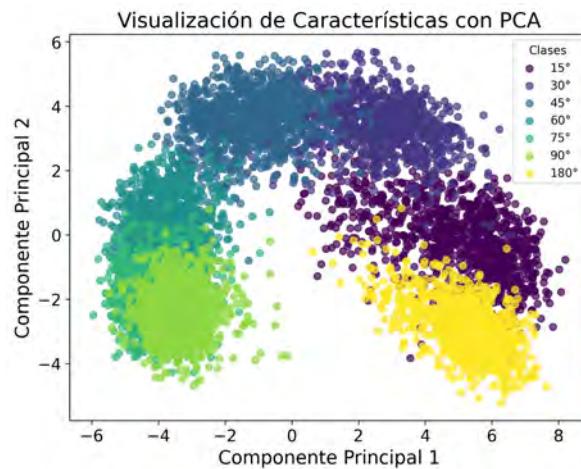
Por otra parte, LDA nos indica que tanto maximiza la separación entre las siete clases. El valor máximo obtenido fue de 0.9990 con ResNet50 y de Inceptionv3 con un valor de 0.9977, indicando que las características extraídas maximizan la distinción en casi todas las clases establecidas a diferencia de la VGG16 que no logra maximizar la separación de las clases al tener un valor de 0.3332. Las gráficas de LDA se observan en la Figura 39 donde ResNet50 (Figura 39b) e InceptionV3 (Figura 39c) presentan una clara dispersión entre clases con una proyección curveada hacia abajo donde los extremos son lineales. En ResNet50 se observa una separación bien definida entre las clases y además se visualiza una linealidad, sugiriendo que las características extraídas permiten diferenciar las clases de forma efectiva al identificar patrones relevantes para la tarea de clasificación. Para el caso de InceptionV3 es visible una separación del agrupamiento de las clases, sin embargo, la superposición se presenta entre los ángulos de 90°, 75° y 60°, tal como se observó en la gráfica de PCA (Figura 38c). Mientras que la proyección de VGG16, se superponen todas las clases y no se visualiza una separación de los ángulos. Este comportamiento sugiere que las características extraídas de los dos modelos con mayor índice de LDA (ResNet50 e InceptionV3) son más discriminativos para la clasificación de ángulos en imágenes tipos GEI.



(a) VGG16.

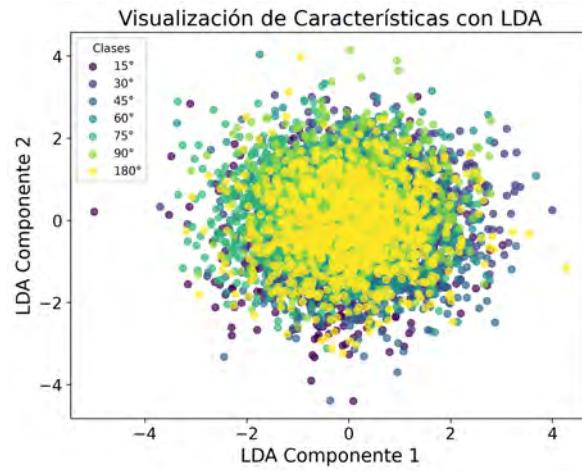


(b) ResNet50.

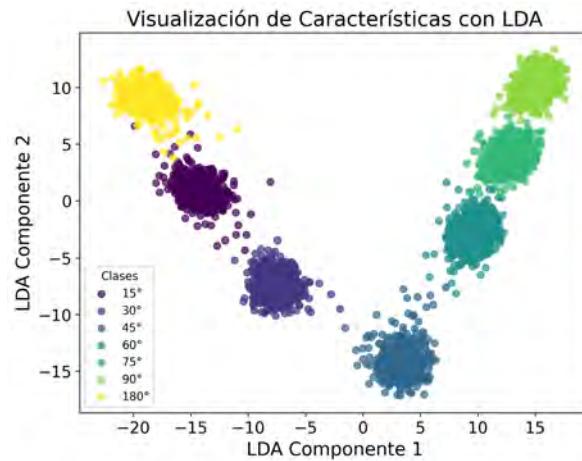


(c) InceptionV3.

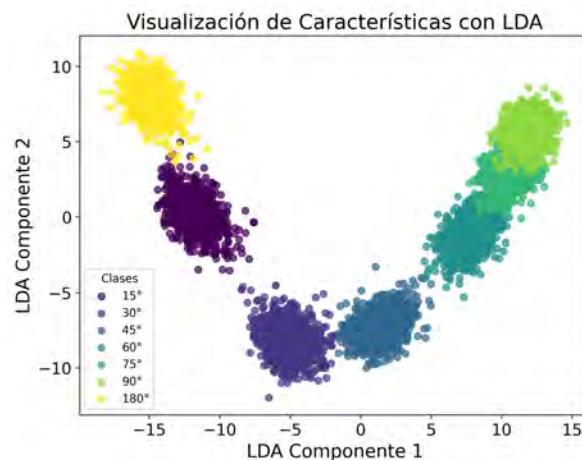
Figura 38: Gráficas de PCA de los modelos entrenados.



(a) VGG16.



(b) ResNet50.



(c) InceptionV3.

Figura 39: Gráficas de LDA de los modelos entrenados.

El análisis de componentes canónicos (CCA) busca la relación lineal entre las etiquetas y los embeddings, es decir, una correlación entre las etiquetas y las proyecciones de las características extraídas. Un índice alto, refleja una correlación alta, es decir, las características son informativas para la clasificación. El modelo con mayor magnitud de resultado fue nuevamente la ResNet50 con 0.9996 y 0.9901 para InceptionV3, con VGG16 mostrando una correlación baja de 0.2853 como se ilustra en la Tabla 5. Estos resultados reflejan, que los primeros dos modelos se encuentran mejor alineados con las etiquetas, mientras que VGG16 podría necesitar ajustes para la captura de las relaciones entre las clases y las características extraídas. Finalmente, el CCR, se centra en la correlación de vectores de características en lugar de conjuntos de variables como lo hace CCA. Un CCR alto, indica una interrelación fuertemente entre las características extraídas. InceptionV3 tuvo un valor de 1 y 0.9990 para ResNet50, lo que indica que las características extraídas en estos dos modelos son correlacionadas y coherentes. Mientras que VGG16 tuvo un CCR de 0.1462, indicando que la interrelación entre las características no es fuerte y por ende no se puede obtener una relación.

6.1.2. Clasificación

Como se menciono con anterioridad la extracción de características forman parte de la clasificación de los modelos VGG16, ResNet50 e Inceptionv3. A continuación, en la Tabla 6 y la Figura 40 se muestra el rendimiento de los modelos de clasificación y las predicciones por clase respectivamente. Tomando en cuenta que el conjunto de validación fue de aproximadamente 1000 individuos por cada una de las clases.

La matriz de confusión revela que el modelo ResNet50 (Figura 40c) presenta una diagonal efectiva con la mayoría de los verdaderos positivos en las clases propuestas y un bajo número de falsos positivos. En las clases de 180° y 45° cuentan con un mayor índice de verdaderos positivos de 1004 y 1000 respectivamente. Sin embargo, la matriz de confusión de Inceptionv3 (Figura 40b) llega a sufrir mínimas variaciones en los falsos positivos e incrementa los falsos negativos, los cuales es notoria su incremento en el ángulo de 60° con un índice de 68 que las confunde con el ángulo de 75° en comparación con ResNet50. Los ángulos con un mayor número de predicciones correctas respecto a su etiqueta real son 180° con 1004 y 30° con 990. En la matriz de confusión del modelo de VGG16 (Figura 40a), se evidencia que cuenta con un índice alto de falsos positivos y falsos negativos, mostrando un comportamiento errático en todas las clases. No obstante, los ángulos de 75° , 90° y 180° llegan a tener un valor alto en predicciones incorrectas de 116, 162 y 158 respectivamente. Las clases de 60° y 90° cuentan con un valor relativamente alto de verdaderos positivos de 157 y 152 respectivamente.

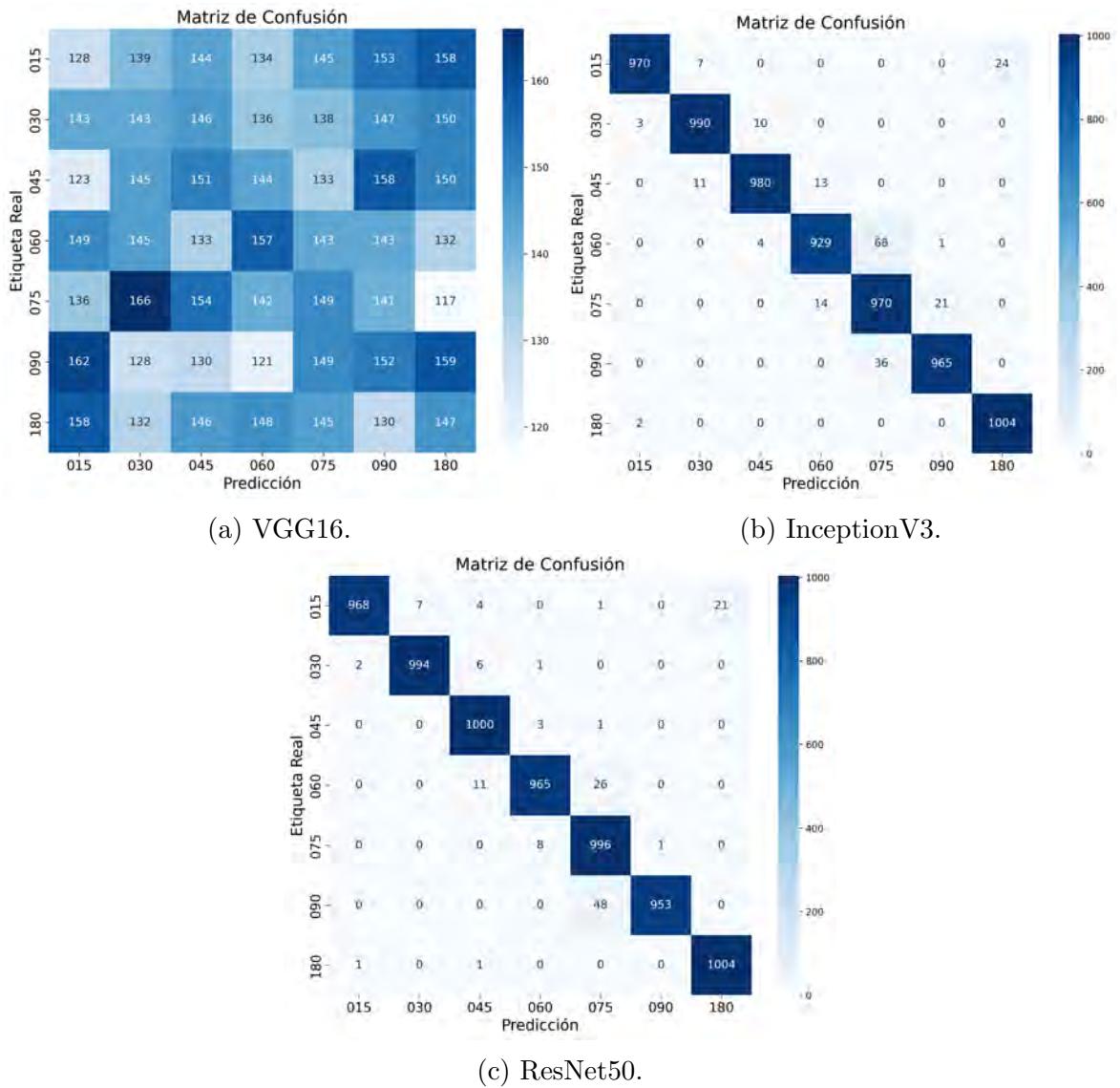


Figura 40: Matrices de confusión de los modelos de clasificación.

Complementando la información de las matrices de confusión, el reporte de clasificación detallado en la Tabla 6, es notorio que el modelo que predice la mayoría de las clases de manera positiva es ResNet50 al tener una precisión mayor a 0.98 en seis clases y un valor de 0.93 en el ángulo de 75°. Ocurre algo similar con InceptionV3, solo que el umbral mayor es de 0.97 y el más bajo es de 0.90 que corresponde al ángulo de 75°, lo cual corresponde a los falsos negativos vistos en su matriz de confusión. Sin embargo, la precisión en las clases del modelo de VGG16 se encuentran en un índice muy bajo comparada a los dos modelos restantes, teniendo valores entre 0.13 a 0.16, siendo este último perteneciente al ángulo de 60°.

Adicional a esto, el Recall o sensibilidad mide las instancias clasificadas como positivas, un valor alto de Recall indica que pocas imágenes GEI pertenecientes a las clases

Modelo	Clase	Precisión	Recall	F1-score
VGG16	15°	0.13	0.13	0.13
	30°	0.14	0.14	0.14
	45°	0.15	0.15	0.15
	60°	0.16	0.16	0.16
	75°	0.15	0.15	0.15
	90°	0.15	0.15	0.15
	180°	0.15	0.15	0.15
ResNet50	15°	1.00	0.97	0.98
	30°	0.99	0.99	0.99
	45°	0.98	1.00	0.99
	60°	0.99	0.96	0.98
	75°	0.93	0.99	0.96
	90°	1.00	0.95	0.97
	180°	0.98	1.00	0.99
InceptionV3	15°	0.99	0.97	0.98
	30°	0.98	0.99	0.98
	45°	0.99	0.99	0.98
	60°	0.97	0.93	0.95
	75°	0.90	0.97	0.93
	90°	0.98	0.96	0.97
	180°	0.98	1.00	0.99

Tabla 6: Reporte de clasificación de los modelos VGG16, ResNet50 e InceptionV3.

especificadas fueron pasadas por alto o clasificadas como una clase diferente. Tanto Inceptionv3 como ResNet50, fueron los que tuvieron los mayores números de Recall, especialmente en las clases de 30°, 45° y 180°. En último lugar, el F1-score, combina las métricas de precisión y Recall para obtener un equilibrio al identificar las clases correctas, donde el equilibrio es mayor en cuatro clases: 15°, 30°, 45° y 180°.

Las gráficas de entrenamiento de la Figura 41 representan el comportamiento de la pérdida y precisión de los modelos empleados durante el entrenamiento y la validación. Las gráficas de entrenamiento muestran que VGG16 tiene uno de los mejores comportamientos tanto para el entrenamiento como en la validación, puesto que, la pérdida va descendiendo desde 0.2 hasta converger a 0 (Figura 41a), indicando que no hay un sobreentrenamiento o sobreajuste en ambos conjuntos. Lo mismo sucede con la precisión (Figura 41b), la cual es suavizado el comportamiento y de forma exponencial hasta converger a 1. Sin embargo, ResNet50 e InceptionV3, cuentan con comportamientos diversos, ya que, el conjunto de validación sufre de saltos, en especial la parte de la pérdida en ResNet50 (Figura 41c). Esto puede deberse a la estructura de la red, debido a que, al compararlo con las matrices de confusión de dichos modelos, se observa que efectivamente realiza una clasificación asertiva en su mayoría. En el caso de InceptionV3, se observa que sufre de picos muy marcados entre la época 10-12 en la pérdida y precisión (Figura 41e y 41f), esto se debe a la configuración establecida para las primeras 10 épocas y las últimas 10, cambiando el optimizador y la tasa de aprendizaje

como se menciona en el capítulo de metodología.

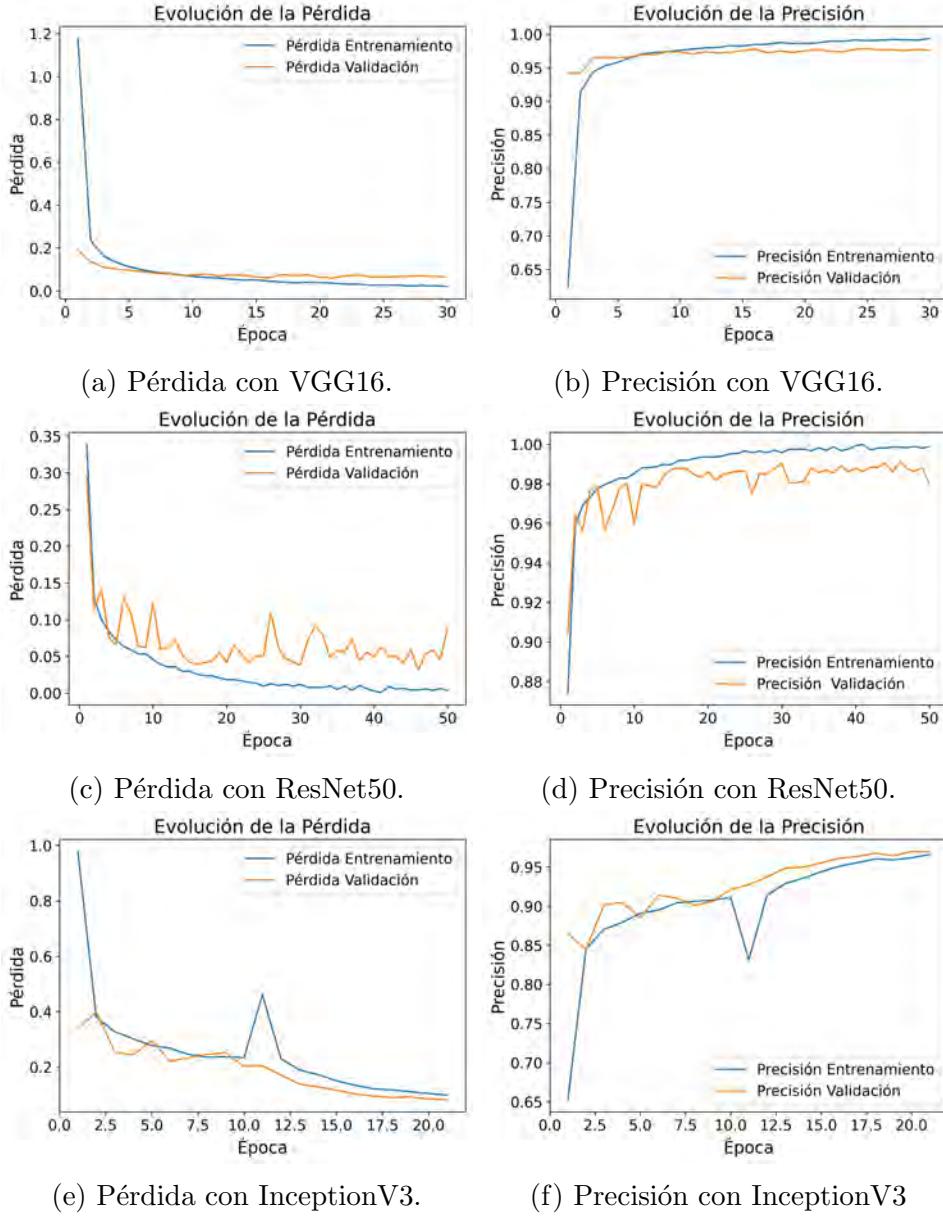


Figura 41: Evolución de la pérdida y precisión de los modelos de VGG16, ResNet50 e Inceptionv3 respectivamente.

6.1.3. Detección de anomalías de datos para la detección de una mala postura

La detección de anomalías de datos consistió en analizar mediante métricas estadísticas las bases de datos generadas. Las cuales cuentan con un total de 4005 individuos.

El mapa de calor mostrado en la Figura 42, se observa la dispersión de los datos entre el promedio del tipo de simetría (SI) que corresponde a las cinco articulaciones específicas: tobillo (ankle), pie (foot), cadera (hip), rodilla (knee) y hombro (shoulder) respecto a los siete ángulos establecidos como clases en los modelos de clasificación. Cabe destacar que el mapa de calor muestra una dispersión de datos en diferentes regiones de los conjuntos de datos analizados y los colores más intensos representan una mayor desviación estándar o dispersión en esa zona. Este comportamiento se puede observar en la simetría de la rodilla con el ángulo de 15° con un valor de 0.0005758, con 0.0004871 en la simetría de la rodilla en un ángulo de 60° , en la simetría del pie con el ángulo de 90° cuenta con un valor de 0.0004404, para el caso de la simetría de la cadera cuenta con una magnitud de 0.0004549 en el ángulo de 180° y finalmente, la simetría del hombro tiene una mayor variabilidad respecto al ángulo de 15° con un 0.0002822, esta alta variabilidad marcada en los tipos de simetría podría indicar una inestabilidad o anomalía.

Por otra parte, los colores suaves que se presentan de color crema en el mapa de calor indican que hay una menor desviación estándar y, por ende, los datos son más constantes en la relación del tipo de simetría y el ángulo. Estos tipos de datos se concentran en las partes centrales de los ángulos como es el caso de 45° , 60° y 75° en la mayoría de los tipos de simetría, que podría indicar que una posible estabilidad o comportamiento repetitivo.

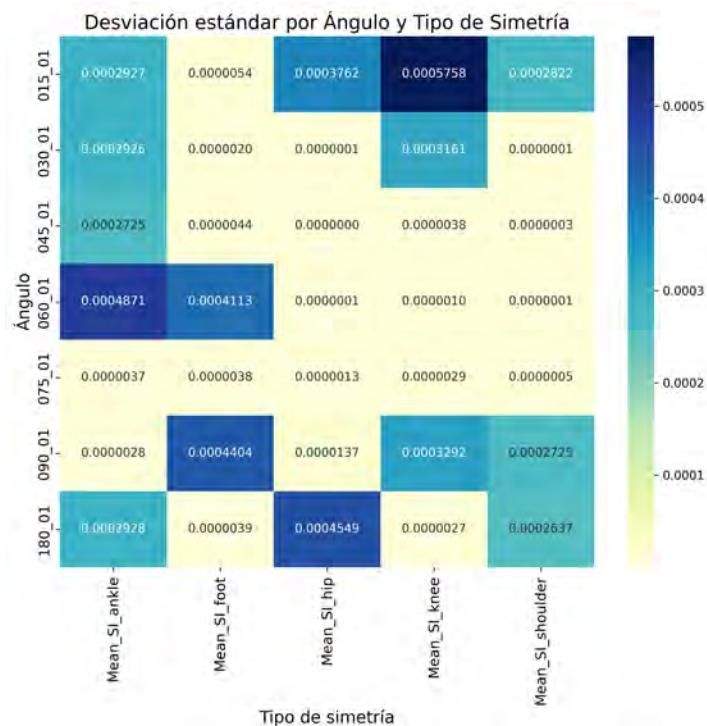


Figura 42: Mapa de calor de desviación estándar.

La matriz de correlación en la Figura 43, muestra los coeficientes entre múltiples

variables. En este caso son los promedios de los cinco tipos de simetría, en el cual se identifican las relaciones lineales entre las variables. En la matriz se observan en color rojo una diagonal con una magnitud igual a 1, que indica una correlación perfecta al estar correlacionada consigo misma el tipo de simetría. Los valores fuera de la diagonal indican una correlación entre pares del tipo de simetría donde se observa que los valores cercanos a cero indican una débil relación dependiendo si cuentan con un $-$, indica que una variable sube mientras la otra baja y viceversa cuando tiene un $+$.

De forma general, es notorio que la correlación entre los tipos de simetría es casi nula al tener valores cercanos a cero y por lo tanto, no comporten una relación lineal.

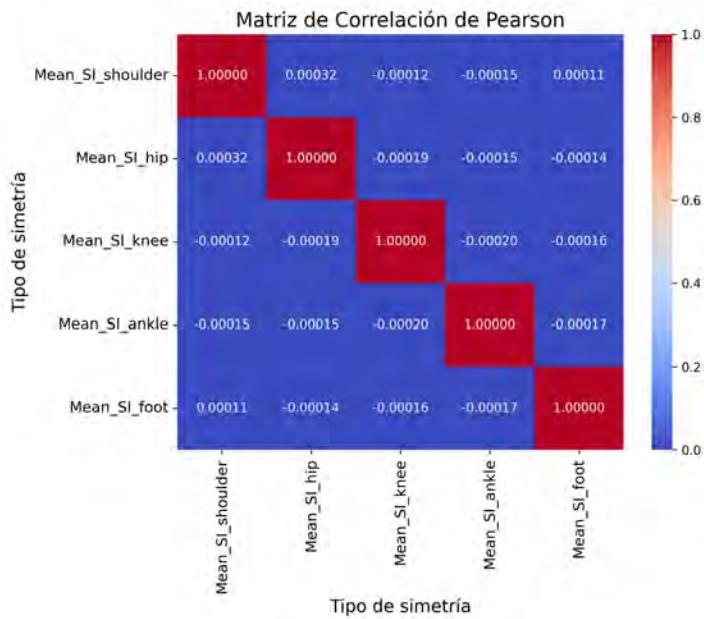


Figura 43: Correlación de Pearson

En la Figura 44, se muestra la distribución de los valores atípicos (outliers) según el ángulo donde los valores globales indican el número total de valores atípicos por ángulo sin importar si los IDs son repetidos. A diferencia de la distribución por IDs únicos, en el que, únicamente se toman en cuenta los IDs que se repiten con valores atípicos. En la Tabla 7, se detallan los índices para cada uno de los ángulos. Esta distribución, refleja a grandes rasgos, un porcentaje aproximado de individuos (IDs) que pueden poseer de alguna anomalía de datos para hacer una detección de una mala postura.

Los valores atípicos fueron visualizados primeramente con Boxplot, un tipo de gráfica que permite ver la distribución y las observaciones atípicas para cada tipo de simetría según el ángulo. Las gráficas generadas con esta función se observan en la Figura A.2 en el apartado de Anexos. Estas gráficas exponen los puntos asilados que representan los valores atípicos, ya que, muestran una distribución completa de los datos que son: mínimo, primer cuartil (Q1), mediana (Q2), tercer cuartil (Q3), máximo y posibles valores atípicos [80].

Ángulo	Valores atípicos totales	Cantidad de IDs únicos con valores atípicos
15°	2558	1996
30°	2146	1721
45°	2127	1656
60°	2182	1658
75°	2471	1838
90°	3475	2395
180°	3435	2451

Tabla 7: Cantidad de valores atípicos (outliers) por ángulo.

La segunda comparación de valores atípicos se observa en la Figura 45, en la cual se compara los porcentajes de valores atípicos por ángulo y por tipo de simetría. Comparación que muestra un índice de que tanto varían los valores atípicos según el punto de vista y en qué articulación predomina más para ser un factor en la detección de la mala postura.

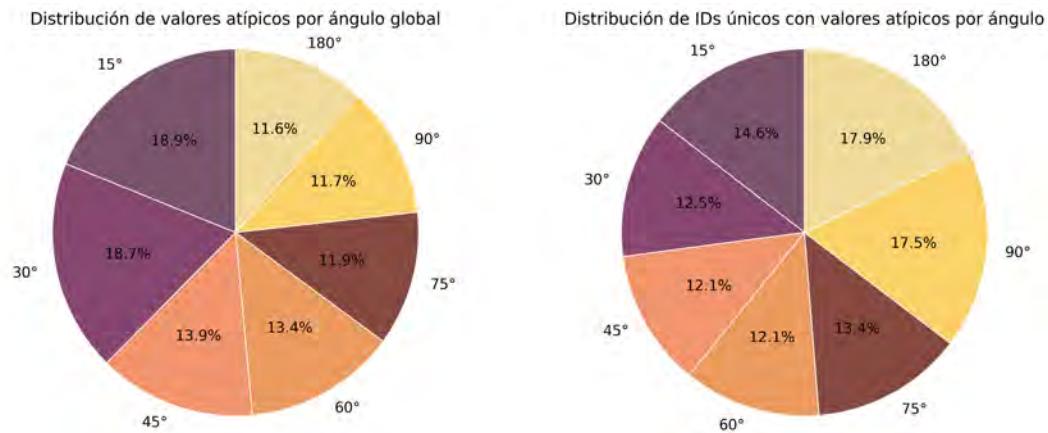


Figura 44: Distribución de los valores atípicos.

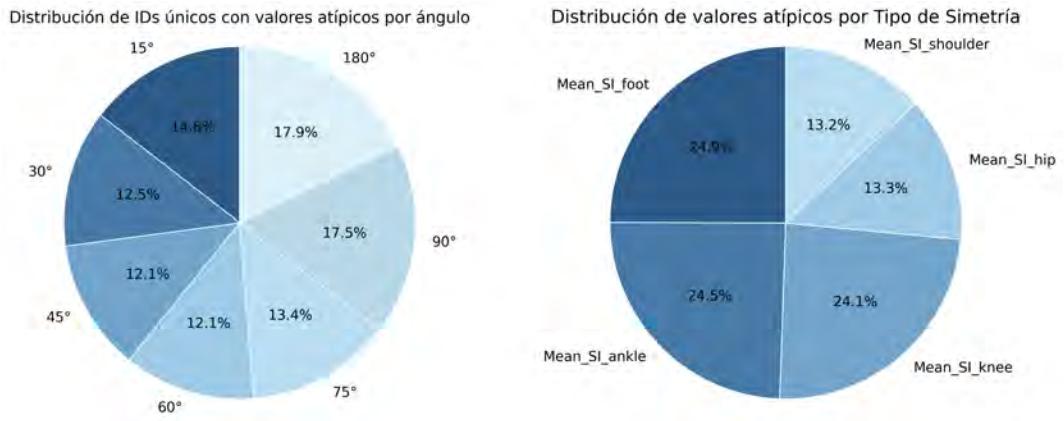


Figura 45: Comparación de valores atípicos respecto al tipo de ángulo y el tipo de simetría.

6.2. Discusión

Los resultados obtenidos en la clasificación de ángulos de imágenes tipo GEI con las arquitecturas de VGG16, ResNet50 e InceptionV3 revelan que estas arquitecturas son funcionales para la clasificación y extracción de características de este tipo de imágenes, principalmente enfocado en las texturas. Sin embargo, las extracciones encontradas con los modelos descritos no permiten extraer la información suficiente para la detección de una mala o buena postura a partir de la silueta de la imagen. Debido a las propiedades de la imagen GEI, que al tratarse de una imagen en escala de grises y con un barrido de secuencias que provocan que la silueta tenga como sombras, provoca una interferencia con la información que se quiere extraer y que no permite emplear plantillas de la estimación de la pose como el conjunto de datos MS COCO o Mediapipe. Plantillas que hacen una calibración en la imagen con un mapa de calor para detectar los puntos de las articulaciones, estas plantillas son efectivas en imágenes a color o en imágenes sintéticas en donde es notorio la posición de las partes del cuerpo que a la vez hacen referencia a algunas de las articulaciones o puntos clave con el que cuente la plantilla. Regiones que no son vistas en una imagen de tipo GEI como se muestran en la Figura A.3a y A.3b del apartado de Anexos.

Por otra parte, ResNet50 e InceptionV3 demostraron ser los modelos más efectivos en cuanto a la extracción de características, según los resultados de PCA, LDA, CCA y CCR. Ambos modelos, logran una buena separación clases a pesar de tener una distribución compleja como se observa en la Figura 39.

Los resultados obtenidos para la detección de anomalías de datos para la detección de una mala postura en este estudio se revelan patrones significativos entre los tipos de simetría y los ángulos. Donde este último se analiza como un ángulo de visión para poder relacionar como afecta la variabilidad con la simetría en un punto específico, es decir, en una de las articulaciones especificadas que son: hombros, cadera, rodilla, tobillo y pie. La articulación del hombro se escogió para verificar que el individuo se encontrara nivelado

o estuviera en equilibrio, esto fue muy notorio en las métricas aplicadas para solo un fotograma que corresponde al del ángulo de 180°, haciendo alusión a un plano sagital y donde ambos pies se encuentran sobre el piso obteniendo para todos los individuos analizados un balance igual a cero, es decir, hay un equilibrio. Además, se obtuvo un promedio de 1, una desviación estándar de 0.000019 y una varianza de 3.480592e-10 para la simetría del hombro, lo que nos indica que el fotograma escogido y los individuos tienen un comportamiento estable y homogéneo, que en biomecánica se conocería como es reposo o en un estado en equilibrio.

En el mapa de calor de la Figura 46, se muestran las métricas de promedio, desviación estándar, varianza y correlación de Pearson medidas entre la distancia de la cabeza a la base de la columna (D H-BS) y la distancia entre las piernas (D legs) entre el dataset de 180° que corresponde a un estado en equilibrio y el dataset global que abarca todos los fotogramas que tiene el individuo. Los patrones de estabilidad en el dataset de 180° es notorio, puesto que, las magnitudes de los valores de sus métricas son bajas comparadas al dataset global. Tal es el caso de la distancia entre las piernas que se observa un aumento en el promedio de 0.1754 a 0.2410, este incremento indica un cambio considerable en la caminata y que la posición de las piernas cambia y que tiene una variabilidad al estar en movimiento, es decir, la zancada de las piernas cambia conforme al tiempo (los fotogramas vistos). En cambio, la correlación entre las dos variables es relativamente más fuerte en el dataset de 180° con 0.3330 a diferencia del global con -0.2835, lo que podría ser un indicio que al abarcar todos los fotogramas y el individuo se encuentra en constante movimiento va perdiendo la relación entre estados variables y que a al vez podría realizar un movimiento al caminar como es agachar la cabeza a comparación de cuando se encuentra en un estado de equilibrio que la relación de las articulaciones se encuentran alineadas.

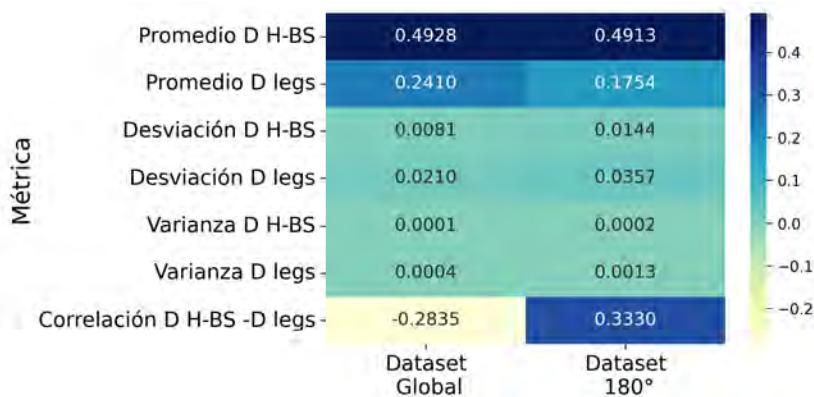


Figura 46: Mapa de calor que compara las métricas de desviación estándar, varianza y correlación de Pearson entre dos variables: D H-BS (la distancia entre la cabeza y la base de la columna) y D legs (la distancia entre las piernas) de dos datasets diferentes. El primer dataset es de datos globales y el segundo es de solo una vista o fotograma de un individuo que es en el ángulo de 180°.

Al comparar los resultados del dataset global con el dataset de 180°, se comprende con mayor profundidad el mapa de calor del tipo de simetría respecto a los ángulos de la Figura 42, en el cual los patrones que indican una mayor estabilidad tanto en ángulo como en el tipo de articulación son la del pie, la cadera y rodilla. Articulaciones donde hay una mayor actividad muscular para la caminata de una persona y que coinciden en el ángulo de 45° y 75°. Esto podría indicar una estabilidad y donde los patrones son mayores para considerar una buena simetría entre las articulaciones, y por ende una postura que se mantiene y se puede clasificar como buena. Sin embargo, valores de desviación estándar, indican una posible anomalía o asimetría entre la articulación como es el caso de la rodilla, el tobillo y la cadera en el ángulo de 15°. No obstante, la articulación del tobillo sufre de mayores índices de anomalías en los ángulos de 15°, 30°, 45°, 60° y 180°, lo que podría implicar que al ser una articulación donde se deposita gran parte de la fuerza de apoyo al momento de caminar, esta sufre de variaciones en sus ángulos al realizar el movimiento y por ello se ve afectado en los ángulos.

Adicional a esto, la simetría en el pie tiene un mayor porcentaje de 24.9 % de los valores atípicos junto con el tobillo y la rodilla con un 24.5 % y 24.1 % respectivamente como se ilustra en la Figura 45. Como se mencionó con anterioridad, esto se debe principalmente a que conforma las articulaciones de los miembros inferiores y que cargan con el mayor número de porcentaje durante la marcha y sufren de extensiones o flexiones, provocando que los cambios sean radicales y al estudiarlos por fotogramas indique posibles anomalías al obtener asimetrías en esas articulaciones. Esta información se complementa con la distribución de los ángulos, en el que se observa que hay un mayor índice de encontrar las anomalías en el ángulo de 180° y 90° al contar con un porcentaje de 17.9 % y 17.5 % respectivamente.

A partir de los resultados de las distribuciones de los valores atípicos de los ángulos y los tipos de simetría, se obtuvo el promedio global de simetría para clasificar de forma global los datos definiendo un umbral por debajo del 25 % de los datos correspondientes a la simetría de la rodilla. De acuerdo con los resultados obtenidos, la articulación de la rodilla es un centro fundamental en el miembro inferior y la cual tienen una estabilidad en cuatro ángulos diferentes (45°, 60°, 75° y 180°). Siendo el más apto para definir los casos anómalos si hay una simetría baja. Una vez definido el umbral, se entrenó un clasificador básico con RandomForest en el que sólo se establecieron dos clases: buena y anomalía. La primera clase es el etiquetado de una buena simetría y por ende una buena postura a diferencia de la segunda clase que denota una asimetría, es decir, una posible anomalía al caminar y pude ser un indicio de una mala postura de una persona. La matriz de confusión se observa en la Figura 47, donde se evidencia que de 28025 datos analizados que pertenecen a 4003 individuos, de los datos de prueba, predice que 6097 datos son verdaderos positivos, es decir, 870 individuos tienen una buena postura. Para la clase de anomalía, predice 345 datos como verdaderos positivos que son solamente 49 individuos con una mala postura. En cambio, los falsos negativos son mayores en la clase de anomalía con 1965 datos que corresponden a 280 individuos son predichos con una buena postura, obteniendo un total de 1199 individuos para la parte de validación que corresponde a un 30 % del conjunto de testeо y 70 % para el conjunto de entrenamiento.

En la gráfica de la Figura 48, se muestra la distribución en porcentaje, destacando una buena postura con un 73.2% y un 26.8% con una anomalía.

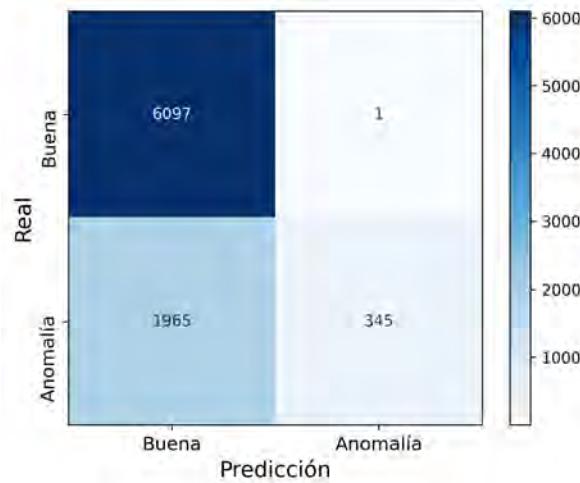


Figura 47: Matriz de confusión de la clasificación global de simetría para las clases de Buena y Anomalía.



Figura 48: Distribución global de la clasificación Buena vs Anomalía

7. Conclusiones

Al implementar modelos de redes neuronales convolucionales para la extracción de características de imágenes tipo GEI, se observó que las propiedades de la imagen delimitan los métodos de extracción de características para un estudio de la mala postura basándose solamente en la silueta de la forma humana. Por lo que, complementa la información al emplear el uso de esqueletos para la representación de la silueta a partir de imágenes GEI. Con el cual, se puede estudiar mediante los datos de las coordenadas al aplicar métricas específicas del esqueleto como es la altura de la persona, distancias entre articulaciones, simetría entre una articulación, simetría de la marcha, entre otras. Métricas que permiten complementar la información y obtener valores atípicos para poder aplicar una clasificación para la detección de una mala postura basándose en las anomalías de los datos generadas.

La metodología presentada para la extracción de características de los esqueletos permite aplicar métricas y evaluar la relación que hay entre los tipos de simetría y los ángulos según las articulaciones especificadas. El modelo que obtuvo un mayor índice en cuatro de seis variables medidas fue ResNet50, que al comparar con las gráficas anteriores. El comportamiento de este modelo refleja un rendimiento consistente según el tipo de imagen analizada. Además, el estudio de ablación propuesto con los cambios en los hiperparámetros demuestra que, con el espacio de búsqueda, ResNet50 sigue teniendo un resultado notable más bien no garantiza que sea el mejor en comparación con los dos modelos restantes: VGG16 e InceptionV3.

La detección de anomalías de datos para la detección de una mala postura mediante análisis estadísticos, permitió relacionar las articulaciones involucradas al caminar que son: la cadera, rodilla, tobillo y pie.

7.1. Limitaciones

A pesar de los resultados satisfactorios en función de la extracción de características y detección de anomalías de datos para la detección de una mala postura, este estudio tiene varias limitaciones. Principalmente el conjunto de la base de datos con la que se trabajó en la parte de las imágenes y las coordenadas para la estimación del esqueleto en 2D, OU-ISIR MVLP. Base de datos que es eficaz, sin embargo, para la orientación del proyecto presenta retos al trabajar con estas dos fuentes de entrada. La información no fue suficiente para aplicar una detección en la imagen tipo GEI al extraer características con métodos manuales y con métodos aplicando inteligencia artificial que fueron modelos de redes neuronales convolucionales. Además, los modelos propuestos para la extracción de características, aunque fueron satisfactorios, no son suficientes para determinar una mala postura basándose en una imagen al recrear la información de toda la silueta mediante los mapas de características y requieren una mayor validación de la postura de la persona. Adicionalmente, la detección de datos anomalías, requieren una validación en entornos clínicos reales, así como las métricas obtenidas para la medición de la simetría de la marcha. Estas últimas pueden no traducirse directamente a una

practica clínica debido a factores externos en la biomecánica de la marcha que influyen en el comportamiento de los datos y su interpretación en un entorno más especializado como es la rehabilitación y la biomecánica del cuerpo humano. En futuros trabajos se deberían implementar ambos tipos de entrada para el estudio biomecánico de la postura mediante la marcha, implementando modelos para la extracción de características y estimación de la pose.

7.2. Trabajo a futuro

Para trabajos posteriores, se plantean mejoras en la detección de la mala postura al realizar la detección mediante captura de imágenes propias y su respectiva estimación de la pose con ayuda de estimadores más profundos principalmente, con el objetivo de crear una base de datos más sólida y con información enfocada en biomecánica para un trabajo más completo en la detección de una buena o mala postura. Se sugiere que investigaciones futuras se conozca los datos con el que se pueda relacionar a un porcentaje de la masa corporal de un individuo para tener modelamiento del cuerpo humano y a partir de esta etapa desarrollar un modelo biomecánico más completo con ayuda de la antropometría (un análisis más profundo sobre sus centros de masa, masa de los segmentos de las extremidades, las longitudes de los segmentos, los centros de rotación, los ángulos de tracción de los músculos, momentos de inercia, entre otros). Adicional a esto, es posible parametrizar las métricas de una clasificación con anomalía o no mediante métricas del esqueleto más detalladas a partir de los estimadores y el modelo biomecánico, mejorando las mediciones y compresión del estudio de la mala postura y el ciclo de marcha con conceptos biomecánicos. Sumando a ello, la implementación de modelos de redes neuronales convolucionales profundos específicos para la detección de la mala postura, ampliando el alcance del proyecto para futuras investigaciones sobre la movilidad del cuerpo humano tanto en la parte de imagen como del esqueleto y exportarlo a aplicaciones móviles o estudios del deporte o rehabilitación.

Referencias bibliográficas

- [1] C. Serrano, “Sistema musculoesquelético,” <https://www.kenhub.com/es/library/anatomia-es/sistema-musculoesqueletico>, Octubre 2023, accedido en noviembre de 2023.
- [2] P. M. McGinnis, *Biomechanics of Sport and Exercise*, 3rd ed. Champaign, IL: Human Kinetics, 2013.
- [3] P. T. Henning, *Anatomy and Kinesiology of the Lower Limb*. Springer Publishing Company, 2013, pp. 1–8.
- [4] X. Hongye and H. Zhuoya, “Gait recognition based on gait energy image and linear discriminant analysis,” in *2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2015, pp. 1–4.
- [5] C. Aracena, F. Villena, F. Arias, and J. Dunstan, “Aplicaciones de aprendizaje automático en salud,” *Revista Médica Clínica Las Condes*, vol. 33, no. 6, pp. 568–575, 2022.
- [6] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” 2015.
- [7] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, “Convolutional neural networks with transfer learning for pneumonia detection,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, 2022.
- [8] A. C. Alonso and E. M. Jara, “Visión por computadora: identificación, clasificación y seguimiento de objetos.” *FPUNE Scientific*, no. 10, 2016.
- [9] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, “Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition,” *IPSJ Transactions on Computer Vision and Applications*, vol. 10, no. 1, 2018.
- [10] X. Li, Y. Makihara, C. Xu, and Y. Yagi, “Multi-view large population gait database with human meshes and its performance evaluation,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 2, pp. 234–248, 2022.
- [11] R. G. Birdal, A. Sertbaş, and B. Mihendisligi, “Human identification based on gait analysis: A survey,” in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 2018, pp. 489–493.
- [12] K. Bashir, T. Xiang, and S. Gong, “Gait recognition without subject cooperation,” *PATTERN RECOGNITION LETTERS*, vol. 31, no. 13, pp. 2052–2060, 2010.

- [13] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng, “GaitSet: Cross-view gait recognition through utilizing gait as a deep set,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3467–3478, 2022.
- [14] C. Wan, L. Wang, and V. V. Phoha, “A survey on gait recognition,” *ACM COMPUTING SURVEYS*, vol. 51, no. 5, 2019.
- [15] P. Connor and A. Ross, “Biometric recognition by gait: A survey of modalities and features,” *Computer Vision and Image Understanding*, vol. 167, pp. 1–27, 2018.
- [16] Y. Zhang, C. Tao, H. Wang, and Y. Fan, “Biomechanical effects of human-mobility aid interaction: A narrative review,” *Gait Posture*, vol. 118, pp. 1–12, 2025.
- [17] E. Alijanpour and D. M. Russell, “Gait phase normalization resolves the problem of different phases being compared in gait cycle normalization,” *Journal of Biomechanics*, vol. 173, p. 112253, 2024.
- [18] C. Vora, V. Katkar, and M. Lunagaria, “GAIT analysis based on GENDER detection using pre-trained models and tune parameters,” *Discover Artificial Intelligence*, vol. 4, no. 1, p. 19, 2024.
- [19] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 4th ed. Hoboken, New Jersey: John Wiley Sons, 2009.
- [20] G. Li and M. Hayashibe, “Motor synergy and energy efficiency emerge in whole-body locomotion learning,” *Scientific Reports*, vol. 15, no. 1, p. 712, 2025.
- [21] N. A. A. Moreno and O. T. Aranza, “Frecuencia de maloclusiones y su asociación con problemas de postura corporal en una población escolar del estado de México,” *Boletín médico del Hospital Infantil de México*, vol. 70, no. 5, pp. 364–371, 2013.
- [22] R. R. Rodríguez, M. de la Caridad Socorro Santana, M. J. Padrón, G. G. Peñate, and L. M. S. Jordán, “Dolor de espalda y malas posturas, ¿un problema para la salud?” *Revista Médica Electrónica*, vol. 40, no. 3, pp. 833–838, 2018.
- [23] K. S. Saladin, *Anatomía y fisiología: la unidad entre forma y función*, 6th ed. México, D. F.: McGRAW-HILL INTERAMERICANA EDITORES, 2013.
- [24] P. Balthazard, D. Curat, and F. Degache, “Fundamentos de biomecánica,” *EMC - Kinesiterapia - Medicina Física*, vol. 36, no. 4, pp. 1–8, 2015.
- [25] A. M. Nogueras, J. O. R. J L. Calvo Arenillas, F. J. B. Iglesias, and C. S. Sánchez, “Fases de la marcha humana,” *Revista Iberoamericana de Fisioterapia y Kinesiología*, vol. 2, no. 1, pp. 44–49, 1999.
- [26] R. D. Gregorio and L. Vocenas, “Identification of gait-cycle phases for prosthesis control,” *Biomimetics*, vol. 6, no. 2, 2021.

- [27] P. Grass, J.-M. Casillas, V. Dulieu, and J.-P. Didier, “La marcha,” *Kinesiterapia - Medicina física*, pp. 44–49, 1999.
- [28] J. V. Campellone, “Anomalías en la forma de caminar,” <https://medlineplus.gov/spanish/ency/article/003199.htm>, enero 2023, accedido en noviembre de 2023.
- [29] D. J. Magee and R. C. Manske, “Chapter 14: Assessment of gait,” in *Orthopedic physical assessment*, 7th ed. Elsevier, 2021.
- [30] J. E. Mason, I. Traoré, and I. Woungang, *Introduction to Gait Biometrics*. Cham: Springer International Publishing, 2016, pp. 1–7.
- [31] Z. O. Abu-Faraj, G. F. Harris, P. A. Smith, and S. Hassani, *Human gait and Clinical Movement Analysis*. John Wiley Sons, Ltd, 2015, pp. 1–34.
- [32] A. Tharatipyakul, T. Srikaewsw, and S. Pongnumkul, “Deep learning-based human body pose estimation in providing feedback for physical movement: A review,” *Heliyon*, vol. 10, no. 17, 2024.
- [33] R. Yue, Z. Tian, and S. Du, “Action recognition based on rgb and skeleton data sets: A survey,” *Neurocomputing*, vol. 512, pp. 287–306, 2022.
- [34] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1302–1310.
- [35] M. Mundt, Z. Born, M. Goldacre, and J. Alderson, “Estimating ground reaction forces from two-dimensional pose data: A biomechanics-based comparison of alphapose, blazepose, and openpose,” *Sensors*, vol. 23, no. 1, 2023.
- [36] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, “Deep high-resolution representation learning for visual recognition,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, 2021.
- [37] A. Toshev and C. Szegedy, “Deeppose: Human pose estimation via deep neural networks,” in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2014, p. 1653–1660.
- [38] X. Li, Y. Makihara, C. Xu, and Y. Yagi, “End-to-end model-based gait recognition using synchronized multi-view pose constraint,” in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 4089–4098.
- [39] Z. Liu and S. Sarkar, “Improved gait recognition by gait dynamics normalization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 863–876, 2006.

- [40] A. O. Lishani, L. Boubchir, and A. Bouridane, “Haralick features for gei-based human gait recognition,” in *2014 26th International Conference on Microelectronics (ICM)*, 2014, pp. 36–39.
- [41] N. Múnera, C. Alvarez, S. Sastoque, and M. Iregui, “Human features extraction by using anatomical and low level image descriptors from whole body images,” in *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, 2016, pp. 1–6.
- [42] X. Huang, D. Zhu, H. Wang, X. Wang, B. Yang, B. He, W. Liu, and B. Feng, “Context-sensitive temporal feature learning for gait recognition,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 12 889–12 898.
- [43] S. A. F. Manssor, S. Sun, and M. A. M. Elhassan, “Real-time human recognition at night via integrated face and gait recognition technologies,” *Sensors*, vol. 21, no. 13, 2021.
- [44] A. Dutta, R. Lal, D. S. Raychaudhuri, C.-K. Ta, and A. K. Roy-Chowdhury, “Poise: Pose guided human silhouette extraction under occlusions,” in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 6141–6151.
- [45] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, “A study on gait-based gender classification,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905–1910, 2009.
- [46] A. Sepas-Moghaddam and A. Etemad, “Deep Gait Recognition: A Survey ,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 45, no. 01, pp. 264–284, 2023.
- [47] I. R. T. Alvarez and G. Sahonero-Alvarez, “Cross-view gait recognition based on u-net,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–7.
- [48] Z. Huang, D. Xue, X. Shen, X. Tian, H. Li, J. Huang, and X.-S. Hua, “3d local convolutional neural networks for gait recognition,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 14 900–14 909.
- [49] J. P. Singh, S. Jain, S. Arora, and U. P. Singh, “Vision-based gait recognition: A survey,” *IEEE Access*, vol. 6, pp. 70 497–70 527, 2018.
- [50] G. Ascenso, M. H. Yap, T. Allen, S. S. Choppin, and C. P. and, “A review of silhouette extraction algorithms for use within visual hull pipelines,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 6, pp. 649–670, 2020.

- [51] Y. Bo, “Gait Silhouette Extraction Algorithm Using Gauss Model,” *Chinese Journal of Sensors and Actuators*, 2008.
- [52] J. Chen and Y. Zhang, “Silhouette extraction from human gait images sequence using cosegmentation,” in *Optoelectronic Imaging and Multimedia Technology II*, vol. 8558, 2012, p. 85580U.
- [53] H. Ng, W.-H. Tan, H.-L. Tong, J. Abdullah, and R. Komiya, “Extraction of human gait features from enhanced human silhouette images,” in *2009 IEEE International Conference on Signal and Image Processing Applications*, 2009, pp. 425–430.
- [54] J. H. Hollman, E. M. McDade, and R. C. Petersen, “Normative spatiotemporal gait parameters in older adults,” *Gait and Posture*, vol. 34, no. 1, pp. 111–118, 2011.
- [55] W. Isa, “Analysis on spatial and temporal features of gait kinematics,” in *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID’05)*, 2005, pp. 130–133.
- [56] Y. Chen and X. Li, “Gait feature learning via spatio-temporal two-branch networks,” *Pattern Recognition*, vol. 147, p. 110090, 2024.
- [57] J. Wan, H. Zhao, R. Li, R. Chen, and T. Wei, “Omni-domain feature extraction method for gait recognition,” *Mathematics*, vol. 11, no. 12, 2023.
- [58] R. B. Ruiz and J. D. Velásquez, “Inteligencia artificial al servicio de la salud del futuro,” *Revista Médica Clínica Las Condes*, vol. 34, no. 1, pp. 84–91, 2023.
- [59] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.
- [60] M. M. Taye, “Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions,” *Computation*, vol. 11, no. 3, 2023.
- [61] J. Qiu, J. Liu, and Y. Shen, “Computer vision technology based on deep learning,” in *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 2, 2021, pp. 1126–1130.
- [62] N. Zhou, “Image recognition in depth: comparative study of CNN and Pre-trained VGG16 architecture for classification tasks,” in *Second International Conference on Physics, Photonics, and Optical Engineering (ICPPOE 2023)*, Y. Liu, Ed., vol. 13075, International Society for Optics and Photonics. SPIE, 2024, pp. 553–559.

- [63] P. Gayathri, A. Dhavileswarapu, S. Ibrahim, R. Paul, and R. Gupta, “Exploring the potential of vgg-16 architecture for accurate brain tumor detection using deep learning,” *Journal of Computers, Mechanical and Management*, vol. 2, no. 2, p. 13–22, 2023.
- [64] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [65] A. H. M. Linkon, M. M. Labib, T. Hasan, M. Hossain, and Marium-E-Jannat, “Deep learning in prostate cancer diagnosis and gleason grading in histopathology images: An extensive study,” *Informatics in Medicine Unlocked*, vol. 24, p. 100582, 2021.
- [66] T. Soni and S. Gupta, “Inceptionv3-based human action detection: An extensive analysis of deep learning methods for enhanced classification performance,” in *2024 Global Conference on Communications and Information Technologies (GCCIT)*, 2024, pp. 1–5.
- [67] P. Shourie, V. Anand, R. Chauhan, G. Verma, and S. Gupta, “A deep dive into gender classification using inception v3: Performance and insights,” in *2023 Global Conference on Information Technologies and Communications (GCITC)*, 2023, pp. 1–6.
- [68] ——, “A deep dive into gender classification using inception v3: Performance and insights,” in *2023 Global Conference on Information Technologies and Communications (GCITC)*, 2023, pp. 1–6.
- [69] Z. Li, “Practice of gesture recognition based on resnet50,” *Journal of Physics: Conference Series*, vol. 1574, no. 1, p. 012154, jun 2020.
- [70] F. Rohani, K. Khoshhal Roudposhti, H. Taheri, A. Mashhadi, and A. Mueller, “Extracting gait and balance pattern features from skeleton data to diagnose attention deficit/hyperactivity disorder in children,” *The Journal of Supercomputing*, vol. 80, no. 6, pp. 8330–8356, 2024.
- [71] K. K. Sahu and K. Satao, “Image compression methods using dimension reduction and classification through pca and lda: A review,” *International Journal of Science and Research*, vol. 5, no. 5, pp. 2277–2280, 2016.
- [72] C. Li and B. Wang, “Fisher linear discriminant analysis,” *CCIS Northeastern University*, vol. 6, 2014.
- [73] I. González, S. Déjean, P. G. P. Martin, and A. Baccini, “Cca: An r package to extend canonical correlation analysis,” *Journal of Statistical Software*, vol. 23, no. 12, pp. 1–14, 2008.

- [74] Z. Sedaghatjoo, H. Hosseinzadeh, and B. S. Bigham, “Local binary pattern(lbp) optimization for feature extraction,” 2024.
- [75] S. Viteckova, P. Kutilek, Z. Svoboda, R. Krupicka, J. Kauler, and Z. Szabo, “Gait symmetry measures: A review of current and prospective methods,” *Biomedical Signal Processing and Control*, vol. 42, pp. 89–100, 2018.
- [76] B. E. Rapp, “Chapter 7 - vector calculus,” in *Microfluidics: Modelling, Mechanics and Mathematics*, ser. Micro and Nano Technologies, B. E. Rapp, Ed. Oxford: Elsevier, 2017, pp. 137–188.
- [77] M. Khosravy, N. Nitta, K. Nakamura, and N. Babaguchi, “Chapter 1 - compressive sensing theoretical foundations in a nutshell,” in *Compressive Sensing in Healthcare*, ser. Advances in ubiquitous sensing applications for healthcare, M. Khosravy, N. Dey, and C. A. Duque, Eds. Academic Press, 2020, pp. 1–24.
- [78] W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introducción a la probabilidad y estadística*, 13th ed. México, D.F: Cengage Learning Editores, 2010.
- [79] J. J. Berman, “Chapter 4 - understanding your data,” in *Data Simplification*, J. J. Berman, Ed. Boston: Morgan Kaufmann, 2016, pp. 135–187.
- [80] J. Hunter, D. Dale, E. Firing, M. Droettboom, and the Matplotlib development team, “matplotlib.pyplot.boxplot,” https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.boxplot.html, 2012.
- [81] B. Su, C. Smith, and E. G. Farewik, “Gait phase recognition using deep convolutional neural network with inertial measurement units,” *Biosensors*, vol. 10, no. 9, pp. 1–18, 2020.
- [82] A. Shafee, D. Kim, and D. Kim, “Simple 1d cnn model for accurate classification of gait patterns using grf data.” *Journal of Information Communication Convergence Engineering*, vol. 22, no. 4, pp. 303–309, 2024.
- [83] B. Su, C. Smith, and E. Gutierrez Farewik, “Gait phase recognition using deep convolutional neural network with inertial measurement units,” *Biosensors*, vol. 10, no. 9, 2020.
- [84] A. Dhillon and G. K. Verma, “Convolutional neural network: a review of models, methodologies and applications to object detection,” *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

A. Anexos

A.1. Figuras complementarias

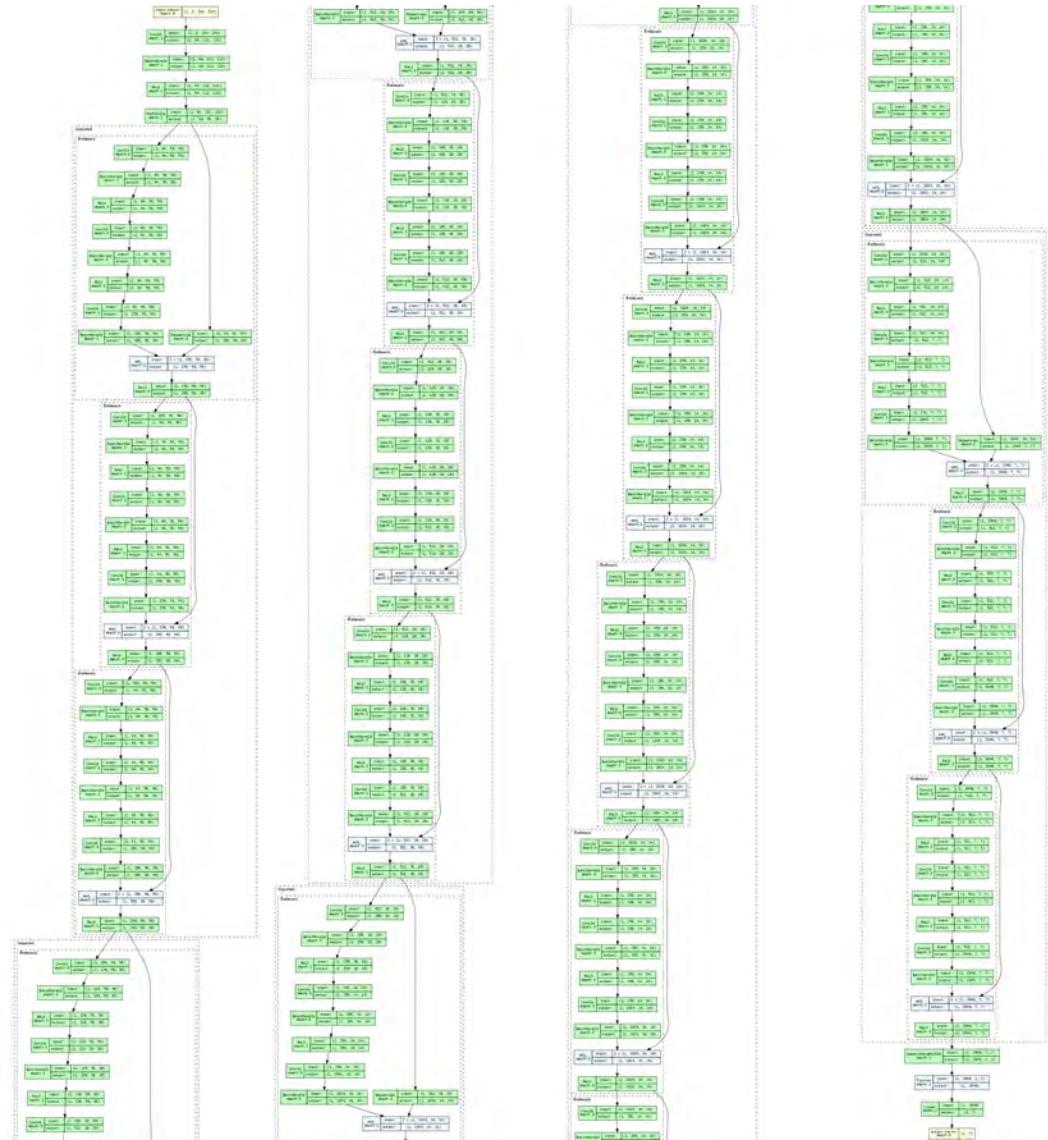
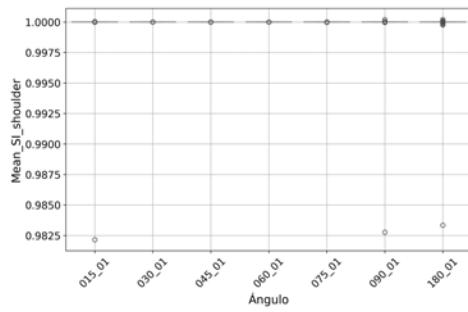
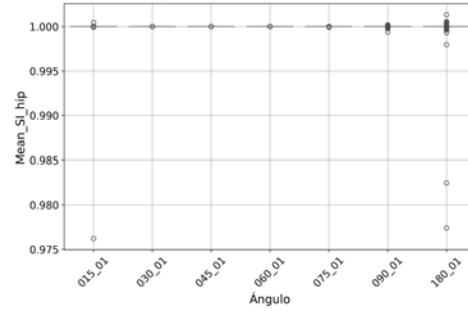


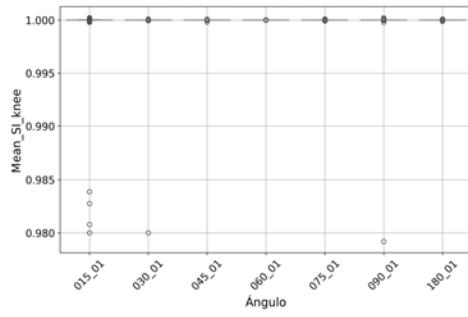
Figura A.1: Arquitectura ResNet50



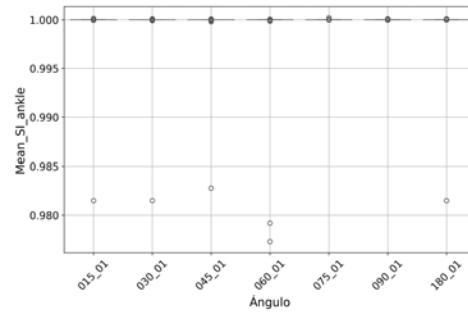
(a) SI del hombro.



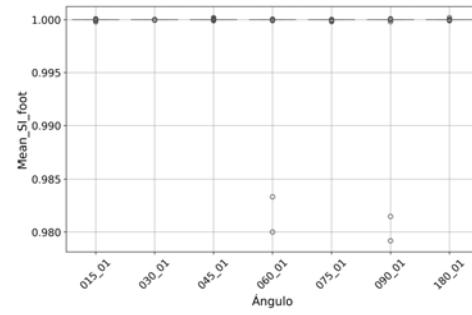
(b) SI de la cadera.



(c) SI de la rodilla.

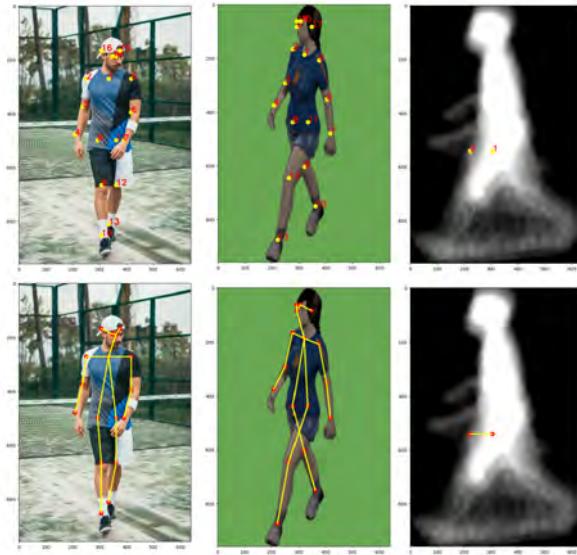


(d) SI del tobillo.

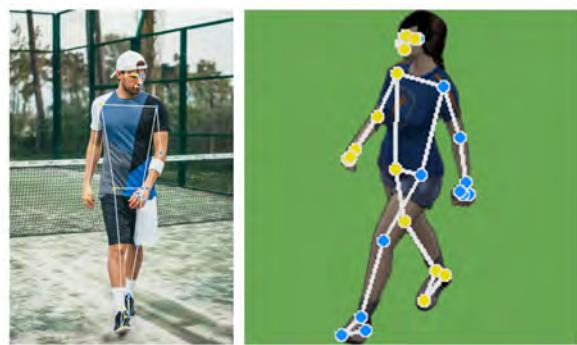


(e) SI del pie.

Figura A.2: Distribución de los tipos de simetría.



(a) Estimación de la pose con COCO.



(b) Estimación de la pose con Mediapipe.

Figura A.3: Estimación de la pose con estimadores de COCO y Mediapipe. Imágenes de creación propia.

A.2. Documentos



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

PEREZ CARRASCO PERLA

Presentó el **Examen de Manejo de la Lengua** efectuado el día dieciocho de septiembre de dos mil veinticuatro, en el cual obtuvo la siguiente calificación:

8-

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día veinticuatro de septiembre de dos mil veinticuatro.

Atentamente,
"Enlazar Culturas por la Palabra"




DRA. MA. DE LOURDES RICO CRUZ

MLRC/mgoa*CL*FLL-C.-44

UAQ CRECER EN LA DIVERSIDAD

fil.uq.mx
442 192 12 00 EXT. 61010

Campus Aeropuerto, Anillo Vial Fray Junípero Serra s/n,
Santiago de Querétaro, Qro. México. C.P. 76140.

FOLIO: 65313

Figura A.4: Constancia de manejo de la lengua inglés.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

PEREZ CARRASCO PERLA

Presentó y acreditó el **Examen de Comprensión de Textos en Inglés** efectuado el día cinco de junio de dos mil veinticuatro.

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día doce de septiembre de dos mil veinticuatro.

A tentamente,
"Enlazar Culturas por la Palabra"




DRA. MA. DE LOURDES RICO CRUZ

MLRC/idos*CL*FLL-C.-38434

UAQ | CRECER EN LA DIVERSIDAD

fil.uq.mx
442 192 12 00 EXT. 61010

Campus Aeropuerto, Anillo Vial Fray Junípero Serra s/n,
Santiago de Querétaro, Qro. México. C.P. 76140.

FOLIO: 64892

Figura A.5: Constancia de comprensión de textos en inglés.

Revisión de enfoques de técnicas empleadas para la extracción de características y reconocimiento de la marcha humana

Review of approaches to human gait feature extraction and recognition techniques

Perla
Perez Carrasco*
*Universidad Autónoma de
Querétaro
Querétaro, México
perla.perezco@gmail.com*

M.C. Osmar Antonio
Espinosa Bernal
*Universidad Autónoma de
Querétaro
Querétaro, México
antonio.be565@gmail.com*

Dr. Jesús Carlos
Pedraza Ortega
*Universidad Autónoma de
Querétaro
Querétaro, México
caryoko@yahoo.com*

Dr. Saúl
Tovar Arriaga
*Universidad Autónoma de
Querétaro
Querétaro, México
saul.tovar@uag.mx*

Resumen— El reconocimiento de la marcha es una técnica biométrica que tiene como objetivo identificar a las personas en función de su forma de caminar, dado que son características únicas que definen a cada uno de los individuos. El aprendizaje profundo ha transformado los estudios y análisis de la marcha al emplear inteligencia artificial para la detección de esta. Los métodos de reconocimiento de la marcha basados en el aprendizaje profundo dominan el estado del arte en el campo y han fomentado aplicaciones en el mundo real. Este artículo revisa

I. INTRODUCCIÓN
La locomoción humana es un rasgo físico y conductual único para cada individuo, el cual puede considerarse un patrón biométrico debido a su singularidad debido a que la marcha manifiesta el estilo de caminar de cada persona y este reconocimiento significa autenticar a una persona [1]-[3], [7]. El interés en el estudio de la marcha tiene raíces en la antigua

Figura A.6: Artículo presentado en el Decimosexto Congreso Internacional de Cómputo en Optimización y software (CICOS 2024).



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



REVISTA PROGRAMACIÓN MATEMÁTICA Y SOFTWARE

Cuernavaca, Morelos, 29 de enero de 2025.

**Perla Pérez Carrasco, Osmar Antonio Espinosa Bernal,
Jesús Carlos Pedraza Ortega, Saúl Tovar Arriaga**

PRESENTES

Les informo que el artículo de investigación titulado “Revisión de enfoques de técnicas empleadas para la extracción de características y reconocimiento de la marcha humana”, con número de identificación: PROGMAT-2024-438, fue aceptado para publicarse en la revista electrónica **Programación Matemática y Software, editada por la Universidad Autónoma del Estado de Morelos** y formará parte de las bases de datos que mantienen indizada a la revista. Este artículo fue sometido a una exhaustiva revisión por parte de un comité de expertos en el área, con la conclusión de que el manuscrito cumple con la calidad necesaria para su publicación en la revista. El artículo PROGMAT-2024-438 será publicado en el volumen 18, número 1 en el mes de febrero del 2026.

Sin más que agregar y en espera de seguir contando con aportaciones de la misma calidad para su publicación en subsecuentes números de la revista, me despido con mis más grandes cortesías.


Atentamente
Dr. Marco Antonio Cruz Chávez
Director Editorial PROGMAT
ISSN: 2007-3283
Dirección de Reservas de Derecho
04-2014-070114141100-203

C.i.p. – Archivo



Av. Universidad 1001 Col. Chamilpa, Cuernavaca Morelos, México, 62209, Edificio 48
Tel. (777) 3297084, (777) 3297000, Ext. 6215/ mcruz@uaem.mx <https://progmat.uaem.mx/progmat/>

UAEM
RECTORÍA
2023-2029

Figura A.7: Carta de aceptación para la publicación del artículo presentado en CICOS 2024.



Ciudad de México, a 06 de enero de 2025

**Mtra. Perla Perez Carrasco
PRESENTE**

CARTA DE CONSTANCIA

El Programa Adopte un Talento (PAUTA) del Instituto de Ciencias Nucleares de la UNAM desarrolla estrategias que permiten a las y los investigadores vincularse activamente con la sociedad, especialmente con infancias y juventudes que realizan proyectos de investigación con impacto social.

En este contexto, se hace constar su participación al impartir la ponencia científica titulada “*¿Nuestros músculos tienen electricidad?*”, dirigida a niñas y niños de las escuelas de USEBEQ, la cual se llevó a cabo el pasado **6 de diciembre de 2024**.

Por ello, se extiende la presente constancia como evidencia de su valiosa colaboración en actividades de divulgación científica orientadas a fomentar vocaciones científicas. Su participación incluyó **6 horas de trabajo voluntario**.

De acuerdo con las evaluaciones de impacto realizadas por PAUTA, sabemos que su contribución ha sido fundamental para inspirar a niñas, niños y jóvenes a explorar y proyectarse hacia un futuro en la ciencia.

Le expresamos nuestro más profundo agradecimiento y esperamos seguir contando con su valiosa participación en próximas actividades.

**Dr. Jorge G. Hirsch Ganievich
Presidente del Consejo Directivo
Programa Adopte un Talento**

Figura A.8: Retribución social: clase a niños de primaria en el Programa Adopte un Talento (PAUTA).