

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, B.C.).

La investigación fue financiada por el CONACYT (Consejo Nacional de Ciencia y Tecnología).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos Autor.

CICESE © 2023, Todos los Derechos Reservados, CICESE

# Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



---

## Maestría en Ciencias en Ciencias de la Computación

---

### Reconocimiento de emociones a partir de patrones en la marcha humana.

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestro en Ciencias

Presenta:

**Yulith Vanessa Altamirano Flores**

Ensenada, Baja California, México

2023

Tesis defendida por

**Yulith Vanessa Altamirano Flores**

y aprobada por el siguiente Comité

---

Dr. Irvin Hussein López Nava

Codirector de tesis

---

Dr. Iván González Díaz

Codirector de tesis

Dr. Ubaldo Ruiz López

Dr. Humberto Pérez Espinosa



---

Dr. Pedro Gilberto López Mariscal

Coordinador del Posgrado en Ciencias de la Computación

---

Dr. Pedro Negrete Regagnon

Director de Estudios de Posgrado

Resumen de la tesis que presenta Yulith Vanessa Altamirano Flores como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación.

### **Reconocimiento de emociones a partir de patrones en la marcha humana.**

Resumen aprobado por:

---

Dr. Irvin Hussein López Nava

Codirector de tesis

---

Dr. Iván González Díaz

Codirector de tesis

El análisis de la marcha humana se ha utilizado ampliamente en el campo clínico, por ejemplo, para el diagnóstico temprano de algunas enfermedades. Por otro lado, es posible asociar patrones de movimiento durante la marcha con varios comportamientos humanos, como las emociones. El objetivo principal de esta tesis es generar modelos para clasificar tres emociones discretas: feliz, triste y enojado, considerando el estado neutral como una clase adicional, utilizando un conjunto de características extraídas de la posición 3D del esqueleto humano durante sesiones de marcha. Se realizó un análisis descriptivo de los datos para seleccionar los mejores subconjuntos de articulaciones para reconocer estas emociones. Los modelos se construyeron con los algoritmos: KNN, Random Forest y un meta-clasificador (Boosting). Los mejores resultados se obtuvieron con Boosting con el método de selección PCA con un mAP de 0.840 para datos balanceados y Random Forest con el método de selección UVA con un mAP de 0.886 para datos desbalanceados. Los resultados fueron prometedores al utilizar métodos basados en aprendizaje automático superficial.

**Palabras clave: Reconocimiento de emociones, emociones, clasificación de emociones, marcha humana, aprendizaje máquina**

Abstract of the thesis presented by Yulith Vanessa Altamirano Flores as a partial requirement to obtain the Master of Science degree in Computer's Science .

### **Recognition of emotions from patterns in the human gait.**

Abstract approved by:

---

Dr. Irvin Hussein López Nava

Thesis Co-Director

---

Dr. Iván González Díaz

Thesis Co-Director

Human gait analysis has been widely used in the clinical field, for example, for the early diagnosis of some diseases. On the other hand, it is possible to associate movement patterns during gait with various human behaviors, such as emotions. The main objective of this thesis is to generate models to classify three discrete emotions: happy, sad and angry, considering the neutral state as an additional class, using a set of features extracted from the 3D position of the human skeleton during walking sessions. A descriptive analysis of the data was performed to select the best subsets of joints to recognize these emotions. The models were built with the algorithms: KNN, Random Forest and a meta-classifier (Boosting). The best results were obtained with Boosting with the PCA selection method with a mAP of 0.840 for balanced data and Random Forest with the UVA selection method with a mAP of 0.886 for unbalanced data. The results were promising when using methods based on shallow machine learning.

**Keywords:** Emotions, emotion recognition, emotion classification, human gait, machine learning

## Dedicatoria

**A mi padre, por ser el pilar más importante y por demostrarme siempre su cariño y apoyo incondicional. A mi madre, por su apoyo en mi maestría sin importar nuestras diferencias de opiniones. A mi hermano, por ser un gran amigo para mí, que junto a sus ideas hemos pasado momentos inolvidables. A mis amigos, que me han apoyado en las buenas y en las malas.**

## Agradecimientos

Al Centro de Investigación Científica y de Educación Superior de Ensenada Baja California por la oportunidad de realizar los estudios de posgrado en el.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme el apoyo económico para realizar mis estudios de maestría. No. de becario: 20264680

A mis directores de tesis **Dr. Irvin Hussein López Nava y Dr. Iván Díaz González** por su esfuerzo y dedicación. Sus conocimientos, sus orientaciones, su manera de trabajar, su persistencia, sobre todo paciencia y su motivación han sido fundamentales para mi formación como investigador. A su manera, han sido capaz de ganarse mi lealtad y admiración, así como sentirme en deuda con ellos por todo lo recibido durante el periodo de tiempo que ha durado la tesis.

A mi comité de tesis Dr. Ubaldo Ruiz López y al Dr. Humberto Pérez Espinosa por sus comentarios y observaciones que ayudaron a mejorar mi trabajo.

# Tabla de contenido

	Página
Resumen en español .....	ii
Resumen en inglés .....	iii
Dedicatoria .....	iv
Agradecimientos .....	v
Lista de figuras .....	viii
Lista de tablas .....	x
<b>Capítulo 1. Introducción</b>	
1.1. Motivación .....	1
1.2. Planteamiento del problema .....	2
1.3. Objetivos .....	2
1.3.1. Objetivo general .....	2
1.3.2. Objetivos específicos .....	2
1.4. Preguntas de investigación .....	3
1.5. Propuesta de solución .....	3
1.6. Metodología .....	3
1.7. Contribuciones esperadas. ....	4
1.8. Organización de la tesis .....	4
<b>Capítulo 2. Fundamentos teóricos</b>	
2.1. Emociones .....	6
2.1.1. Tipos de emociones .....	6
2.2. Cinemática del movimiento humano .....	7
2.2.1. Marcha humana .....	7
2.2.2. Biomecánica de la marcha humana .....	8
2.3. Dataset .....	10
2.3.1. SIG .....	10
2.3.2. BML .....	10
2.3.3. The CMU graphics lab dataset .....	10
2.3.4. ICT .....	10
2.3.5. EWalk .....	11
2.3.6. Human 3.6M .....	11
2.3.7. Resumen de datasets .....	11
2.4. Técnicas .....	12
2.4.1. Métodos de selección automática de características .....	13
2.4.1.1. Métodos de filtros .....	13
2.4.1.2. Métodos de wrappers .....	14
2.4.2. Algoritmos de inferencia .....	14
<b>Capítulo 3. Trabajo relacionado</b>	
3.1. Emociones humanas .....	16

3.2.	Patrones de la marcha para reconocer emociones humanas . . . . .	16
3.3.	Tipos de datos . . . . .	21
3.4.	Algoritmos de aprendizaje de máquina . . . . .	21
3.5.	Algoritmos de aprendizaje profundo . . . . .	22
<b>Capítulo 4. Metodología</b>		
4.1.	Descripción de datos . . . . .	26
4.2.	Análisis descriptivo . . . . .	27
4.3.	Extracción de características . . . . .	29
4.4.	Selección de características . . . . .	31
4.5.	Clasificación de emociones . . . . .	31
<b>Capítulo 5. Resultados</b>		
5.1.	Análisis descriptivo . . . . .	34
5.1.1.	Análisis relativo de movimiento . . . . .	34
5.1.2.	Análisis inter-emociones . . . . .	35
5.1.3.	Selección de subconjuntos de articulaciones relevantes. . . . .	39
5.2.	Clasificación de emociones . . . . .	41
5.2.1.	Algoritmos de aprendizaje de máquina tradicionales . . . . .	41
5.2.1.1.	Datos balanceados . . . . .	41
5.2.1.2.	Datos desbalanceados . . . . .	43
5.2.2.	Algoritmos de aprendizaje profundo . . . . .	44
5.2.2.1.	Datos balanceados . . . . .	45
5.2.2.2.	Datos desbalanceados . . . . .	45
5.2.3.	Replicación de algoritmos del estado del arte . . . . .	46
5.3.	Comparativa de los mejores clasificadores . . . . .	47
5.3.1.	Mejores clasificadores para datos balanceados . . . . .	47
5.3.2.	Mejores clasificadores para datos desbalanceados . . . . .	47
5.3.3.	Discusión . . . . .	49
<b>Capítulo 6. Conclusiones y trabajo a futuro</b>		
6.1.	Conclusiones . . . . .	52
6.2.	Contribuciones . . . . .	53
6.3.	Limitaciones . . . . .	53
6.4.	Trabajo a futuro . . . . .	53
<b>Literatura citada</b> . . . . .		55
<b>Apéndice A. Listado de características extraídas</b> . . . . .		
.1.	Características temporales . . . . .	60
.2.	Características estadísticas . . . . .	60
.3.	Características espectrales . . . . .	61
<b>Apéndice B. Detalle de resultados: Tablas de los diferentes subconjunto de articulaciones.</b> . . . . .		
		63

## Lista de figuras

Figura	Página
1. Metodología del trabajo de investigación. . . . .	4
2. Ciclo de la marcha normal. Tomada de Sedgley (2013). . . . .	8
3. Representación de los escenarios de captura de los diferentes datasets. . . . .	12
4. Representación del cálculo del volumen. . . . .	17
5. Representación del cálculo de áreas. . . . .	18
6. Representación del cálculo de distancias. . . . .	18
7. Representación del cálculo de ángulos. . . . .	19
8. Representación del cálculo de las características para reconocer emociones. De izquierda a derecha: (a) El ejemplo para calcular el seno del ángulo del cuello, a partir de la diferencia de las coordenadas del cuello y orejas (triángulos cafés).(b) Para calcular el seno del ángulo del hombro derecho se cálculo con diferencia del hombro derecho con respecto al cuello (triángulos azules). (c) En el cálculo para la inclinación de la cabeza se obtuvo con las distancias de los hombros dividido sobre los datos del cuello (triángulos verdes). (d) Para calcular el ángulo del codo derecho se obtiene restando $m_1 - m_2$ , donde $m_1$ es la diferencia del hombro derecho y el codo derecho y $m_2$ representa la resta del codo derecho menos la muñeca (ángulos amarillos). (e) El cálculo del ángulo derecho de la rodilla se obtiene restando $m_3 - m_4$ , donde $m_3$ es la diferencia de la cadera derecha y de la rodilla derecha y $m_4$ es la resta de la rodilla derecha menos el tobillo derecho (ángulos rojos). . . . .	20
9. Red convolucional de grafos espacio-temporales. Primero realiza una estimación de pose en vídeos y se construye un grafo espacio temporal en secuencias de esqueleto. Después se aplican múltiples capas de convolución al grafo espacio temporal (STGCN) y se generarán gradualmente mapas de características de nivel superior en el gráfico. Luego, el clasificador estándar de Softmax lo clasificará en la clase correspondiente. Tomada de (Yan et al., 2018). . . . .	23
10. Red neuronal de grafos dirigidos. (a) Es el grafo original. (b) Muestra el procedimiento de actualización de vértices, donde el atributo del vértice en ( $v_2$ ) y los atributos de su borde entrante ( $e_1$ ) y bordes salientes ( $e_2$ y $e_3$ ) se combinan para obtener un vértice actualizado ( $v'_2$ ). (c) Muestra el procedimiento de actualización de bordes, donde el atributo de la propia arista ( $e_1$ ) y los atributos de su vértice de origen ( $v_1$ ) y vértice de destino ( $v'_2$ ) se combinan para obtener una arista actualizada ( $e'_1$ ). El círculo azul representa el borde (o vértice) que se está actualizando. El círculo naranja y el círculo rojo representan el vértice de origen (o borde de entrada) y el vértice de destino (o borde de salida) que están involucrados en la actualización. Tomada de (Shi et al., 2019). . . . .	23
11. Arquitectura STEP: clasificador para predecir emociones percibidas a partir de marcha humana. Tomada de Bhattacharya et al. (2020a). . . . .	24
12. Arquitectura empleando LSTM para la clasificación de emociones a partir de la marcha. Tomada de (Randhavane et al., 2019). . . . .	25
13. Arquitectura para la clasificación semi-supervisada de emociones discretas percibidas a partir de la marcha. Tomada de (Bhattacharya et al., 2020b). . . . .	25

Figura	Página
14. Metodología implementada. . . . .	26
15. Esqueleto representado por un conjunto de 21 articulaciones. . . . .	27
16. Distribución de clases para datos desbalanceados $D$ . . . . .	27
17. Datos de muestra balanceados agrupados por par de clases. . . . .	27
18. Enfoque para calcular la cantidad de movimiento promedio relativo. Inter-joint: diferentes articulaciones, mismo tiempo (18a). Intra-joint: misma articulación, tiempos diferentes (18b). . . . .	28
19. Arquitectura de las redes profundas basadas en CNN'S. . . . .	33
20. Enfoque inter-joint: Los mapas de calor representan la cantidad de movimiento de todas las articulaciones. . . . .	35
21. Enfoque intra-joint: Los mapas de calor representan la cantidad de movimiento de todas las articulaciones. . . . .	35
22. Enfoque inter-joint: Los mapas de calor representan la diferencia de movimiento de todas las articulaciones por pares de emociones. . . . .	36
23. Enfoque intra-joint: Los mapas de calor representan la diferencia de movimiento de todas las articulaciones por pares de emociones. . . . .	36
24. Enfoque inter-joint: Las gráficas de caja y pivote representan la diferencia de movimiento de todas las articulaciones por pares de emociones. . . . .	37
25. Enfoque intra-joint: Las gráficas de caja y pivote representan la diferencia de movimiento de todas las articulaciones por pares de emociones. . . . .	37
26. Enfoque inter-joint: Comparación de la relevancia de las articulaciones. . . . .	38
27. Enfoque intra-joint: Comparación de la relevancia de las articulaciones. . . . .	38
28. Correlación de Pearson aplicada a las 21 articulaciones para todas las emociones. . . . .	39
29. Subconjuntos de la representación completa. . . . .	40
30. Matrices de confusión para los mejores clasificadores usando datos balanceados. . . . .	49
31. Matrices de confusión para los mejores clasificadores usando datos desbalanceados. . . . .	50
32. Datos balanceados. . . . .	50
33. Datos desbalanceados. . . . .	50
34. Comparación de clasificadores de cada emoción. . . . .	50

## Lista de tablas

Tabla	Página
1. Rango de movimiento en las fases de la marcha humana. . . . .	9
2. Resumen de los conjuntos de datos. . . . .	12
3. Características extraídas de acuerdo a los grados de libertad. . . . .	19
4. Conjunto de 21 articulaciones para datos balanceados. . . . .	42
5. Subconjunto de 9 articulaciones para datos balanceados. . . . .	43
6. Conjunto de 21 articulaciones para datos desbalanceados. . . . .	44
7. Subconjunto de 9 articulaciones para datos desbalanceados. . . . .	44
8. Subconjunto de 21 articulaciones para datos balanceados para algoritmos de aprendizaje profundo. . . . .	45
9. Subconjunto de 21 articulaciones para datos desbalanceados para algoritmos de aprendizaje profundo. . . . .	45
10. Resultados reportados en el Estado del Arte. . . . .	46
11. Replicación de algoritmos y parámetros del Estado del Arte para 21 articulaciones. . . . .	46
12. Subconjunto de 14 articulaciones para datos balanceados. . . . .	63
13. Subconjunto de 12 articulaciones para datos balanceados. . . . .	63
14. Subconjunto de 10 articulaciones para datos balanceados. . . . .	63
15. Subconjunto de 6 articulaciones para datos balanceados. . . . .	63
16. Subconjunto de 5 articulaciones para datos balanceados. . . . .	64
17. Subconjunto de 4 articulaciones para datos balanceados. . . . .	64
18. Subconjunto de 14 articulaciones para datos desbalanceados. . . . .	64
19. Subconjunto de 12 articulaciones para datos desbalanceados. . . . .	64
20. Subconjunto de 10 articulaciones para datos desbalanceados. . . . .	65
21. Subconjunto de 6 articulaciones para datos desbalanceados. . . . .	65
22. Subconjunto de 5 articulaciones para datos desbalanceados. . . . .	65
23. Subconjunto de 4 articulaciones para datos desbalanceados. . . . .	65
24. Subconjunto de 14 articulaciones estado del arte. . . . .	66
25. Subconjunto de 12 articulaciones estado del arte. . . . .	66
26. Subconjunto de 10 articulaciones estado del arte. . . . .	66
27. Subconjunto de 9 articulaciones estado del arte. . . . .	66
28. Subconjunto de 6 articulaciones estado del arte. . . . .	67
29. Subconjunto de 5 articulaciones estado del arte. . . . .	67

30. Subconjunto de 4 articulaciones estado del arte. . . . . 67

# Capítulo 1. Introducción

---

En este capítulo se presenta una visión general de la investigación. Se inicia exponiendo la motivación que inspiró este trabajo, además se presenta el planteamiento del problema y los objetivos planteados al inicio de la investigación. Posteriormente se expone la hipótesis de la investigación y la propuesta de solución. Además se presenta la metodología a seguir al inicio de la investigación y por último se presenta brevemente la organización de la tesis.

## 1.1. Motivación

El análisis de la marcha humana se ha utilizado principalmente en el campo clínico, por ejemplo, para el diagnóstico temprano de algunas enfermedades como el Alzheimer (Beauchet et al., 2008) o el Parkinson (Muro-De-La-Herran et al., 2014). Sin embargo, se han llevado a cabo otros estudios en los que los patrones de movimiento en la marcha se pueden asociar con el comportamiento humano, incluidos los estados emocionales como ira, felicidad, tristeza o miedo (Roether et al., 2009a).

Actualmente el reconocimiento de emociones se ha convertido en un área de investigación muy popular, surgiendo un gran número de trabajos para reconocer emociones a partir de las expresiones faciales (Mase, 1991). El rostro se encarga de comunicar no solo pensamientos o ideas, sino también emociones (Frank y Stennett, 2001). Estas emociones se pueden identificar a través de capturas estáticas (imágenes) de dichos gestos faciales. El lenguaje gestual ocupa un importante lugar en la comunicación humana, a través de la cual se intercambian y construyen significados con otros individuos, a lo largo de todo el ciclo de la vida. Los individuos adquieren la habilidad de utilizar este lenguaje para manifestar sus necesidades, expresar sus sensaciones y sentimientos, y vincularse con otras personas así como con el mundo que los rodea (Acredolo y Goodwyn, 1985). Existen otros gestos a partir de los cuales se comunica el individuo, como el llamado lenguaje de señas, la frecuencia del parpadeo y postura corporal (Glowinski et al., 2008). Sin embargo, hay muchos casos en los que la captura de lenguaje corporal puede ser poco fiable, por situaciones en las que haya oclusión total o parcial, cuestiones relacionadas con la cercanía y lejanía (escala) y por cuestiones relacionadas con la iluminación, entre otras (Wang et al., 2020).

Las emociones no solo están involucradas en el proceso cognitivo de las personas, sino que son importantes para la regulación de las interacciones sociales humanas (Xu et al., 2020). Teniendo en cuenta que una de las principales actividades de la vida diaria es caminar, existe la posibilidad de procesar una gran cantidad de información a partir de diferentes fuentes y sistemas de sensores heterogéneos útiles

para analizar la marcha humana (o el ciclo de la marcha humana) y que pueden relacionarse con las emociones humanas (Schmelzer, 2020).

Una vez que se decide poner el foco de atención en la relación entre el estado emocional y la manera de caminar, siendo ésta analizada empleando sistemas de visión para la captura de movimiento, se deben estudiar las técnicas basadas en aprendizaje automático (ML) y aprendizaje profundo (DL), las cuales permiten construir modelos de clasificación capaces de aprender de diferentes tipos de datos; como imágenes, vídeos o series temporales. Los sistemas de captura de movimiento (MoCaps) permiten la reconstrucción y el seguimiento en 3D de las articulaciones del cuerpo humano. Recientemente, se han implementado arquitecturas profundas que permiten el seguimiento del movimiento como lo hacen los MoCaps. Independientemente de cómo se capture el movimiento, el cuerpo humano se puede representar mediante puntos clave 3D en el tiempo. En esta medida, estas series temporales se pueden utilizar para el entrenamiento de los modelos ML y DL.

## **1.2. Planteamiento del problema**

La complejidad de reconocer emociones radica en que, al momento, no hay un acuerdo en cómo se definen las emociones, pese a que se han propuesto modelos discretos y continuos, más aún en cómo las emociones se pueden medir de forma objetiva. Por lo que la interpretación de cada emoción tendrá asociado un grado de subjetividad. Por un lado, las personas expresan las emociones de modos distintos, estas variaciones dependen principalmente de la personalidad, de la cultura étnica, de la fisiología de la persona así como su rango de edad. Por otro lado, los patrones que representan cada emoción pueden variar de manera considerable de acuerdo al tipo de movimiento corporal, como gestos faciales, poses o incluso la marcha humana.

## **1.3. Objetivos**

### **1.3.1. Objetivo general**

Estudiar las emociones humanas y su relación con la marcha humana de sujetos jóvenes (entre un rango de edad 22-35 años) a partir del seguimiento de sus articulaciones.

### **1.3.2. Objetivos específicos**

- Preparar un conjunto de datos conformado por las articulaciones extraídas de vídeos de marcha humana de sujetos jóvenes.

- Generar modelos explicativos, que permitan seleccionar el mejor subconjunto de articulaciones al analizar la marcha humana.
- Proponer algoritmos de aprendizaje de máquina y aprendizaje profundo para que sean entrenados con los datos extraídos previamente.
- Evaluar los modelos generados con nuevos datos.

#### 1.4. Preguntas de investigación

Con base en la problemática expuesta anteriormente surgen las siguientes preguntas de investigación: ¿Qué emociones se pueden reconocer en la marcha humana de sujetos jóvenes a partir del seguimiento de las articulaciones usando vídeo? y ¿Con que precisión se puede reconocer un conjunto de emociones a partir de datos de la marcha empleando modelos de aprendizaje automático y profundo?

#### 1.5. Propuesta de solución

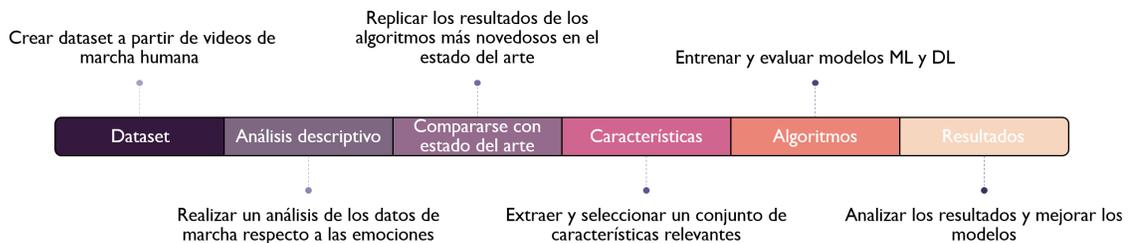
Para atender al problema planteado se propone la extracción de un vector de características que incorpore patrones de la marcha humana, con el objetivo de entrenar diferentes algoritmos ML y DL para clasificar tres emociones básicas: felicidad, tristeza y enojo, además del estado neutral. Finalmente, los resultados obtenidos en los experimentos serán comparados con los resultados reportados en el Estado del Arte, con el fin de verificar si es posible reducir la cantidad de características sin perder precisión.

#### 1.6. Metodología

La metodología seguida durante el trabajo de investigación se divide en 6 fases, que se describen a continuación (Figura 1).

- **Fase 1: Selección del conjunto de datos.** Se realizó una búsqueda exhaustiva para conformar un conjunto de datos que incluyera vídeos de marcha humana de sujetos jóvenes con un estilo asociado a las emociones de interés.
- **Fase 2: Análisis descriptivo.** Se analizaron los datos desde múltiples perspectivas con el fin de buscar la información relevante con respecto al conjunto de articulaciones que mejor caracterizan cada emoción.

- **Fase 3: Comparación con el Estado del Arte.** Se realizó una búsqueda de algoritmos propuestos para el reconocimiento de emociones a partir de la marcha, para replicarlos con el mismo conjunto de datos, y con fines de comparación.
- **Fase 4: Extracción y análisis de características.** Se realizó una extracción y selección de características temporales, estadísticas y espectrales, a partir de las articulaciones representadas por series de tiempo.
- **Fase 5: Entrenamiento de modelos de clasificación.** Con las características extraídas se entrenaron los modelos de aprendizaje de máquina y aprendizaje profundo para la clasificación de emociones.
- **Fase 6: Análisis y comparación de resultados.** Finalmente, se evaluaron los resultados obtenidos con los modelos de clasificación a partir de diferentes métricas.



**Figura 1.** Metodología del trabajo de investigación.

## 1.7. Contribuciones esperadas.

Como resultados de éste trabajo se esperan las siguientes contribuciones:

- Análisis descriptivo de la relación de un conjunto de emociones humanas respecto a la marcha humana de sujetos jóvenes.
- Generación de modelos de aprendizaje a partir de un conjunto de características que sean robustas al problema de clasificación con clases desbalanceadas.

## 1.8. Organización de la tesis

El resto de la tesis se desglosa en los siguientes capítulos:

En el **Capítulo 2** (*Fundamentos teóricos*) se introducen los conceptos de emociones, marcha humana, las características del conjunto de datos y las técnicas utilizadas a lo largo del presente trabajo de tesis.

En el **Capítulo 3** (*Trabajo relacionado*) se revisan los trabajos más importantes en el área del reconocimiento de emociones.

En el **Capítulo 4** (*Metodología*) se aborda la metodología implementada para el reconocimiento de emociones a partir de patrones de la marcha.

En el **Capítulo 5** (*Resultados*) se presentan los resultados obtenidos en los experimentos al implementar la metodología propuesta.

Por último, en el **Capítulo 6** (*Conclusiones y trabajo a futuro*) se presentan las conclusiones derivadas del trabajo de tesis, las contribuciones, limitaciones y el trabajo a futuro.

## Capítulo 2. Fundamentos teóricos

---

En este capítulo se introducen los conceptos básicos y las técnicas utilizadas en el presente trabajo de tesis. En la primera y segunda sección se describen los conceptos básicos de emociones y de la marcha humana, respectivamente. En la tercera sección se describen los conjuntos de datos que se emplearon. Por último, en la cuarta sección se describen los conceptos relacionados a las técnicas de aprendizaje automático y profundo utilizadas para tareas de clasificación.

### 2.1. Emociones

Las emociones son reacciones psicofisiológicas que representan modos de adaptación del individuo cuando percibe un objeto, persona, lugar, suceso o recuerdo importante. El tipo de emoción que experimenta una persona está determinado por la circunstancia que desencadena la emoción (Trampe et al., 2015).

#### 2.1.1. Tipos de emociones

Existen seis emociones básicas que son universales en todas las culturas humanas: miedo, disgusto, enojo, sorpresa, felicidad y tristeza (Ekman, 1999). Conforme pasaron los años se agregaron nuevos métodos para clasificar las emociones, por ejemplo en 1980, Robert Plutchik introdujo un sistema de clasificación de emociones conocido como la rueda de las emociones. Este modelo demostró cómo se pueden combinar o mezclar diferentes emociones, de la misma manera que un artista mezcla colores primarios para crear otros colores (Plutchik, 1984). Hay varias teorías sobre cuántos tipos de emociones experimentan los seres humanos. Las emociones universales que propuso el psicólogo Paul Ekman son las siguientes:

- **Felicidad:** Normalmente las personas persiguen esta emoción, debido a que produce una sensación de bienestar y satisfacción. La felicidad a menudo se expresa sonriendo, hablando en un tono de voz optimista. Las personas tienden a caminar más rápido y más erguidas, balancean los brazos y se mueven más hacia arriba y hacia abajo, balanceándose menos de lado a lado que las personas tristes o deprimidas (Reddy, 2014).
- **Tristeza:** Todos los seres humanos experimentan tristeza de vez en cuando. Alguien puede externar tristeza de diversas formas, por ejemplo, una velocidad de marcha reducida, un balanceo de brazos limitado, movimientos verticales de cabeza reducidos, movimientos de balanceo lateral más grandes en la parte superior del cuerpo y una postura más caída (Michalak et al., 2009).

- Miedo: Esta emoción acelera la frecuencia cardíaca, causa pensamientos acelerados o desencadena la respuesta de lucha o huida. Puede ser provocada por reacción a amenazas reales o percibidas. La persona puede tender a quedarse inmóvil, sin embargo, se pueden presentar patrones de temblor en ciertas partes del cuerpo (Kozłowska et al., 2015).
- Disgusto: Se desencadena por una experiencia física, como ver u oler alimentos podridos, sangre o mala higiene. El disgusto moral puede ocurrir cuando alguien ve a otra persona haciendo algo que encuentra inmoral o desagradable. En general se traduce en gestos faciales evidentes (Oaten et al., 2018).
- Enojo: Se puede manifestar con expresiones faciales como fruncir el ceño, gritar o como comportamiento en la marcha ambulando con la cabeza en alto, hombros hacia atrás, pecho hacia afuera, pasos largos y rápidos. También puede manifestarse en algunos casos con rigidez muscular en algunas partes del cuerpo, como los pies durante la marcha (Martin, 2020).
- Sorpresa: La sorpresa puede ser agradable o desagradable. Al experimentar esta emoción se desencadenan reacciones que repercuten en el rostro, como expresiones en la boca. La sorpresa, como el miedo, puede desencadenar la respuesta de lucha o huida. Al contrario que otras emociones, la movilidad corporal se libera con algunas manifestaciones evidentes como saltos (Zhu et al., 2019).

## 2.2. Cinemática del movimiento humano

El cuerpo humano se entiende como una estructura arquitectónica con capacidad de movimiento. La cinemática se define como el movimiento independiente de las articulaciones o segmentos del cuerpo. Este movimiento ocurre sin que una fuerza externa actúe sobre la articulación o el segmento del cuerpo (Seidenberg y Beutler, 2008).

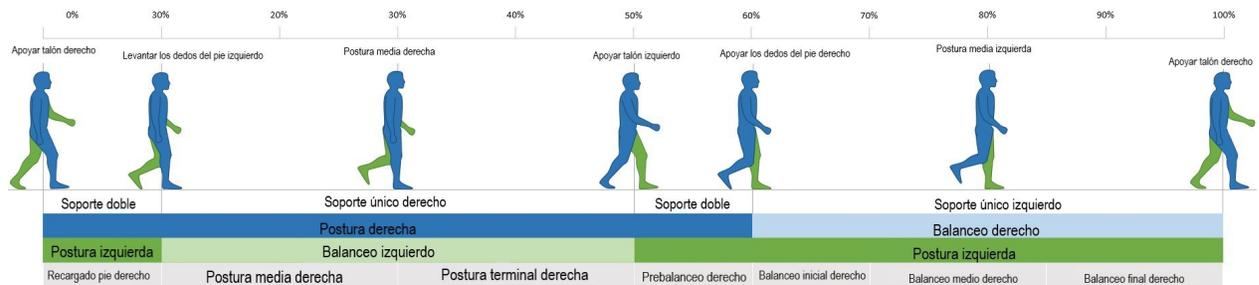
### 2.2.1. Marcha humana

La marcha humana se define como una serie de movimientos alternos de las extremidades inferiores en un movimiento rítmico que resulta en la progresión hacia adelante del cuerpo con un gasto mínimo de energía (Tekscan, 2021). Cabe aclarar que la marcha humana no solo involucra las extremidades inferiores si no también las extremidades superiores y el resto del cuerpo.

La evaluación de la marcha ayuda a comprender asimetrías específicas relacionadas con la forma en que nos movemos. La medición de la marcha se puede realizar utilizando información cinética (fuerza) y

cinemática (espacial / temporal) (Tekscan, 2020), debido a la naturaleza de los datos nos enfocaremos principalmente en los datos cinemáticos.

La cinemática es el estudio del movimiento del cuerpo y de los segmentos del mismo sin referencia a las fuerzas que pueden estar actuando sobre el cuerpo. Por ejemplo, durante la marcha normal (no patológica) hay una división evidente entre el tiempo que el pie está en contacto con el suelo y el tiempo en que no lo está. Estos se conocen como la “fase de apoyo o postura” (aproximadamente el 60 % del ciclo de la marcha) y la “fase de oscilación o balanceo” (aproximadamente el 40 % del ciclo de la marcha), respectivamente. Mientras tanto, los brazos se desplazan hacia delante y hacia atrás en dirección opuesta a la de las piernas, por ejemplo, el brazo izquierdo se desplaza con la pierna derecha hacia delante, mientras el brazo derecho lo hace hacia atrás. El ciclo de la marcha normal (no patológica) se ilustra en la Figura 2.



**Figura 2.** Ciclo de la marcha normal. Tomada de Sedgley (2013).

### 2.2.2. Biomecánica de la marcha humana

A continuación se describen los arcos de movilidad que permiten las articulaciones principales y después cómo estas articulaciones se conjuntan para el desplazamiento durante la marcha humana, ver Tabla 1 (Physiopedia, 2022).

- Contacto inicial (golpe de talón)
  - Ocurre al 0 % del ciclo de la marcha.
- Plantar pie (respuesta de carga)
  - Ocurre en el 8-10 % del ciclo de la marcha.
- Postura media (apoyo de una sola pierna)
  - El trocánter mayor está verticalmente por encima del punto medio del pie.

- Levantar tobillo  
Ocurre alrededor del 30-40 % del ciclo de la marcha.
- Despegar dedos del suelo (prebalanceo)  
Ocurre al 60 % del ciclo de la marcha (fase final de la postura).
- Balanceo temprano  
Ocurre en el 60-75 % del ciclo de la marcha (comienzo de la fase de balanceo).
- Balanceo medio  
Ocurre en el 75-85 % del ciclo de la marcha.
- Balanceo tardío  
Ocurre en el 85-100 % del ciclo de la marcha.

**Tabla 1.** Rango de movimiento en las fases de la marcha humana.

Fase	Articulación	Rango de movimiento
Contacto inicial	Tobillo	0° (posición neutral)
	Rodilla	0° (extensión completa)
	Cadera	media 20° de flexión
Plantar pie	Tobillo	0-5° flexión plantar
	Rodilla	15° de flexión
	Cadera	15° de flexión (cadera hacia la extensión)
Postura media	Tobillo	5° de dorsiflexión
	Rodilla	5° de flexión
	Cadera	0° de flexión (posición neutra)
Levantar tobillo	Tobillo	0° (posición neutral)
	Rodilla	0° de flexión (extensión completa)
	Cadera	10-20° de hiperextensión
Despegar dedos del suelo (Prebalanceo)	Tobillo	20° de flexión plantar
	Rodilla	30° de flexión
	Cadera	10-20° de hiperextensión
Balanceo temprano	Tobillo	10° de flexión plantar
	Rodilla	60° de flexión
	Cadera	se mueve a 20° de flexión
Balanceo medio	Tobillo	0° (posición neutral)
	Rodilla	se mueve a 30° de flexión
	Cadera	30° de flexión (cadera hacia la extensión)
Balanceo tardío	Tobillo	0° (posición neutral)
	Rodilla	0° (extensión completa)
	Cadera	30° de flexión

## 2.3. Dataset

Se seleccionaron seis conjuntos de datos disponibles públicamente que incluyen dos tipos de marchas. En la primera (BML, ICT y Human 3.6M), se les proporcionaba a los sujetos instrucciones específicas de cómo representar una emoción, a estas, nos referiremos como marchas actuadas o solicitadas. En el segundo tipo (CMU, SIG y EWalk), los sujetos representaron la emoción de forma natural sin instrucciones, a estas no referiremos como marchas no actuadas e improvisadas. Además se incluyeron marchas neutrales, es decir, que no hubo una emoción manifestada aparente. Las principales características de estos conjuntos de datos se presentan a continuación.

### 2.3.1. SIG

El conjunto de datos SIG comprende marchas sintéticas generadas usando mezclas locales de modelos autorregresivos, para generar diferentes estilos de movimiento. Para replicar los estilos de caminar, capturaron datos con un sistema Vicon MoCap. El conjunto contiene una amplia variedad de actividades que registraron diferentes estilos pero solo son de interés para este trabajo: enojado, triste y neutral (Xia et al., 2015).

### 2.3.2. BML

El conjunto de datos contiene diferentes vídeos de marchas de 30 actores no profesionales (15 mujeres y 15 hombres) con una edad promedio de 22 años, utilizando el sistema Falcon Analog optical MoCap. Los sujetos caminaron en un triángulo durante 30 segundos, girando en el sentido de las agujas del reloj y luego en el sentido contrario a las agujas del reloj, con 4 estilos diferentes de caminar en ambas direcciones. Capturaron cuatro clases: feliz, triste, enojado y el estado neutral (Ma et al., 2006).

### 2.3.3. The CMU graphics lab dataset

Contiene 49 vídeos de personas (edad promedio de 22 años) caminando con diferentes estilos. Los sujetos no recibieron instrucciones sobre qué emociones debían realizar, sino que decidieron representar las emociones que tenían instantáneamente mientras realizaban diferentes actividades (Carnegie, 2001).

### 2.3.4. ICT

El conjunto de datos comprende 22 vídeos de marcha de sujetos universitarios con una edad promedio de 27 años (11 mujeres y 11 hombres). Se instruyó a los sujetos para que representaran cuatro emociones:

enojado, triste, feliz y neutral, y se les pidió que etiquetaran sus propios vídeos (Narang et al., 2017).

### **2.3.5. EWalk**

El conjunto de datos EWalk contiene vídeos RGB de 24 estudiantes (16 hombres y 8 mujeres) con una edad media de 34.5 años. Se sugirió a los sujetos que asumieran que estaban experimentando una determinada emoción y caminaran en consecuencia. Los vídeos fueron grabados tanto en ambientes interiores como exteriores (Randhavane et al., 2019).

### **2.3.6. Human 3.6M**

El dataset Human 3.6M contiene grabaciones de vídeo de actores profesionales (5 mujeres y 6 hombres) que realizaron diferentes actividades usando cuatro cámaras calibradas trabajando simultáneamente. Los sujetos recibieron instrucciones específicas sobre cómo representar las emociones de enojo y felicidad mientras caminaban (Ionescu et al., 2013).

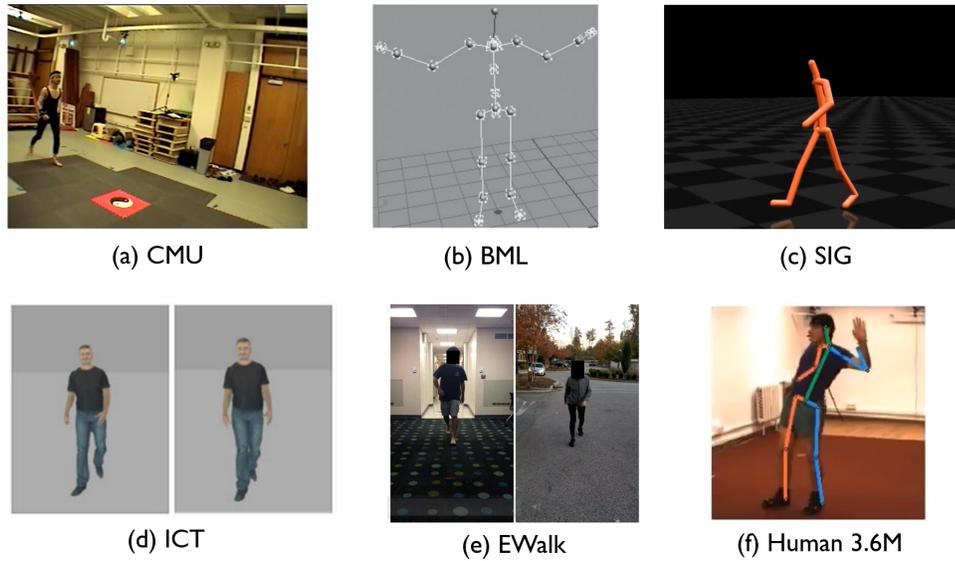
### **2.3.7. Resumen de datasets**

En la Tabla 2 se presenta un resumen de los seis conjuntos de datos empleados, mostrando las características más relevantes para la siguiente pregunta de investigación planteada:

¿Qué emociones se pueden reconocer en la marcha humana de sujetos jóvenes a partir del seguimiento de las articulaciones usando vídeo?

Los datasets se capturaron bajo diferentes condiciones las cuales se describen a continuación:

En el dataset CMU se realizó la captura de vídeos colocando cámaras en un área rectangular de aproximadamente 3m x 8m en el centro de la sala, como se muestra en la Figura 3a. Para el dataset BML el área designada para que los sujetos caminaran fue un triángulo y caminaron durante 30 segundos, para tener diferentes perfiles del sujeto ellos girando en el sentido de las agujas del reloj y luego en el sentido contrario (ver Figura 3b). En el dataset SIG emplearon los datos CMU para poder sintetizar datos de la marcha, ver Figura 3c. Los autores del dataset ICT pidieron a los sujetos caminar 10m en línea recta y en un círculo de 3m de radio, así como giros en el lugar, caminar a un lado, entre otros (ver Figura 3d). En EWalk los sujetos caminaron en un área designada de 7m en dirección a una única cámara fija, (ver Figura 3e). Por último en el dataset Human 3.6M el área designada fue de 6m de diámetro, donde la captura efectiva abarcaba un diámetro de 4 metros (ver Figura 3f).



**Figura 3.** Representación de los escenarios de captura de los diferentes datasets.

**Tabla 2.** Resumen de los conjuntos de datos.

Referencias	Conjunto de datos	# Marchas	# Sujetos	Promedio de edad	Sensor	Emoción	Tipo de captura	Etiquetado
Carnegie (2001)	CMU	49	17	22 años	Sistema de captura de movimiento Vicon	Presentaron las emociones que sintieron en ese momento.	No actuado	No etiquetado
Ma et al. (2006)	BML	120	30	22 años	Falcon Analog optical MoCap	Feliz Triste Enojado Neutral	Actuado	Etiquetado
Xia et al. (2015)	SIG	41	Datos sintéticos	Datos sintéticos	Modelos MAR	Datos sintéticos	Datos sintéticos	Etiquetado
Narang et al. (2017)	ICT	24	22	27.13 años	Modelos	Feliz Triste Enojado Neutral	Actuado	Etiquetado
Randhavane et al. (2019)	EWALK	84	24	34.5 años	Cámara	Feliz Triste Enojado Neutral	No actuado	Etiquetado
Ionescu et al. (2013)	Human 3.6M	14	11	25.8 años	Cámara	Feliz Enojado	Actuado	No etiquetado

Se extrajeron las posiciones de las articulaciones a partir de los vídeos de marcha y se normalizaron a 21 articulaciones. Posteriormente se segmentaron y etiquetaron para que tuvieran el mismo número de *frames*, con el objetivo de que los datos estuvieran bajo las mismas condiciones sin importar los diferentes protocolos de captura. Finalmente el conjunto de datos resulto en 1856 segmentos de vídeos de 240 *frames* que representan las marchas.

## 2.4. Técnicas

En esta sección, se describen las técnicas empleadas para la selección automática de características y los algoritmos de inferencia utilizados en el presente trabajo.

### 2.4.1. Métodos de selección automática de características

La selección de características se emplea debido a que se quieren obtener las características más importantes y/o relevantes del conjunto de datos, con el objetivo de mejorar el rendimiento de predicción de los modelos de clasificación, proporcionar predictores más rápidos y más rentables, y proporcionar una mejor comprensión del proceso que generó los datos.

Normalmente cuando se cuenta con una alta dimensionalidad de características se procede a aplicar un método de selección para reducir dichas dimensiones. Esto se realiza para evitar que el exceso de información pueda actuar como ruido principalmente para los modelos de aprendizaje de máquina, provocando un rendimiento extremadamente bajo, además de que el modelo pueda tardar más tiempo en entrenarse y se requiera una asignación de recursos innecesarios en características no relevantes. Para la selección de características se tomaron dos categorías filtros y wrappers, los cuales se explican a continuación.

#### 2.4.1.1. Métodos de filtros

Los métodos se dividen según sus puntajes estadísticos que determinan la correlación de las características con respecto a la variable de clase (Gonzalez, 2020). Los métodos de selección de características que utilizan filtros son los siguientes:

- **PCA**

El análisis de componentes principales, es un método estadístico que permite reducir la dimensionalidad de grandes conjuntos de datos, transformando un gran conjunto de variables en uno más pequeño que aún contiene la información relevante, permitiendo que el tiempo para entrenamiento se reduzca considerablemente. Supóngase que existe una muestra con  $n$  individuos cada uno con  $p$  variables  $(X_1, X_2, \dots, X_p)$ , es decir, el espacio muestral tiene  $p$  dimensiones. PCA permite encontrar un número de factores subyacentes ( $z < p$ ) que explican aproximadamente lo mismo que las  $p$  variables originales. Donde antes se necesitaban  $p$  valores para caracterizar a cada individuo, ahora bastan  $z$  valores. Cada una de estas  $z$  nuevas variables recibe el nombre de componente principal (Omuya et al., 2021).

- **ANOVA F-test**

ANOVA viene del acrónimo de análisis de varianza y es una prueba de hipótesis estadística paramétrica que permite determinar si las medias de dos o más muestras de datos provienen de la misma distribución o no.

F-test, es una clase de prueba estadísticas que calcula la relación entre los valores de varianzas, como la varianza de dos muestras diferentes o la varianza explicada e inexplicable mediante una prueba estadística, como ANOVA.

El método estadístico ANOVA F-test se emplea con variables categóricas, las cuales permiten eliminar características siempre y cuando las variables sean independientes de la variable destino (Rodrigo, 2017).

#### 2.4.1.2. Métodos de wrappers

Estos métodos realizan la selección de características mediante un algoritmo de inferencia, es decir, van tomando el mejor subconjunto mediante las inferencias que se extraen de un algoritmo, para agregar o eliminar características de su subconjunto (Gonzalez, 2020). A continuación, se explica el método de importancia de *Gini* que se utilizó para seleccionar las características usando *wrappers*.

El método de importancia de *Gini* o disminución media de la impureza (MDI) calcula la importancia de cada característica como la suma sobre el número de divisiones (en todos los árboles) que incluyen la característica, proporcionalmente al número de muestras que la divide (Lee, 2020).

#### 2.4.2. Algoritmos de inferencia

- **k-Nearest Neighbors (kNN)**

Es un método no paramétrico que no necesita ningún modelado o entrenamiento explícito antes de la fase de clasificación. Para clasificar una nueva observación, el algoritmo kNN calcula las distancias entre la nueva observación y todos los puntos de entrenamiento. La nueva observación se asigna a la clase más común según el voto mayoritario de sus k-vecinos más cercanos (Fix y Hodges, 1989).

- **Random forest (RF)**

Es una implementación de la estrategia *bagging* que utiliza un ensamble de árboles de clasificación. Cada uno de los árboles de clasificación se construye utilizando una muestra *bootstrap* del conjunto de entrenamiento, y en cada división, el conjunto de atributos candidatos es un subconjunto aleatorio de atributos. Cada clasificador contribuye con un solo voto para la asignación de la clase más frecuente (Ho, 1995).

- **Boosting**

Produce una regla de predicción precisa al combinar reglas aproximadas y moderadamente inexactas. Este meta-algoritmo aplica un algoritmo de clasificación base a versiones repetidamente modificadas de los datos, y las predicciones de todos ellos se combinan a través de un voto de mayoría ponderada para producir la predicción final (Freund y Schapire, 1997). El algoritmo base seleccionado fue *Support Vector Machines*, un método de aprendizaje estadístico que es capaz de clasificar información no vista derivando características seleccionadas y construyendo un hiperplano de alta dimensión para separar los puntos de datos en dos clases para formar un modelo de decisión (Cortes y Vapnik, 1995).

- **Convolutional Neural Networks (CNN)**

Es una red neuronal multicapa la cual consiste en un conjunto finito de capas de procesamiento que pueden aprender varias características de los datos de entrada (por ejemplo, imágenes) con múltiples niveles de abstracción. Las capas de iniciación aprenden y extraen las características de alto nivel (con menor abstracción), y las capas más profundas aprenden y extraen las características de bajo nivel (con mayor abstracción) (Ghosh et al., 2020).

## Capítulo 3. Trabajo relacionado

---

En este capítulo, se revisan los trabajos más importantes en el área de reconocimiento de emociones a partir del análisis de la marcha humana utilizando algoritmos de aprendizaje de máquina y aprendizaje profundo. En la primera sección se presentan estrategias para reconocer las emociones humanas. En la segunda sección se abordan los patrones de la marcha humana utilizados en la literatura para reconocer emociones. En la tercera sección se introducen los niveles de datos para representar la marcha humana. Por último, en la cuarta y quinta sección se presentan los trabajos más recientes en los que se clasifican emociones humanas a partir de datos de la marcha utilizando algoritmos de aprendizaje de máquina y aprendizaje profundo.

### 3.1. Emociones humanas

Como se definió previamente, las emociones son reacciones psicofisiológicas que representan modos de adaptación del individuo cuando percibe un objeto, persona, lugar, suceso o recuerdo importante. Las emociones están presentes en la vida diaria aproximadamente un 90 % del tiempo, son un factor importante que nos permite interactuar con nuestro alrededor, y de existir un desequilibrio emocional puede afectar nuestra toma de decisiones y la convivencia con otras personas (Trampe et al., 2015).

Existen diversas formas de reconocer las emociones, como expresiones faciales (Benitez-Quiroz et al., 2017), movimiento corporal (Ahmed et al., 2019; Crane y Gross, 2007; Ahmed y Gavrilova, 2019), postura (Janssen et al., 2008; Kang y Gross, 2016; Randhavane et al., 2019; Bhattacharya et al., 2020a), gestos (Glowinski et al., 2008), señales fisiológicas (Picard y Klein, 2002), información lingüística (Alm et al., 2005) y características acústicas (Dellaert et al., 1996).

Una alternativa que no ha sido tan explorada como las anteriores es la identificación de emociones a partir del análisis del movimiento y postural en la marcha humana, este trabajo se enfoca principalmente en reconocer emociones (feliz, triste, enojado y el estado neutral) a partir del análisis de movimiento y postural, extrayendo una familia.

### 3.2. Patrones de la marcha para reconocer emociones humanas

Estudios recientes revelan que la marcha transmite información sobre la emoción de las personas, dependiendo de la emoción que presenta cada persona se pueden apreciar diferentes patrones en la misma (Xu et al., 2020).

En la revisión del trabajo relacionado se encontró que para reconocer emociones algunos autores como *Randhavane et al. (2019)* extraen características afectivas que, principalmente, están basadas en la parte psicológica y postural, así como del movimiento. Estas características de postura se basan en trabajos previos de *Crane y Gross (2007)*, en el que emplearon características de la parte superior del cuerpo, como el área del triángulo entre las manos y el cuello, las distancias entre las articulaciones de la mano, el hombro y la cadera, y los ángulos en el cuello y la espalda. Sin embargo, su formulación no considera las características relacionadas con las articulaciones del pie, que pueden transmitir emociones al caminar (*Roether et al., 2009b*). Por ello *Randhavane et al. (2019)* incluyen áreas, distancias y ángulos de las articulaciones de los pies en la extracción de características. Las características posturales que calcularon son las siguientes:

- Volumen: la expansión del cuerpo transmite emociones positivas, mientras que una persona tiene una postura más compacta durante las expresiones negativas, por ello, los autores decidieron delimitar el cuerpo humano utilizando un cubo alrededor del mismo (ver Figura 4).

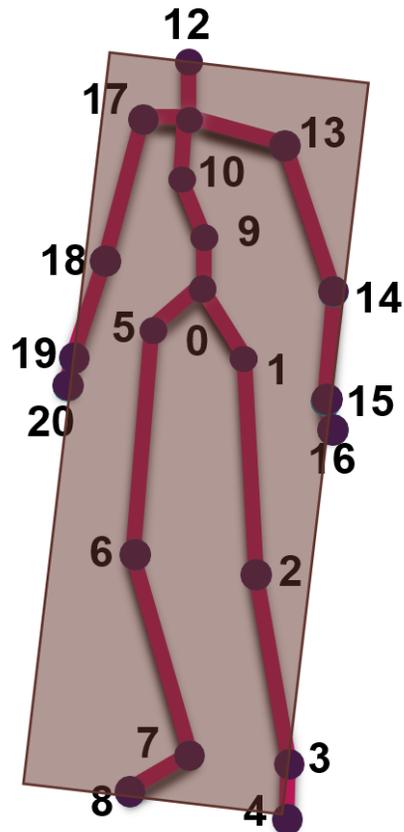


Figura 4. Representación del cálculo del volumen.

- Área: también modelaron la expansión del cuerpo mediante áreas de triángulos entre las manos y la cabeza y entre los pies y la articulación de la tronco (ver Figura 5).

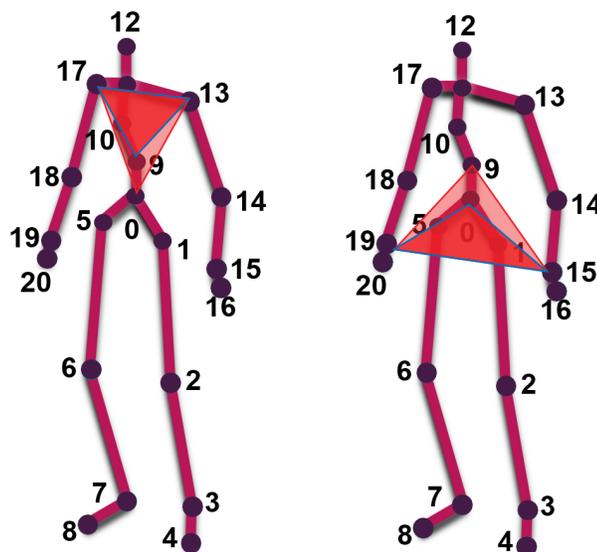


Figura 5. Representación del cálculo de áreas.

- Distancia: las distancias entre los pies y las manos también se pueden utilizar para modelar la expansión del cuerpo (ver Figura 6).

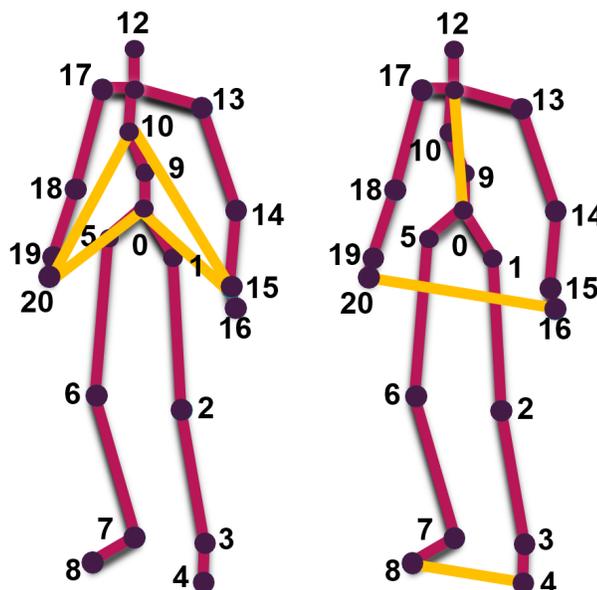


Figura 6. Representación del cálculo de distancias.

- Ángulo: la inclinación de la cabeza se emplea para distinguir entre emociones felices y tristes. También modelaron ángulos de otras articulaciones como el cuello (ver Figura 7).

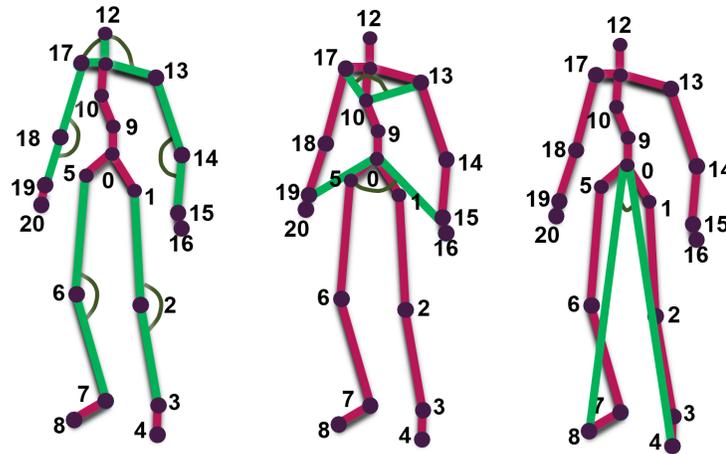


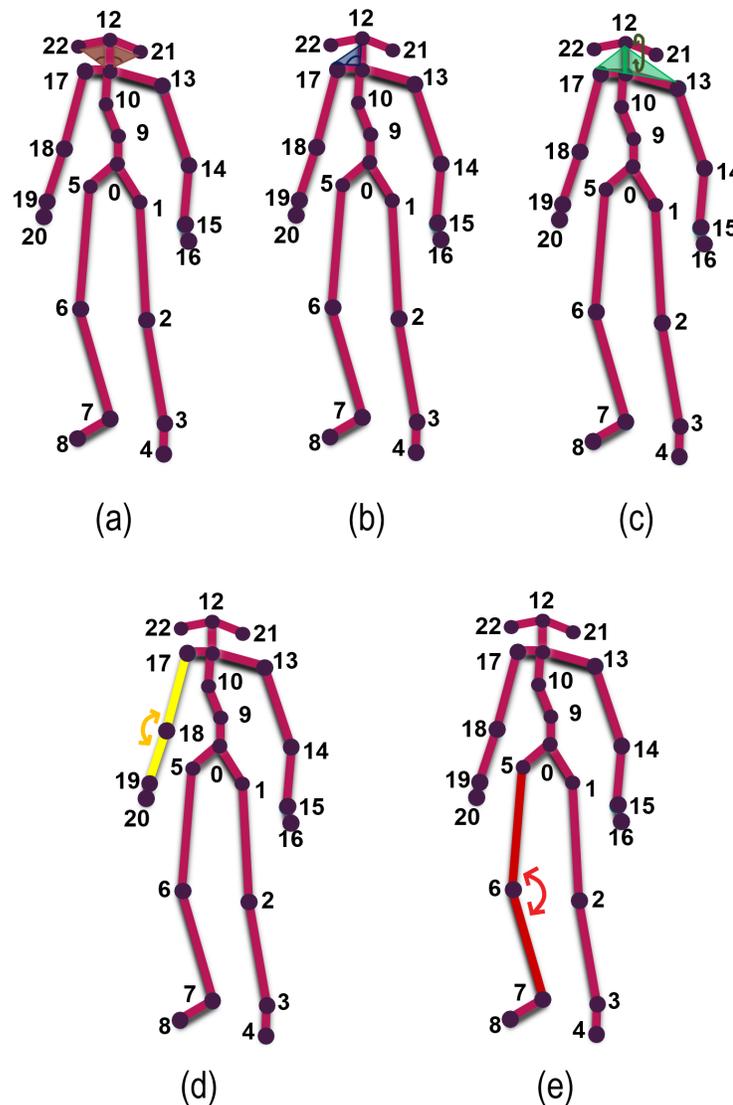
Figura 7. Representación del cálculo de ángulos.

Otro método interesante para reconocer las emociones utilizando la marcha humana fue propuesto por *Venture et al. (2014)* en el que calcularon la cinemática inversa y extrajeron 34 ángulos, velocidades y aceleraciones articulares, así como las 6 coordenadas generalizadas de la parte inferior del torso. Siguiendo recomendaciones de la Sociedad Internacional de Biomecánica usaron un modelo de 34 grados de libertad (DOF) formado por 15 segmentos del esqueleto humano como se muestra en la Tabla 3, donde se puede observar que cuentan con tres tipos distintos de articulaciones: libres (son articulaciones que se pueden mover libremente y pueden girar sobre cualquier eje), esféricas (mayor libertad de movimiento) y giratorias (permiten movimientos giratorios limitados).

Tabla 3. Características extraídas de acuerdo a los grados de libertad.

Nombre de la articulación	Tipo de articulación	Número de DOF
Tronco (torso bajo)	Libre	6
Cuello (cabeza)	Esférico	3
Cintura (torso superior)	Esférico	3
Hombro derecho (brazo superior)	Esférico	3
Codo derecho (brazo)	Giratorio	1
Muñeca derecha (mano)	Esférico	3
Hombro izquierdo (brazo superior)	Esférico	3
Codo izquierdo (brazo)	Giratorio	1
Muñeca izquierda (mano)	Esférico	3
Cadera derecha (muslo)	Esférico	3
Rodilla derecha (pierna)	Giratorio	1
Tobillo derecho (pie)	Esférico	3
Cadera izquierda (muslo)	Esférico	3
Rodilla izquierda (pierna)	Giratorio	1
Tobillo izquierdo (pie)	Esférico	3

Existen algoritmos que se enfocan principalmente en reconocer las emociones extrayendo parámetros espacio-temporales. *Srivastava et al. (2021)* se enfocan en utilizar una escala espacio-temporal para extraer información relevante que proporciona la red profunda *OpenPose* (Cao et al., 2017). Una particularidad de este trabajo, es que se identificó que las características evaluadas también dependían en gran medida de las extremidades superiores del cuerpo, como la caída de los hombros y el cuello, así como el movimiento de las manos, éstos fueron factores evidentes para clasificar las emociones. Los ángulos se tomaron en forma de seno o tangente como se especifica en la Figura 8.



**Figura 8.** Representación del cálculo de las características para reconocer emociones. De izquierda a derecha: (a) El ejemplo para calcular el seno del ángulo del cuello, a partir de la diferencia de las coordenadas del cuello y orejas (triángulos cafés). (b) Para calcular el seno del ángulo del hombro derecho se cálculo con diferencia del hombro derecho con respecto al cuello (triángulos azules). (c) En el cálculo para la inclinación de la cabeza se obtuvo con las distancias de los hombros dividido sobre los datos del cuello (triángulos verdes). (d) Para calcular el ángulo del codo derecho se obtiene restando  $m1 - m2$ , donde  $m1$  es la diferencia del hombro derecho y el codo derecho y  $m2$  representa la resta del codo derecho menos la muñeca (ángulos amarillos). (e) El cálculo del ángulo derecho de la rodilla se obtiene restando  $m3 - m4$ , donde  $m3$  es la diferencia de la cadera derecha y de la rodilla derecha y  $m4$  es la resta de la rodilla derecha menos el tobillo derecho (ángulos rojos).

### 3.3. Tipos de datos

En el contexto del análisis de la marcha, dependiendo del sensor que se emplee, se asocia el tipo de dato a considerar. Se clasifican dos tipos de datos, el primero corresponde a los datos 2D o bidimensionales los cuales se obtiene principalmente de vídeos, imágenes o sensores con mayor facilidad (Liu et al., 2020; Ma et al., 2006). A diferencia del segundo tipos de datos que son 3D o tridimensionales, para esta representación se requiere una construcción de la marcha ya sea en imagen o vídeo y se necesita diferentes posiciones de toma, se requiere emplear sensores más sofisticados. Otros autores optan por utilizar los datos 2D, aplicar un preprocesamiento y agregarles un eje adicional (Narang et al., 2017; Ionescu et al., 2013).

Una vez que se abordó la parte de los tipos de datos se pasa a tratar otro aspecto importante, que son los algoritmos que se emplean para clasificar emociones a partir de los datos de la marcha humana. La clasificación de algoritmos se divide en dos partes, primero están los algoritmos de aprendizaje de máquina y luego los algoritmos de aprendizaje profundo.

### 3.4. Algoritmos de aprendizaje de máquina

En *Li et al. (2016)*, se reconocen emociones a partir de la marcha humana, usando tecnologías portátiles y de menor costo, específicamente, se emplearon dos cámaras Kinect para la captura de datos que registraron los modos de marcha de 59 participantes en estado de ánimo neutral, de enojo y felicidad. Las características de los patrones de marcha se extrajeron de las coordenadas 3D de 14 articulaciones mediante la transformación de Fourier y el análisis de componentes principales (PCA). Se evaluaron los clasificadores *Naive Bayes*, *Random Forest*, *Support Vector Machines (SVM)* y *Sequential Minimal Optimization*, con una validación cruzada de 10 pliegues. La precisión al reconocer el enojo y la felicidad respecto al estado neutral alcanzó el 80.5 % y el 75.4 %, respectivamente.

En el trabajo de *Chiu et al. (2018)*, los autores identificaron características de movimiento en función de la estimación de poses en vídeo. Se capturaron datos de cinco emociones con 11 participantes en un entorno controlado. Se entrenaron varios modelos de aprendizaje de máquina, incluidos *SVM*, *Multilayer Perceptron*, *Naive Bayes*, *Decision Tree*, *Random Forest* y *Logistic Regression*, aplicando técnicas de selección de características y comparando su rendimiento. El mejor modelo entrenado para clasificar las cinco emociones tuvo una precisión del 64 %.

En *Ahmed et al. (2019)*, se presentó un marco de selección de características geométricas y cinemáticas

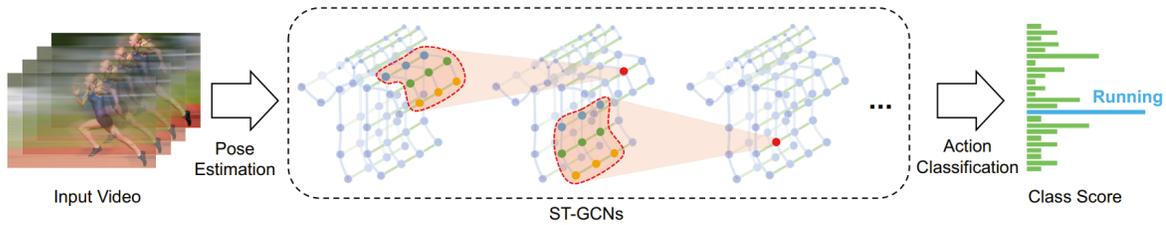
de dos capas para reconocer cuatro emociones: felicidad, tristeza, miedo e ira y el estado neutral. Algunas de estas características se introdujeron previamente para la síntesis, clasificación e indexación de movimiento 3D. En la primera capa, las características irrelevantes fueron eliminadas mediante una combinación de ANOVA y MANOVA. ANOVA fue utilizada para ordenar las características según su relevancia en el reconocimiento de emociones. MANOVA se utilizó para calcular la importancia del grupo y distribuir características entre varios grupos. Se pudo observar que la primera capa no fue suficiente para lograr un rendimiento óptimo del modelo en función de las características relevantes que fueron calculadas. El objetivo de la segunda capa fue encontrar el mejor subconjunto de características que maximiza la tasa de reconocimiento de emociones de los modelos expertos. El sistema fue validado en conjuntos de datos públicos, que contenían datos de marcha de 30 participantes. El sistema propuesto logró una precisión de reconocimiento del 90 %.

### **3.5. Algoritmos de aprendizaje profundo**

Recientemente, Bhattacharya y colaboradores, presentaron varios conjuntos de datos de marcha para el entrenamiento de cinco modelos basados en arquitecturas de aprendizaje profundo. Los modelos fueron capaces de reconocer tres emociones: feliz, triste y enojado y, además, agregaron el estado neutral. Este trabajo se basó principalmente en seleccionar una combinación de características afectivas y características obtenidas como resultado de una arquitectura de aprendizaje profundo entrenada a las cuales llamaremos características profundas.

#### **Red convolucional de grafos espacio-temporales (STGCN)**

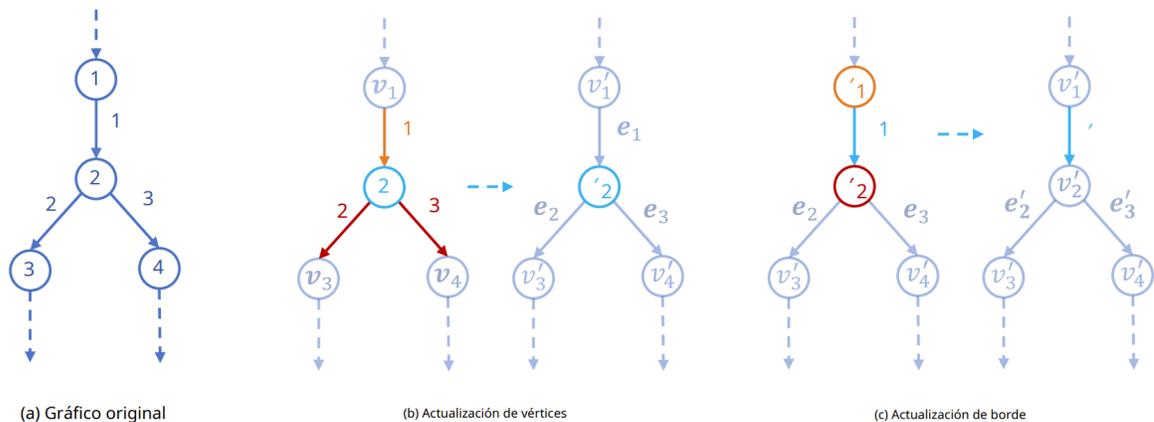
Primero se realizó una estimación de pose en vídeos, por lo general, los datos son una secuencia de *frames*, cada *frame* tendrá un conjunto de coordenadas conjuntas. Dadas las secuencias de las articulaciones del cuerpo en forma de coordenadas 3D, construyeron un grafo espacio temporal con las articulaciones como nodos del grafo y las conectividades naturales tanto en las estructuras del cuerpo humano como en el tiempo como bordes del grafo. La entrada al STGCN son los vectores de coordenadas conjuntas en los nodos del grafo. Luego se aplican múltiples capas de convolución en los datos de entrada y se generan mapas de características de nivel superior en el grafo. Luego, el clasificador estándar de Softmax lo clasifica en la clase correspondiente, ver Figura 9 (Yan et al., 2018). Esta arquitectura tuvo una precisión promedio de todas las clases (mAP) de 0.61.



**Figura 9.** Red convolucional de grafos espacio-temporales. Primero realiza una estimación de pose en vídeos y se construye un grafo espacio temporal en secuencias de esqueleto. Después se aplican múltiples capas de convolución al grafo espacio temporal (STGCN) y se generarán gradualmente mapas de características de nivel superior en el gráfico. Luego, el clasificador estándar de Softmax lo clasificará en la clase correspondiente. Tomada de (Yan et al., 2018).

### Red neuronal de grafos dirigidos (DGNN)

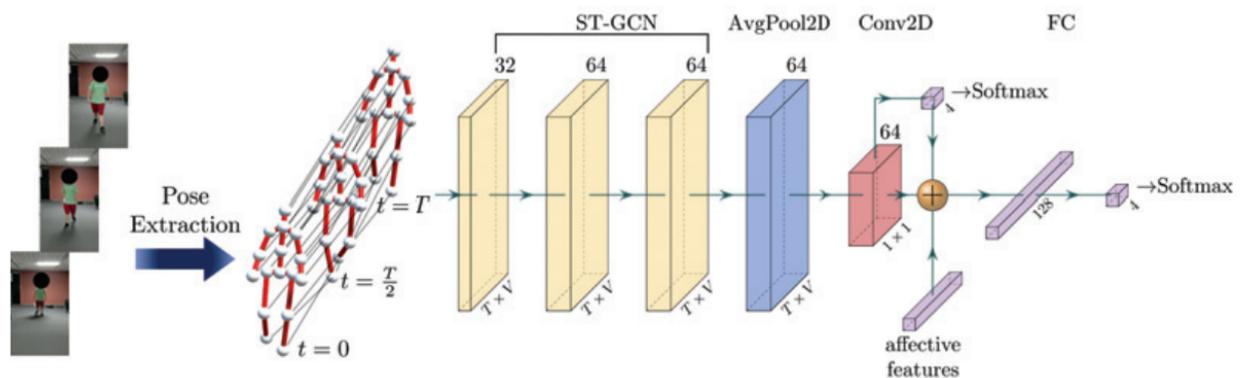
Los datos del esqueleto sin procesar son una secuencia de *frames*, cada uno contiene un conjunto de coordenadas conjuntas. Dada una secuencia de esqueletos, primero se extrae la información de acuerdo con las coordenadas 3D de las articulaciones. Luego, las articulaciones y los segmentos (la información espacial) en cada *frame* se representan como los vértices y los bordes dentro de un grafo acíclico dirigido, que alimenta a la red neuronal (DGNN) para extraer características para el reconocimiento de emociones. Finalmente, la información de movimiento, que se representa con la misma estructura de grafo que se utilizó para la información espacial, se extrae y se combina con la información espacial en un *frame* de dos flujos para mejorar aún más el rendimiento, ver la Figura 10 (Shi et al., 2019). Dicha arquitectura obtuvo un mAP de 0.74.



**Figura 10.** Red neuronal de grafos dirigidos. (a) Es el grafo original. (b) Muestra el procedimiento de actualización de vértices, donde el atributo del vértice en ( $v_2$ ) y los atributos de su borde entrante ( $e_1$ ) y bordes salientes ( $e_2$  y  $e_3$ ) se combinan para obtener un vértice actualizado ( $v'_2$ ). (c) Muestra el procedimiento de actualización de bordes, donde el atributo de la propia arista ( $e_1$ ) y los atributos de su vértice de origen ( $v_1$ ) y vértice de destino ( $v'_2$ ) se combinan para obtener una arista actualizada ( $e'_1$ ). El círculo azul representa el borde (o vértice) que se está actualizando. El círculo naranja y el círculo rojo representan el vértice de origen (o borde de entrada) y el vértice de destino (o borde de salida) que están involucrados en la actualización. Tomada de (Shi et al., 2019).

## Grafo espacial-temporal (STEP)

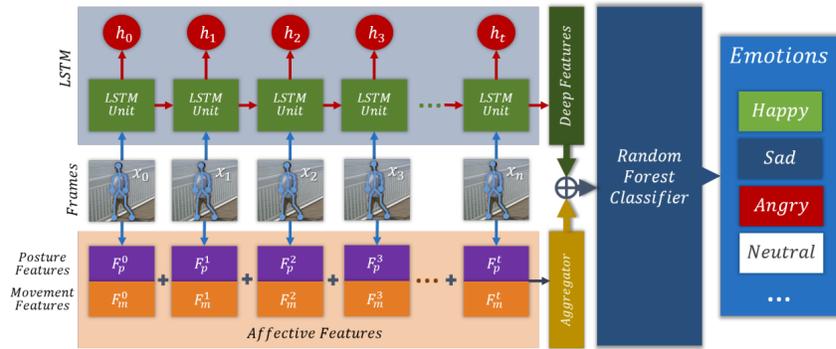
La arquitectura *STEP* para clasificar emociones humanas a partir de la marcha, está basada en una arquitectura de red convolucional de grafo espacio temporal (STGCN). Dicha arquitectura de red clasifica tres emociones feliz, triste, enojado y el estado neutral. La arquitectura necesita como entrada vídeos de marcha humana en la que, a dichos vídeos, se les hace una extracción de articulaciones que da como resultado final poses de esqueletos humanos (series de tiempo). Las series de tiempo obtenidas se proporcionan como entrada a las tres capas STGCN, la salida de la última capa STGCN se agrupa en una capa de *average pool*, el resultado de la capa *average pool* se transmite a través de una capa *Conv2D* en donde su salida pasa a través de la capa de dimensión *fully connected (FC)*, correspondiente a las 4 etiquetas de emoción que se tiene, seguido de una operación *softmax* para generar las etiquetas de clase, ver Figura 11 (Bhattacharya et al., 2020a). La arquitectura STEP logró un mAP de 0.77.



**Figura 11.** Arquitectura STEP: clasificador para predecir emociones percibidas a partir de marcha humana. Tomada de Bhattacharya et al. (2020a).

## Redes de memoria prolongada de corto plazo (LSTM)

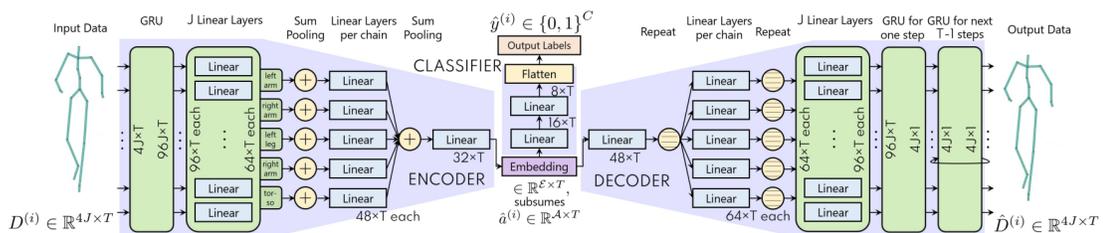
Al algoritmo *LSTM Network* se le introducen las poses 3D para su entrenamiento utilizando múltiples conjuntos de datos de marcha. Calculan características afectivas que responde a postura y movimiento, utilizando caracterización psicológica. Estas se dividen en afectivas y profundas. Se procede a clasificar la combinación de las mismas en las cuatro etiquetas mencionadas anteriormente, usando el clasificador random forest, ver Figura 12 (Randhavane et al., 2019).



**Figura 12.** Arquitectura empleando LSTM para la clasificación de emociones a partir de la marcha. Tomada de (Randhavane et al., 2019).

### Take an Emotion Walk (TAEW)

Y por último la arquitectura *TAEW* se centra en un algoritmo autoencoder por atención, en el que la función principal del codificador es introducir las poses a una unidad recurrente cerrada (GRU) de dos capas para obtener representaciones de entidades para rotaciones en todas las articulaciones en todos los instantes de tiempo. Las salidas se pasan por unidades lineales individuales, en donde la salida de las unidades lineales se agrupan en diferentes partes del cuerpo, por ejemplo, una capa para el brazo derecho, una para el brazo izquierdo, pierna derecha, pierna izquierda y una más para el tronco del esqueleto. Por lo tanto, cada una de estas cinco capas lineales aprende a centrar la atención en una parte diferente del cuerpo humano. Luego, se agruparon las salidas de estas cinco capas lineales en otra capa lineal que, por construcción, centra la atención en los movimientos de todo el cuerpo. Finalmente, el decodificador hace lo mismo que el codificador pero de manera inversa (Bhattacharya et al., 2020b). El resultado de la última capa da como resultado la predicción de una etiqueta que representa una emoción, ver la Figura 13 (Bhattacharya et al., 2020b). Este modelo logró mejor rendimiento para clasificar las emociones con una precisión de 84 %.



**Figura 13.** Arquitectura para la clasificación semi-supervisada de emociones discretas percibidas a partir de la marcha. Tomada de (Bhattacharya et al., 2020b).

## Capítulo 4. Metodología

En este capítulo se abordan las técnicas utilizadas para el análisis de los datos de marcha humana. En primer lugar se describe la recopilación y preparación de datos para varios datasets de marchas humanas. Después se muestran los dos tipos de tratamiento para el análisis descriptivo. Se detallan los diferentes métodos de selección de características. Enseguida se describe la construcción de los modelos de clasificación a partir de los subconjuntos resultantes del análisis anterior. Finalmente, se describen los algoritmos de inferencia a utilizar para la construcción de los modelos de clasificación, cuyos resultados de evaluación se presentan en el siguiente capítulo. La metodología se muestra a bloques en la Fig. 14.

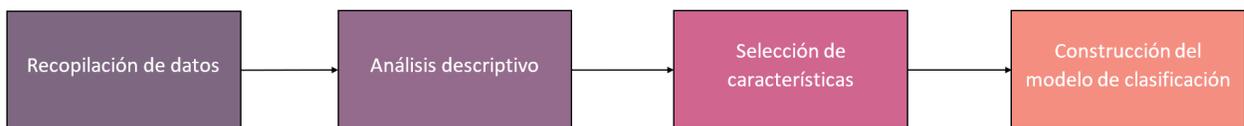


Figura 14. Metodología implementada.

### 4.1. Descripción de datos

Al unir los conjuntos de datos detallados en el Capítulo 2 Subsección 2.3, se obtuvieron 1856 segmentos de videos de marcha. Para los conjuntos de datos actuados, el etiquetado de emociones se realizó utilizando la estrategia de etiquetado de emociones percibidas, que consiste en seleccionar personas aleatorias que se les proporciona la información necesaria para que entienda las emociones empleadas y 10 videos para su etiquetado; posteriormente se realiza un promedio ponderado de sus respuestas (Randhavane et al., 2019). Se extrajeron puntos clave del esqueleto humano normalizado a 21 articulaciones; cada articulación está representada por coordenadas espaciales 3D, ver Fig. 15. Los datos de marcha están etiquetados con cuatro valores de clase ( $C = \text{feliz}, \text{triste}, \text{enojado}, \text{neutral}$ ), y la distribución de clases se muestra en la Fig. 16. Es importante señalar que existe un claro desbalance en los datos, donde la clase mayoritaria (feliz) cuenta con 1019 datos y la clase minoritaria (neutral) tiene 81 datos. Cada marcha comprende múltiples series de tiempo con la siguiente estructura:

$$\text{marcha} = 240 \text{ registros} * 21 \text{ articulaciones} * 3 \text{ ejes} \quad (1)$$

donde un video de una persona caminando (marcha) consta de 240 imágenes y al extraer el esqueleto de cada una, se obtiene un punto en 3D para cada articulación, lo que aumenta considerablemente la representación de las marchas.



Figura 15. Esqueleto representado por un conjunto de 21 articulaciones.

## 4.2. Análisis descriptivo

En la sección anterior se aprecia que se tiene una alta dimensionalidad de los datos, es decir, cada ejemplo de marcha comprende 63 series temporales con 240 puntos. Si bien se puede pensar que la marcha humana solo involucra el estudio del movimiento periódico de las extremidades inferiores, se debe enfatizar que el resto del cuerpo brinda estabilidad y equilibrio durante la locomoción (Shultz et al., 2015). Se propone un análisis descriptivo de los datos para determinar los mejores subconjuntos de articulaciones que aportan información relevante para discriminar las emociones estudiadas.

Se propusieron dos tratamientos para este análisis de datos. El primero consistió en analizar los datos originales y considerar las cuatro clases como se muestra en la Fig. 16. Debido al desequilibrio de datos dado el número de marchas por clase, el segundo tratamiento consistió en seleccionar un subconjunto de datos según el número de marchas de la clase minoritaria ( $c = neutral$ ), por par de emociones, para determinar qué articulaciones son más relevantes para separar cada par de clases, ver Fig. 17.

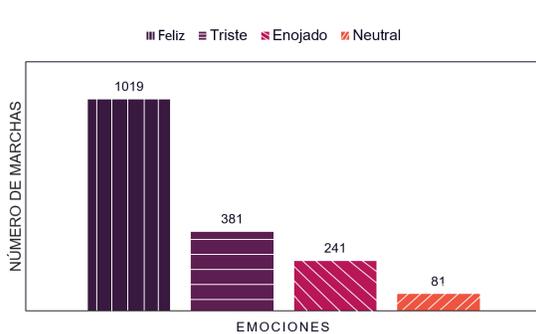


Figura 16. Distribución de clases para datos desbalanceados  $D$ .

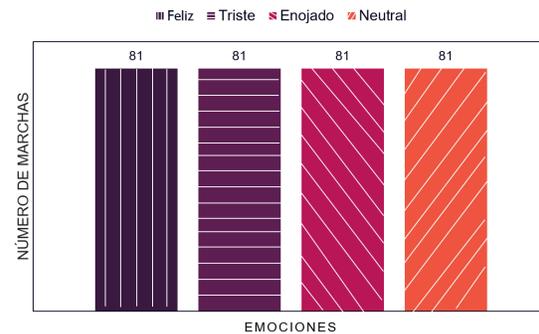
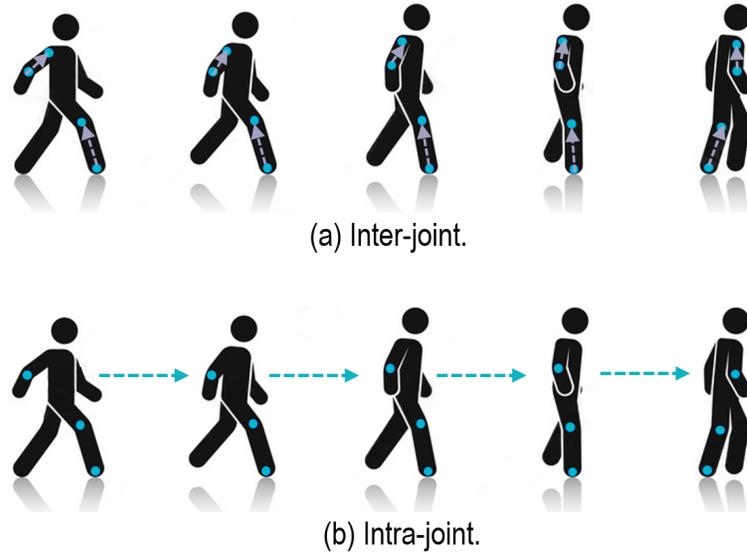


Figura 17. Datos de muestra balanceados agrupados por par de clases.

Se calculó la cantidad de movimiento de promedio relativo (Sapiński et al., 2019) para ambos tratamientos, con el objetivo de medir la cantidad de información que las articulaciones brindan respecto a las emociones. Esta métrica se aplica a series de tiempo, por lo que se puede aplicar a un par de articulaciones diferentes en el mismo *frame* (inter-joint), por ejemplo, el codo y hombro en el tiempo  $t$ , o a la misma articulación en *frames* contiguos (intra-joint), por ejemplo, el hombro en los tiempos  $t$  y  $t - 1$ . La diferencia entre ambos enfoques se ilustra en la Fig. 18a y 18b.



**Figura 18.** Enfoque para calcular la cantidad de movimiento promedio relativo. Inter-joint: diferentes articulaciones, mismo tiempo (18a). Intra-joint: misma articulación, tiempos diferentes (18b).

En el enfoque inter-joint se calcula el movimiento de una articulación  $P_j$  con respecto a su articulación predecesora  $P_{j-1}$  (enlace anterior) conectando todo el esqueleto según el orden mostrado en la Fig. 15, y considerando las conexiones particulares 0-5, 0-9, 11-13 y 11-17. Se aplicó la ecuación 2 para cada emoción en ambos conjuntos de datos.

Por el contrario, en el enfoque intra-joint se calcula el movimiento de cada articulación del esqueleto en el tiempo  $t$  con respecto a la misma articulación en el tiempo  $t - 1$ , de acuerdo a la ecuación 3.

$$inter_{je} = \sum_{ne=1}^{N_e} \frac{|P_{je}(f_{ne}) - P_{(j-1)e}(f_{ne})|}{F_{ne}N_e} \quad (2)$$

$$intra_{je} = \sum_{ne=1}^{N_e} \frac{|P_{je}(f_{ne}) - P_{je}(f_{ne-1})|}{F_{ne}N_e} \quad (3)$$

donde:  $j = 0, \dots, 21$  es el número de articulaciones en un eje;  $e$  es la emoción;  $N_e$  es el número de marchas por emoción  $e$ ;  $ne = 1, \dots, N_e$  es el índice de la emoción  $e$ ;  $F_{ne}$  es el número de *frames* (cuadros)  $n$  de emoción  $e$ ;  $f_{ne} = 2, \dots, F_{ne}$  es el índice de frame por marcha  $n$  de cada emoción  $e$  (excluyendo el primer frame);  $P_{je}(f_{ne})$  es la posición del joint  $j$  en el frame  $f_{ne}$ .

Finalmente y para facilitar la comparación, las señales resultantes en los tres ejes se combinan con diferentes operaciones: vector de magnitud (ecuación 4), promedio (ecuación 5), desviación estándar (ecuación 6), varianza (ecuación 7) y frecuencia (ecuación 8), lo que permitirá analizar la cantidad de movimiento por articulación.

$$mag = \sqrt{X^2 + Y^2 + Z^2} \quad (4)$$

$$promedio = \frac{\sum_{i=0}^N (X_i + Y_i + Z_i)}{N} \quad (5)$$

$$desviación\ estándar = \sqrt{\frac{\sum_{i=0}^N ((X_i - \bar{X}) + (Y_i - \bar{Y}) + (Z_i - \bar{Z}))^2}{N - 1}} \quad (6)$$

$$varianza = \frac{\sum_{i=0}^N ((X_i - \bar{X}) + (Y_i - \bar{Y}) + (Z_i - \bar{Z}))^2}{N - 1} \quad (7)$$

$$potencia = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N |X(i) + Y(i) + Z(i)|^2 \quad (8)$$

Con el fin de apreciar la variabilidad entre los valores de movimiento relativo de cada clase se hizo un análisis inter-emociones por tuplas calculando la diferencia entre emociones, el análisis se muestra en los mapas de calor de la Subsección 5.1.2

Los subconjuntos seleccionados de los datos con base en este análisis se presentan en el Capítulo 5.1.3 de resultados.

### 4.3. Extracción de características

El siguiente paso fue extraer un vector de características de la combinación de los diferentes ejes del sistema de coordenadas de posición, las 21 articulaciones que representan el esqueleto humano y un

conjunto de características que se estimó para cada señal de marcha según la ecuación 1.

Cada uno de los ejes proporciona información relevante durante la marcha, uno de ellos está asociado a la dirección frontal en la que se mueven las personas (eje X), el segundo corresponde a la elevación de las articulaciones con respecto al plano transversal (eje Z), y el tercero es el desplazamiento lateral en el plano sagital (eje Y). Por esta razón, las señales individuales se combinaron en pares y tripletas, para aprovechar la información complementaria entre ejes, tal como se define en la ecuación 9. El conjunto de articulaciones se muestra en la Fig. 15 y en forma de vector de acuerdo a la ecuación 10. Se utilizó un conjunto de 160 características de dominio temporal (18), estadístico (36) y espectral (106) (Barandas et al., 2020), donde temporal se refiere a características del dominio del tiempo, por ejemplo, la energía absoluta de la señal (*abs\_energy*); estadístico a características que se extraen de un análisis estadístico, por ejemplo, la autocorrelación de la señal (*autocorr*) y finalmente espectral que son características que convierten la señal basada en el tiempo al dominio de frecuencia, por ejemplo, la pendiente espectral (*spectral\_slope*).

$$\text{series de tiempo} = [ X, Y, Z, \text{mag}(XY), \text{mag}(XZ), \text{mag}(YZ), \text{mag}(XYZ) ] \quad (9)$$

$$\begin{aligned} & \{ 0 \text{ tronco, } 1 \text{ cadera izquierda, } 2 \text{ rodilla izquierda, } 3 \text{ tobillo izquierdo,} \\ & 4 \text{ pie izquierdo, } 5 \text{ cadera derecha, } 6 \text{ rodilla derecha, } 7 \text{ tobillo derecho,} \\ \text{articulaciones} = & 8 \text{ pie derecho, } 9 \text{ espalda baja, } 10 \text{ espina, } 11 \text{ cuello, } 12 \text{ cabeza, } 13 \text{ hombro} \\ & \text{izquierdo, } 14 \text{ codo izquierdo, } 15 \text{ muñeca izquierda, } 16 \text{ mano izquierda,} \\ & 17 \text{ hombro derecho, } 18 \text{ codo derecho, } 19 \text{ muñeca derecha, } 20 \text{ mano derecha } \} \\ & (10) \end{aligned}$$

En el Apéndice A se muestran de manera detallada las diferentes características de cada grupo (temporal, estadístico y espectral). Una de las peculiaridades es que estos grupos están incorporados dentro de la librería TSFEL que se emplea para datos de series de tiempos, y la estructura que tienen los datos se adaptan para implementar dicha librería, además en la literatura encontrada trabajan principalmente con ángulos, distancias, áreas, grados de libertad, pero no extraen características estadísticas sobre la misma marcha. Algunos autores trabajan sobre los mismos ángulos y calculan promedios pero ninguno los utiliza del mismo modo en que se propone en este trabajo.

Combinando los tres grupos de características se obtienen un total de 23,373 características más una

variable de clase, esto se puede apreciar en las ecuaciones 11 y 12 respectivamente.

$$\text{características finales} = [ 7 (\text{series de tiempo}) * 21 (\text{articulaciones}) * 160 (\text{características}) ] \quad (11)$$

$$\text{clase} = \{ \text{feliz}, \text{triste}, \text{enojado}, \text{neutral} \} \quad (12)$$

Finalmente se calcularon siete vectores de características finales de acuerdo con el análisis descriptivo detallado en la Subsección anterior. Estos se representan en el *vector final* = [*características finales*, *clase*].

#### 4.4. Selección de características

El siguiente paso fue seleccionar de manera automática características del vector final, empleando tres métodos diferentes de selección (ver subsección 2.4.1) como se muestra en el vector de la ecuación 13.

$$\text{seleccion}_{\text{metodo}} = (\text{univariado}, \text{impureza de disminucion media}, \text{componentes principales}) \quad (13)$$

Dando como resultado de la selección de características la ecuación 14

$$\text{características seleccionadas} = \text{seleccion}_{\text{metodo}}(\text{vector final}) \quad (14)$$

#### 4.5. Clasificación de emociones

Los clasificadores se construyeron a partir de los vectores de características finales utilizando tres algoritmos de inferencia de aprendizaje de máquina tradicionales y algoritmos de aprendizaje profundo basados en redes neuronales convolucionales (CNN's), a continuación, se explican los parámetros propuestos de cada algoritmo.

- **k-Nearest Neighbors (kNN)**

A pesar que este algoritmo es no-paramétrico, para encontrar la mejor k en este algoritmo se empleo la técnica del codo, lo cual, nos permitió definir el número de vecinos en ocho y la métrica de distancia que se seleccionó es Minkowski debido a que mostró mejor desempeño.

### ■ Random forest (RF)

Los parámetros que se emplearon fueron propuestos después de varias ejecuciones y que derivaron en parámetros óptimos para el algoritmo *random forest*, para lo cual nos permitió seleccionar el número de estimadores en cien, el criterio seleccionado *gini*, se agregó una restricción con respecto a la profundidad del árbol y se configuró para que los árboles recibieran todas las características dependiendo del método de selección que se eligió (por omisión este parámetro acepta solo el 50 % de las características que le proporcionamos) esta decisión se tomó para que todos los algoritmos estén bajo las mismas condiciones, es decir, que tengan solo las características proporcionadas.

### ■ Boosting

Para este meta-clasificador se utilizó el algoritmo *Support Vector Machines*, con un número de estimadores de cien dando como resultado un mejor rendimiento.

### ■ CNN

En las diferentes variantes propuestas de CNN's se buscaron los parámetros óptimos que dieran como resultado un mejor rendimiento. Para ello se seleccionó un *batch size* de 64, el número de épocas fue 500, una pérdida de *categorical crossentropy* y como optimizador el método *Adam*.

- CNN-1.

La red convolucional cuenta con 9 capas, tres capas convolucionales, dos capas *Max Pooling*, una capa *dropout*, una capa *flatten* y dos capas densas (ver Fig. 19a).

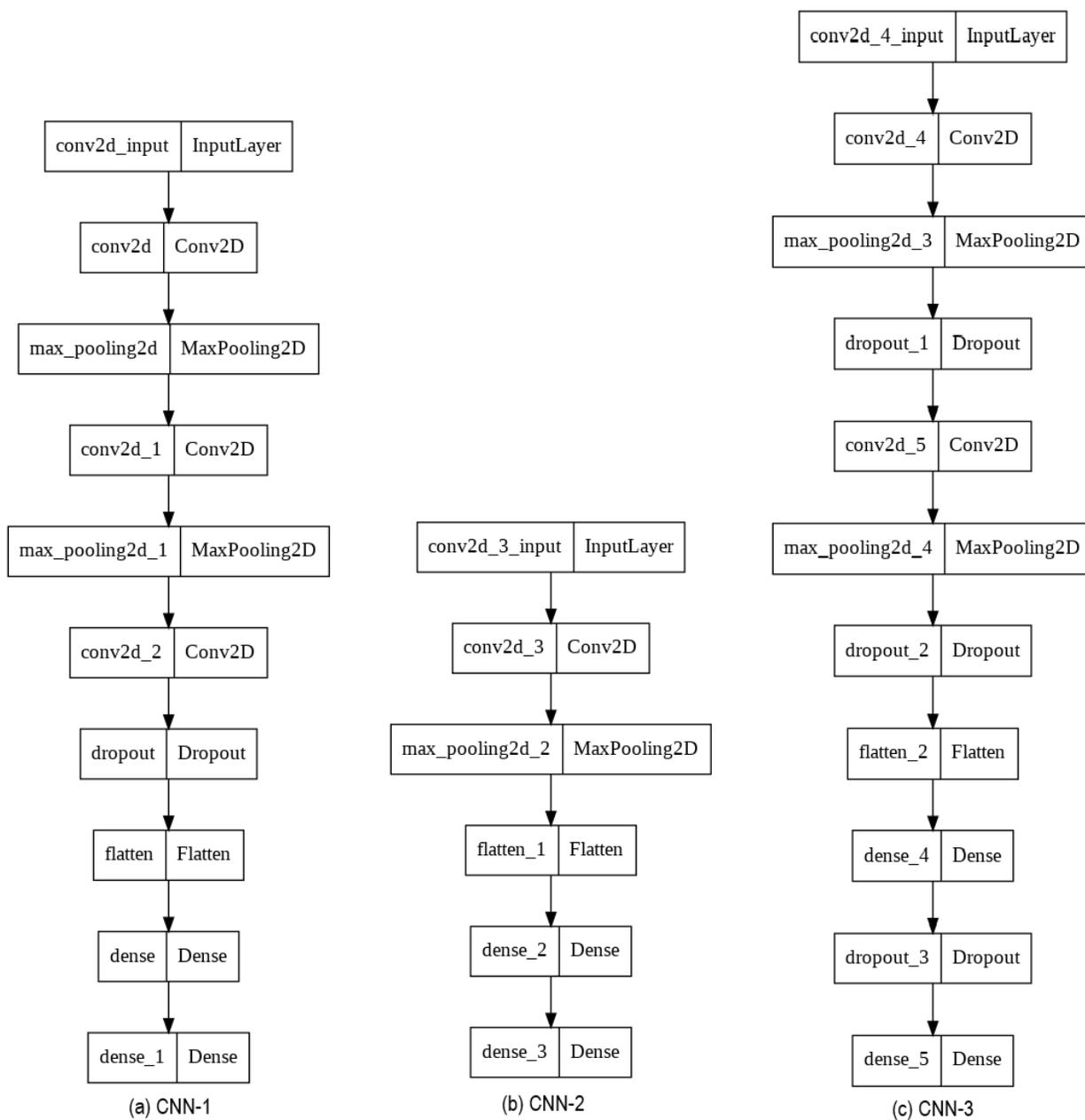
- CNN-2.

La red convolucional cuenta con 5 capas, una capa convolucional, una capa *Max Pooling*, una capa *flatten* y dos capas densas (ver Fig. 19b).

- CNN-3.

La red convolucional 3 cuenta con 10 capas, dos capas convolucionales, dos capas *Max Pooling*, una capa *flatten*, tres capas *dropout* y dos capas densas (ver Fig. 19c).

Los resultados de los modelos propuestos se presentan en el Capítulo 5 Subsección 5.2.



**Figura 19.** Arquitectura de las redes profundas basadas en CNN'S.

## Capítulo 5. Resultados

---

En este capítulo se detallan los experimentos y resultados obtenidos a lo largo de la investigación. Primero se muestran los resultados del análisis descriptivo de los datos (ver Subsección 4.2), se enfoca en el análisis relativo de movimiento y en el análisis inter-emociones, los cuales nos permitieron seleccionar diversos subconjuntos de articulaciones con base en su rango de movilidad y su correlación. Después se reportan los resultados obtenidos con los modelos basados en los algoritmos **kNN**, **random forest** y **boosting** (ver Subsección 4.5). Con los tratamientos obtenidos en el análisis descriptivo se comparan los resultados con el estado del arte y respecto a los resultados de la replicación de los mismos.

### 5.1. Análisis descriptivo

A continuación se muestran los resultados que se obtuvieron en el análisis descriptivo de los datos mencionados en la Subsección 4.2.

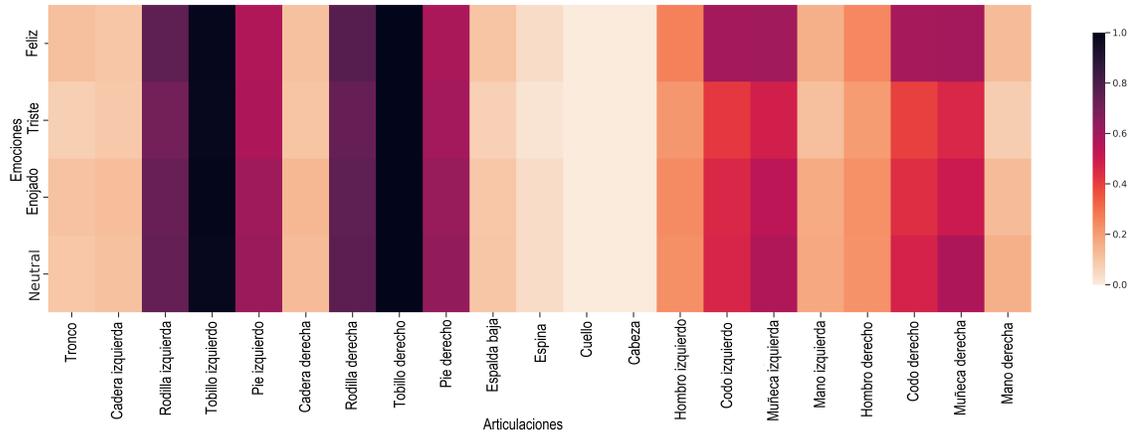
#### 5.1.1. Análisis relativo de movimiento

El análisis relativo de movimiento se centra principalmente en el comportamiento que presentan las articulaciones con respecto al rango de movimiento y la observación de los diferentes patrones que tiene cada una de las emociones, tanto para el enfoque inter-joint (Ecuación 2) como intra-joint (Ecuación 3) con el objetivo de encontrar subconjuntos de articulaciones que mejoren la clasificación en comparación con el uso de la representación completa de 21 articulaciones.

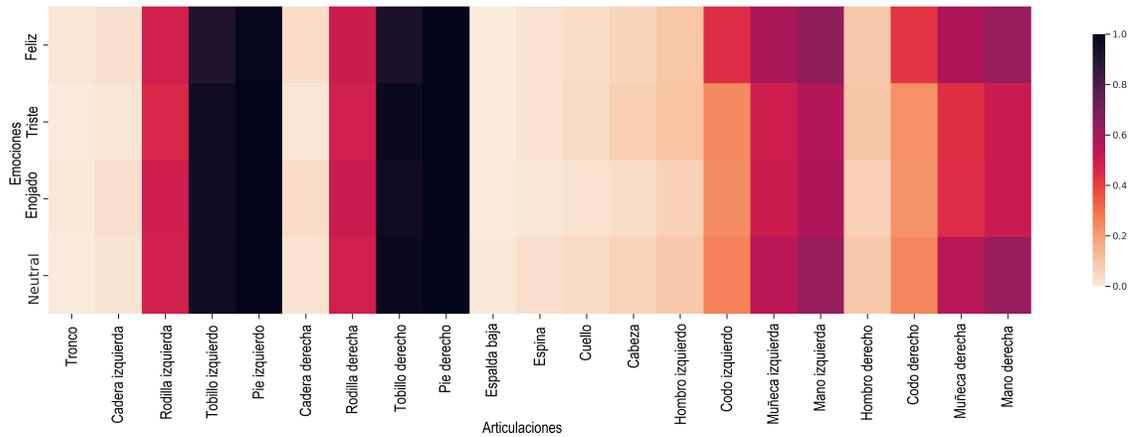
La cantidad de movimiento promedio relativo se calculó para los datos no balanceados, como se ilustra en las Figuras 20 y 21, para todas las articulaciones considerando las emociones feliz, triste, enojado y el estado neutral para los enfoques inter-joint e intra-joint, respectivamente. Utilizando la métrica reportada en la Ecuación 4.

En general, los valores más altos del movimiento relativo se registraron en las articulaciones de la parte inferior; destacando los alcanzados en los tobillos del análisis inter-joint y los tobillos y pies para el análisis intra-joint. En cuanto a las emociones, feliz presenta valores altos en comparación a las demás emociones con respecto a las mismas articulaciones para ambos tratamientos; cuando se compara el resto de las emociones se puede apreciar que enojado y el estado neutral exhiben el mismo patrón de rango de movimiento en todas las articulaciones en el análisis inter-joint y a diferencia del análisis intra-joint las emociones que pueden presentar confusión entre métricas son triste y enojado debido a que

muestran el mismo comportamiento en el rango de movimiento. Los resultados obtenidos en el rango de movimiento relativo para ambos tratamientos nos brinda información distinta entre emociones, por lo cual se realizó un análisis inter-emociones para analizar a profundidad cada emoción mediante pares de emociones.



**Figura 20.** Enfoque inter-joint: Los mapas de calor representan la cantidad de movimiento de todas las articulaciones.

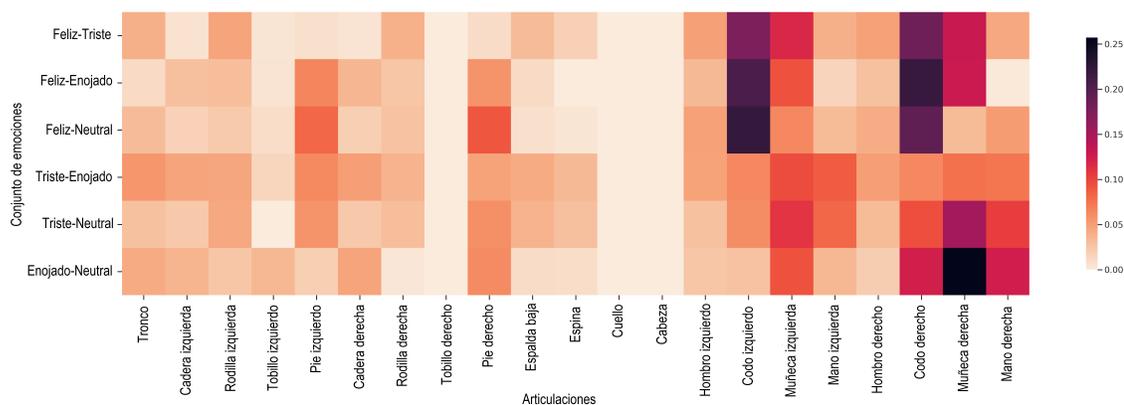


**Figura 21.** Enfoque intra-joint: Los mapas de calor representan la cantidad de movimiento de todas las articulaciones.

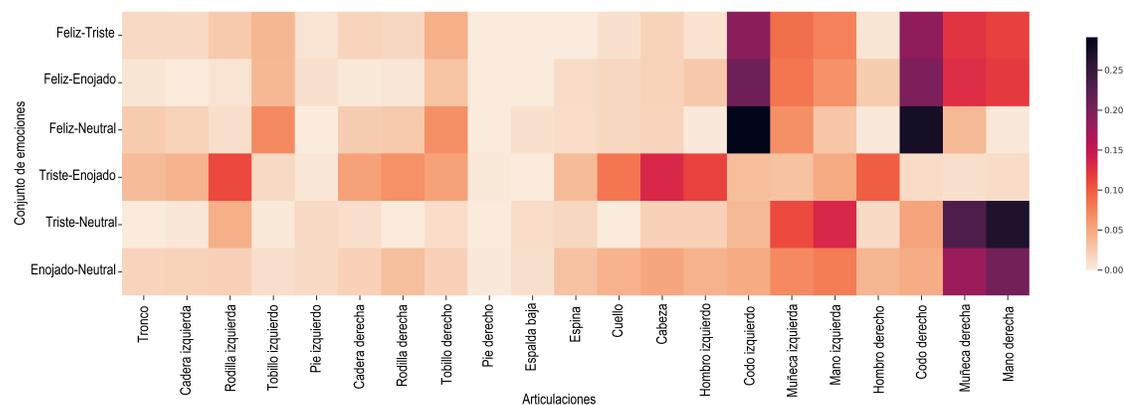
### 5.1.2. Análisis inter-emociones

En este segundo análisis inter-emociones se utilizó una muestra de los datos originales para tener una comparación justa, ver Figura 17. En las Figuras 22 y 23 se ilustra la diferencia de cantidad de movimiento para todas las articulaciones considerando pares de emociones para los enfoques inter-joint e intra-joint. En resumen, los valores más altos se registraron en las articulaciones de las extremidades superiores (codos, muñecas y manos) para ambos enfoques. La articulación de la extremidad inferior con mayor diferencia del movimiento son los pies para el análisis inter-joint y las rodillas y tobillos en el análisis

intra-joint, pero con menores valores respecto a las articulaciones de las extremidades superiores. En cuanto a los pares de emociones, las combinaciones que contienen las emociones feliz y neutral con otras emociones presentan valores altos con respecto a las articulaciones mencionadas con anterioridad; para este análisis sólo se presenta un patrón diferente cuando se comparan las clases triste y enojado del análisis intra-joint.



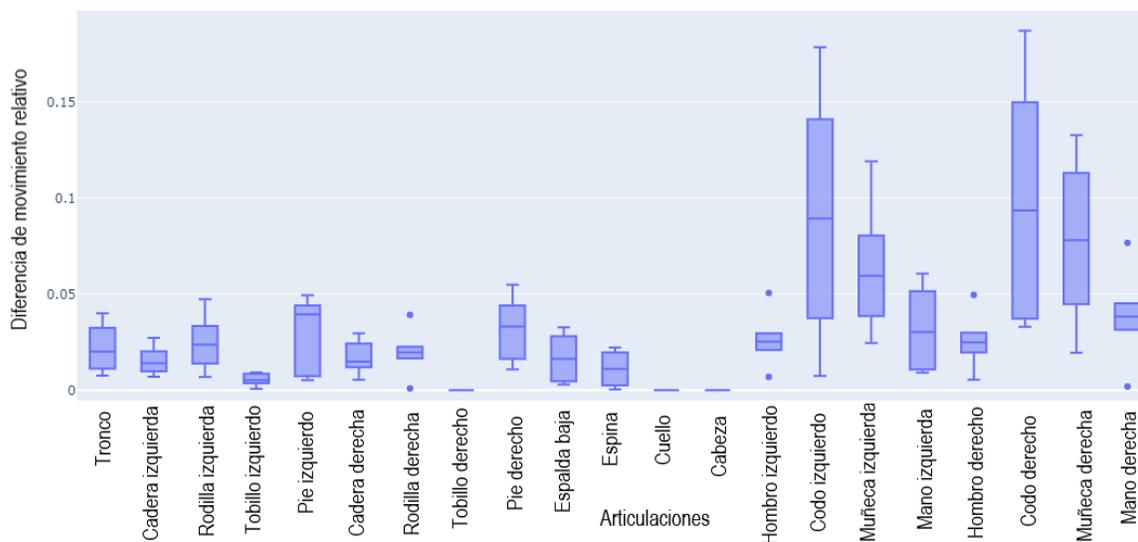
**Figura 22.** Enfoque inter-joint: Los mapas de calor representan la diferencia de movimiento de todas las articulaciones por pares de emociones.



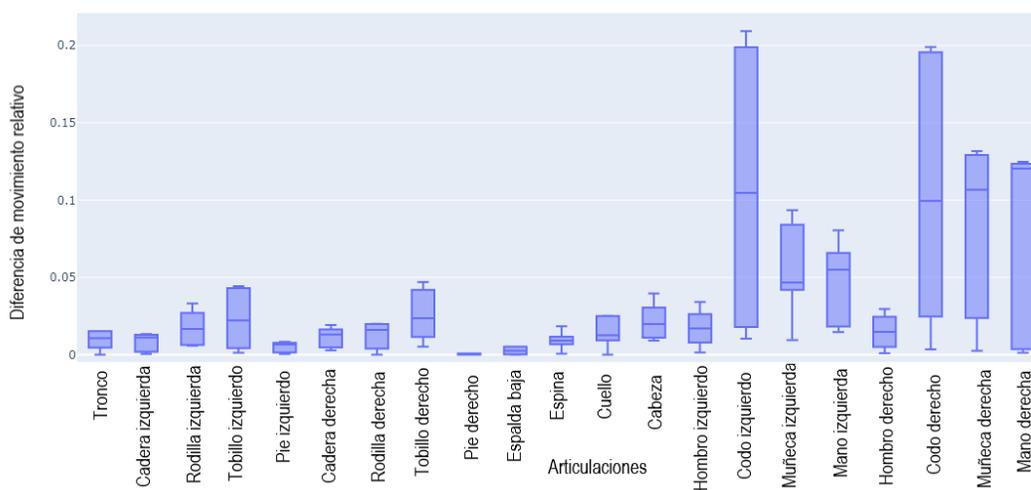
**Figura 23.** Enfoque intra-joint: Los mapas de calor representan la diferencia de movimiento de todas las articulaciones por pares de emociones.

Las gráficas presentadas en las Figuras 24 y 25 muestran los valores numéricos correspondiente al análisis-inter emociones de las Figuras 22 y 23. Estas nos permiten observar cómo se alternan las extremidades del cuerpo al caminar, es decir, extremidades opuestas avanzan de forma conjunta, por ejemplo extremidad inferior izquierda y extremidad superior derecha. Se puede apreciar que las extremidades izquierdas inferiores presentan mayor dispersión que las derechas, por lo cual si las extremidades inferiores izquierdas tienen mayor dispersión, las extremidades superiores derechas deben de presentar menor dispersión. Al

igual que en los mapas de calor se observan valores altos en las articulaciones de los codos y muñecas para el enfoque inter-joint y los codos, muñecas y manos en el enfoque intra-joint, lo que indica que hubo un mayor desplazamiento en ciertas articulaciones respecto al tiempo anterior que respecto a su articulación predecesora, por ejemplo la mano registro un mayor movimiento relativo en el análisis intra-joint. Estos valores son esperados ya que es proporcional al arco, además mientras las articulaciones de esas extremidades estén mas lejos del o tronco los valores son más altos.



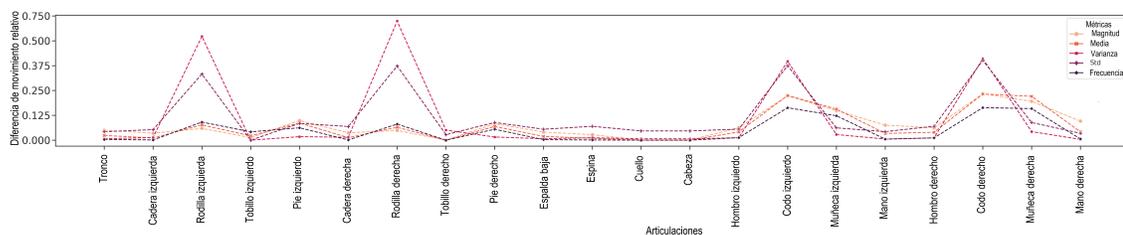
**Figura 24.** Enfoque inter-joint: Las gráficas de caja y pivote representan la diferencia de movimiento de todas las articulaciones por pares de emociones.



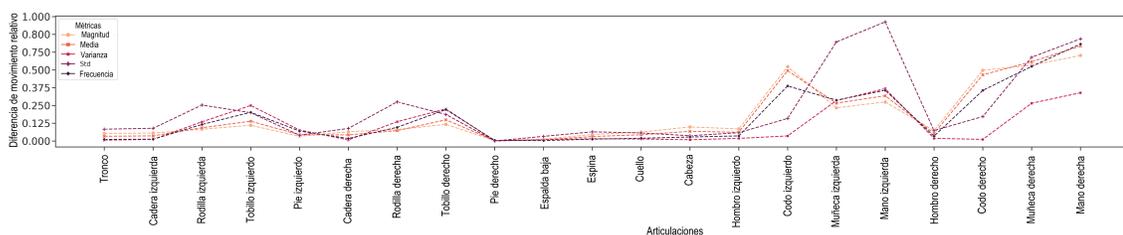
**Figura 25.** Enfoque intra-joint: Las gráficas de caja y pivote representan la diferencia de movimiento de todas las articulaciones por pares de emociones.

Las Figuras 26 y 27 muestran los resultados anteriores promediando los valores para cada articulación

para todas las métricas empleadas, recordando que estas métricas se utilizaron para combinar las señales de posición en 3 dimensiones. Para el enfoque inter-joint, los codos y muñecas presentan los valores más altos, así como las rodillas y los pies pero estos últimos en menor medida. Además se observa que existe simetría entre las articulaciones mencionadas anteriormente con respecto a todas las métricas utilizadas para el análisis inter-emociones. Para el enfoque intra-joint son relevantes la mayoría de las articulaciones de los miembros superiores: codos, muñecas y manos; en particular, la muñeca y mano derecha. También se puede observar que las 3 articulaciones de las extremidades inferiores tienen valores altos, siendo la rodilla la que tiene valores superiores y similares independientemente de la métrica. Estos resultados están asociados al hecho de que los eslabones más lejanos de las extremidades inferiores y superiores se mueven más durante la marcha. Las articulaciones de las extremidades superiores sugieren que pueden ser decisivas para diferenciar entre pares de emociones. Por el contrario, las articulaciones más próximas a la raíz presentan valores muy pequeños de movimiento para ambos enfoques (inter e intra), lo que se puede relacionar con la baja variabilidad entre emociones durante la marcha.



**Figura 26.** Enfoque inter-joint: Comparación de la relevancia de las articulaciones.



**Figura 27.** Enfoque intra-joint: Comparación de la relevancia de las articulaciones.

En el análisis de correlación de Pearson (ver Figura 28), se puede observar cuáles articulaciones están fuertemente correlacionadas, donde se notan patrones en la extremidades del cuerpo, es decir, se contempla en el bloque del centro que las articulaciones que conforman el tronco están muy correlacionadas entre sí. Las siguientes articulaciones que están correlacionadas fuertemente son las de la extremidad superior izquierda y derecha. Finalmente se observó que las articulaciones de las extremidades inferiores son las menos correlacionadas, cuando se comparan con las otras extremidades, se mostró que la extre-

idad derecha tiene más correspondencia que la extremidad inferior izquierda. Esto indica que pudiera haber información redundante entre articulaciones contiguas y tiene que ver con el hecho de que las articulaciones en conjunto conforman una cadena cinemática como se explicó en el Capítulo 2.

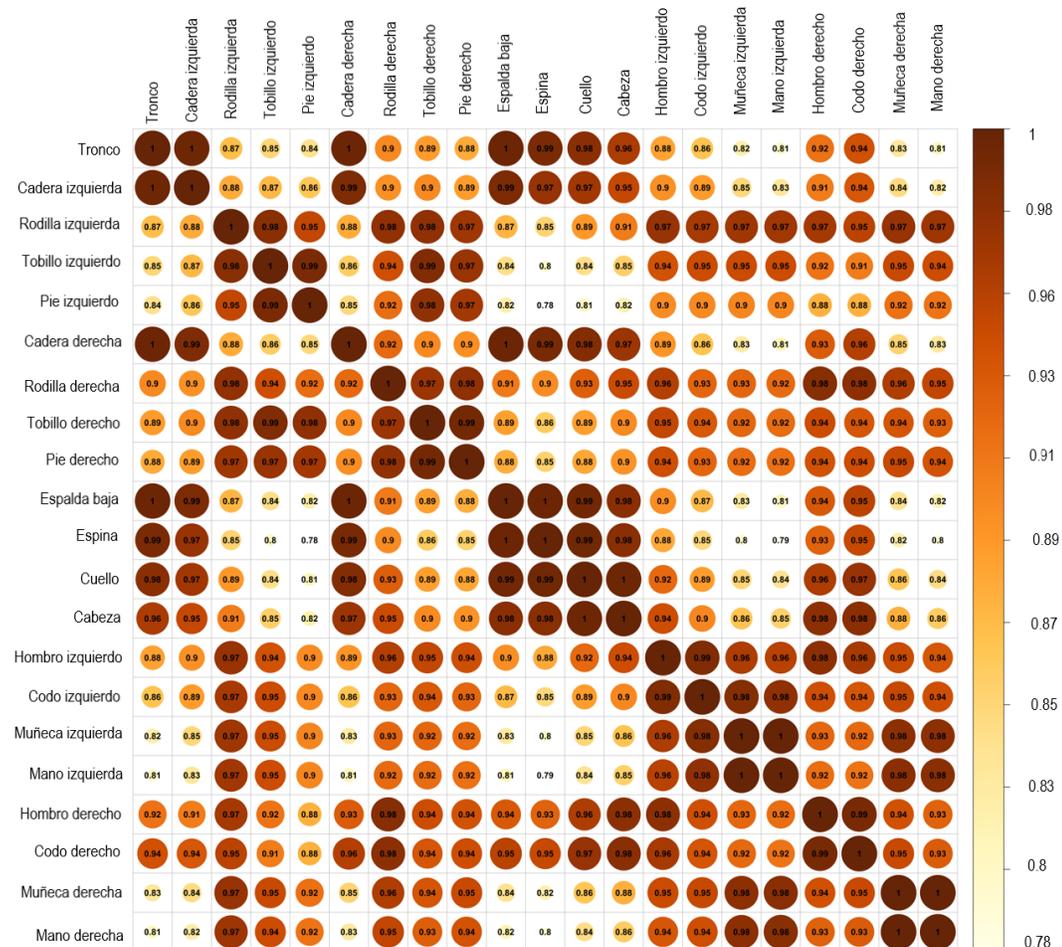


Figura 28. Correlación de Pearson aplicada a las 21 articulaciones para todas las emociones.

### 5.1.3. Selección de subconjuntos de articulaciones relevantes.

Con base en el análisis descriptivo previo y las características de la cinemática del movimiento humano, se seleccionaron siete subconjuntos de articulaciones  $S = \{s_1, s_2, \dots, s_7\}$ , los cuales se muestran en la Figura 29 y se describen a continuación:

- El primer subconjunto ( $S_1$ ) está compuesto por todas las articulaciones que presentan la mayor variabilidad de movimiento entre articulaciones, dando como resultado un subconjunto con 14 articulaciones.
- El segundo subconjunto ( $S_2$ ) está compuesto por las articulaciones con mayor variabilidad, pero solo

considerando las tres articulaciones finales (mano, muñeca y codo) para los miembros superiores e inferiores, dando un modelo de 12 articulaciones.

- El tercer subconjunto ( $S_3$ ) se basa en la idea inversa: la marcha tiene características de movimiento similares independientemente de la emoción, por lo que se consideraron las articulaciones con la menor cantidad de movimiento.
- El cuarto subconjunto se basa en la misma ideología de conservar las articulaciones con la menor cantidad de movimiento relativo entre articulaciones ( $S_4 = articulaciones - S_2$ ).
- El quinto subconjunto ( $S_5$ ) considera los puntos finales del esqueleto (que solo se vinculan con una articulación): manos, pies, cabeza y agregando la articulación de la raíz como referencia.
- La siguiente configuración ( $S_6$ ) solo considera las articulaciones del codo y la rodilla, que son las que aportan mayor relevancia en los miembros superiores e inferiores, y además añade como referencia la raíz.
- Finalmente, el subconjunto más pequeño ( $S_7$ ) comprende las cuatro articulaciones que componen el tronco: raíz, espalda baja, columna y cuello.

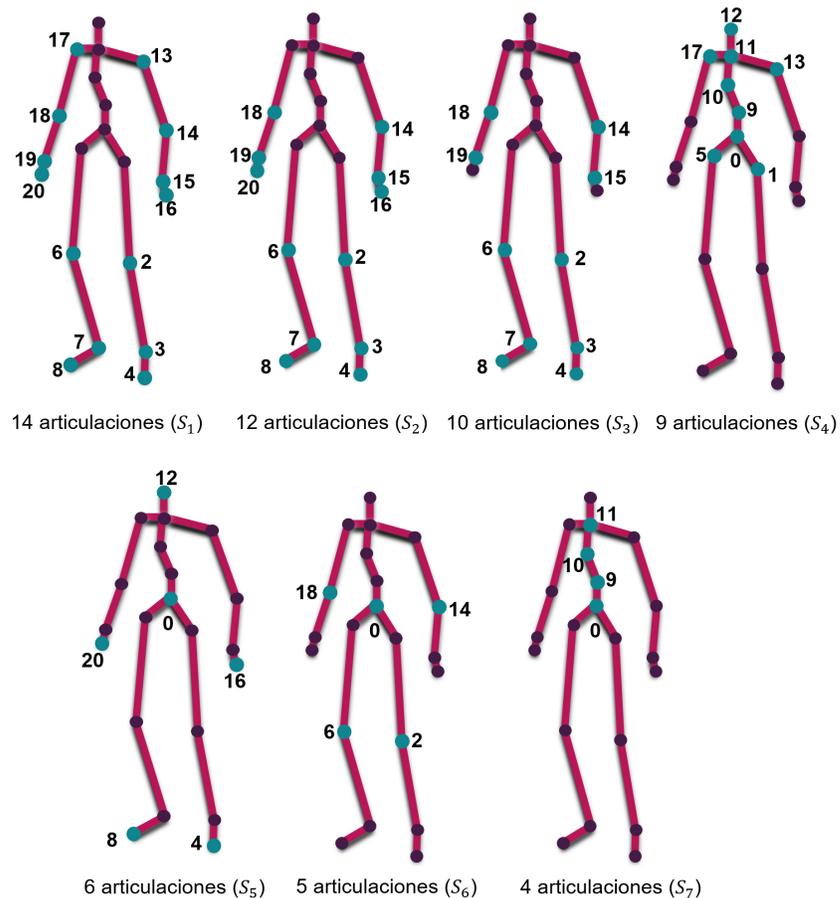


Figura 29. Subconjuntos de la representación completa.

## 5.2. Clasificación de emociones

Se construyeron y evaluaron una serie de modelos de clasificación, basados en los algoritmos de inferencia descritos en la Subsección 4.5, con cada uno de los subconjuntos de características de la Figura 29 considerando dos conjuntos de datos: (i) el conjunto de datos original que está desbalanceado, es decir, 1856 segmentos de marchas, y (ii) una muestra del conjunto de datos original que está balanceado, es decir, 324 marchas. Ambos conjuntos de datos se dividieron en conjuntos de entrenamiento y prueba utilizando una división de datos en 80 % de entrenamiento, 10 % de validación y 10 % de prueba. Se calcularon las métricas de sensibilidad, especificidad y F1-Score por emoción, y la precisión promedio para todas las clases (mAP) para cada clasificador.

### 5.2.1. Algoritmos de aprendizaje de máquina tradicionales

En esta sección se reportan los resultados de las marchas tanto para los datos balanceados y desbalanceados de 21 articulaciones que es la base para comparar con los subconjuntos de articulaciones ( $S_1, \dots, S_7$ ) y además se muestran los resultado del mejor subconjunto para todos los clasificadores. Los resultados restantes para el resto de subconjuntos de articulaciones se encuentran en el Apéndice B.

#### 5.2.1.1. Datos balanceados

Analizando el subconjunto de 21 articulaciones (ver Tabla 4) se observa que el clasificador con el mAP más alto es KNN con el método de selección univariado (UVA) el cual presenta un mAP de 0.784 (número de características = 2,500), para este conjunto de articulaciones se puede ver que el mejor clasificador es KNN sin importar el método de selección automática. Analizando a nivel de emociones el *F1-score*: feliz alcanzó 0,800 para el algoritmo de *random forest* con el método de selección análisis de componentes principales (PCA); triste 0.909 con los clasificadores *boosting* para todos los métodos de selección de características automáticas (AF-SEL); enojado 0.667 con los algoritmos KNN y *random forest* con todos los métodos AF-SEL; y neutral 0.737 con KNN para el método sin selección de características (NSL) y con los métodos de selección UVA y PCA. En resumen, las emociones mejor clasificadas son feliz y triste debido a que sus métricas presentan resultados altos y no hay mucha dispersión entre las métricas que se extrajeron.

El mejor modelo para clasificar las emociones, cuando el número de marchas es el mismo para cada emoción, se basó en el algoritmo *boosting* usando el método de selección PCA con el subconjunto de 9 articulaciones con un mAP de 0.840 (número de características = 29), ver Tabla 5. Sin embargo,

considerando los 3 algoritmos para el subconjunto de 9 articulaciones, el mAP promedio para el método de selección MDI es 0.770 con una desviación estándar de 0.024 (número de características = 1,386), mientras que el mAP para el método Univariado es de 0.754 pero con una desviación estándar de 0.025 (número de características = 2,500), valores similares a los que se usan con el modelo de todas las características (número de características). Esto indica que para esta tarea se puede usar un conjunto de características 805 veces más pequeño en comparación del modelo que emplea todas las características extraídas. Por otro lado, al considerar todas las articulaciones, el algoritmo kNN con el método de selección Univariado tiene el mejor mAP de 0.784, mientras que *boosting* utilizando todas las características tiene el mAP más bajo de 0.629. Tomando en cuenta los 3 algoritmos, el mAP promedio con el método de selección UVA es de 0.734 con una desviación estándar de 0.044 (número de características = 2,500), a diferencia de usar todas las características, donde el promedio mAP es 0.707 y presenta una desviación estándar de 0.070 (número de características = 23,373). Al revisar los mejores mAPs se puede inferir que no es necesario emplear todas las articulaciones para obtener buenos resultados, es decir, ver las métricas equilibradas en todas las clases.

Analizando las emociones individualmente según la puntuación F1 para el subconjunto de 9 articulaciones, el cual es el mejor modelo: feliz alcanzó 0,800 con todos los algoritmos de *boosting* para todo los métodos AF-SEL; triste 0.909 con *boosting* para todo los métodos AF-SEL; enojado 0.929 con KNN con el método de selección Univariado; y neutral 0.737 con KNN para los métodos de selección NSL, UVA y PCA. En general, la tasa positiva verdadera más alta, o sensibilidad, es para feliz con 0.833 (sd = 0.055), y la más baja fue 0,600 (sd = 0) para enojado; mientras que la tasa negativa verdadera más alta, o especificidad, fue para triste con 1 (sd = 0), y la más baja también fue para enojado con 0.843 (sd = 0.018). En resumen, la emoción que peor se puede clasificar es enojado, y además, al ser una de las clases con menor número de instancias los algoritmos no pudieron aprender a diferenciarla.

**Tabla 4.** Conjunto de 21 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
21 JS	NSL	23,373	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.763</b>
		23,373	RF	<b>0.857</b>	0.885	0.750	0.750	0.952	0.818	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	0.667	0.833	0.632	0.729
		23,373	BT	<b>0.857</b>	0.885	0.750	0.833	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.667	0.833	0.632	0.629
	UVA	2,500	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.784</b>
		2,500	RF	0.714	<b>0.923</b>	0.714	0.833	0.952	0.870	<b>0.600</b>	0.893	0.545	0.667	<b>0.875</b>	0.667	0.706
		2,500	BT	<b>0.857</b>	0.885	0.750	0.833	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.667	0.833	0.632	0.711
	MDI	1,386	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.763</b>
		1,386	RF	<b>0.857</b>	0.885	0.750	0.750	0.952	0.818	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	0.667	0.833	0.632	0.729
		1,386	BT	<b>0.857</b>	0.885	0.750	0.833	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.667	0.833	0.632	0.629
	PCA	29	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.763</b>
		29	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	0.833	0.952	0.870	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	0.667	0.833	0.632	0.752
		29	BT	<b>0.857</b>	0.885	0.750	<b>0.857</b>	0.885	0.750	0.200	0.893	0.222	0.667	0.833	0.632	0.629

**Tabla 5.** Subconjunto de 9 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
9 JS	NSL	10,017	KNN	<b>0.857</b>	0.885	0.750	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.929	<b>0.600</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.742
		10,017	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>
		10,017	BT	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>
	UVA	2,500	KNN	<b>0.857</b>	0.885	0.750	0.600	<b>1.000</b>	0.857	<b>0.600</b>	0.600	<b>0.929</b>	<b>0.778</b>	<b>0.875</b>	0.737	0.742
		2,500	RF	0.714	0.923	0.714	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.929	<b>0.600</b>	<b>0.778</b>	0.833	0.700	0.738
		2,500	BT	0.857	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>
	MDI	1,386	KNN	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	0.742
		1,386	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>
		1,386	BT	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>
	PCA	29	KNN	<b>0.857</b>	0.885	0.750	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.929	<b>0.600</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.742
		29	RF	0.714	<b>0.923</b>	0.714	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.929	<b>0.600</b>	<b>0.778</b>	0.833	0.700	0.738
		29	BT	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	0.833	0.700	<b>0.784</b>

### 5.2.1.2. Datos desbalanceados

Analizando el subconjunto de 21 articulaciones (ver Tabla 6) se observa que el clasificador con el mAP más alto es *random forest* con el método de selección PCA que presenta un mAP de 0.829 (número de características = 29), mientras que KNN utilizando todas las características tiene el mAP más bajo con 0.719. Analizando a nivel de emociones el *F1-score*: feliz alcanzó 0,776 con el algoritmo de *random forest* para el método de selección PCA; triste 0.989 con los clasificadores *random forest* para todo los métodos AF-SEL; enojado 0.863 con el algoritmo de *boosting* con todos los métodos AF-SEL; y neutral 0.500 con *random forest* para el método de selección PCA. En resumen, las emociones mejores clasificadas son triste y enojado debido a que sus métricas presentan resultados más altos y no hay mucha dispersión entre ellas.

El mejor modelo para clasificar las emociones cuando los datos están desbalanceados se basó en el algoritmo *random forest* usando el método de selección Univariado con el subconjunto de 9 articulaciones con un mAP de 0.886 (número de características = 2,500), ver Tabla 7. Sin embargo, considerando los 3 algoritmos para el subconjunto de 9 articulaciones, el mAP promedio para el método de selección MDI es 0.788 con una desviación estándar de 0.093 (número de características = 1,386), mientras que el mAP para el método Univariado es de 0.806, pero con una desviación estándar de 0.109 (número de características = 2,500), los valores presentan diferencias notorias entre los modelos que usan todas las características (número de características), es decir, presentan un mAP promedio más pequeño y una dispersión más grande. Esto indica que para esta tarea se puede usar un conjunto de características 9 veces más pequeño a comparación del modelo que emplea todas las características extraídas. Tomando en cuenta los 3 algoritmos, el mAP promedio con el método de selección PCA es de 0.786 con una desviación estándar de 0.059 (número de características = 29), a diferencia de usar todas las características, donde el promedio mAP es 0.772 y presenta una desviación estándar de 0.048 (número de características = 23,373).

Analizando las emociones individualmente con F1-score para el subconjunto de 9 articulaciones, el cual es el mejor modelo: feliz alcanzó 0,824 con RF + método de selección UVA; triste 0.984 con *boosting* + cualquier AF-SEL; enojado 0.872 con *boosting* + cualquier AF-SEL; y neutral 0.667 con *boosting* + cualquier AF-SEL. En general, la sensibilidad más alta fue para triste con 0.961 (sd = 0.028), y la más baja fue 0.464 (sd = 0.202) para neutral; mientras que la especificidad más alta fue para neutral con 0.985 (sd = 0.004), y la más baja para enojado con 0.908 (sd = 0.007). En esta ocasión, las marchas neutrales no pudieron clasificarse correctamente, y puede coincidir que al ser la clase minoritaria, los algoritmos no pudieron aprender a diferenciarla; aunque las marchas mal clasificadas de las otras emociones no se confundieron con la clase neutral.

**Tabla 6.** Conjunto de 21 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
21 JS	NSL	23,373	KNN	<b>0.700</b>	0.916	0.667	0.915	0.949	0.935	0.857	0.924	0.818	0.286	0.988	0.364	0.719
		23,373	RF	0.567	<b>1.000</b>	0.723	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.885	<b>0.825</b>	<b>0.429</b>	0.976	0.429	0.786
		23,373	BT	0.667	0.979	0.755	<b>0.989</b>	0.975	0.984	0.905	<b>0.893</b>	0.809	0.286	<b>0.994</b>	0.400	<b>0.811</b>
	UVA	2,500	KNN	<b>0.700</b>	0.916	0.667	0.915	0.949	0.935	0.857	0.924	0.818	0.286	0.988	0.364	0.719
		2,500	RF	0.600	0.979	0.706	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.905	0.885	0.800	0.286	0.982	0.333	0.741
		2,500	BT	0.667	0.979	0.755	<b>0.989</b>	0.975	0.984	0.905	<b>0.893</b>	0.809	0.286	<b>0.994</b>	0.400	<b>0.811</b>
	MDI	1,386	KNN	0.567	0.986	0.694	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.878	0.804	0.286	0.982	0.333	0.719
		1,386	RF	<b>0.700</b>	0.916	0.667	0.915	0.949	0.935	0.857	0.924	0.818	0.286	0.988	0.364	0.748
		1,386	BT	0.667	0.979	0.755	<b>0.989</b>	0.975	0.984	0.905	<b>0.893</b>	0.809	0.286	<b>0.994</b>	0.400	<b>0.811</b>
	PCA	29	KNN	<b>0.700</b>	0.916	0.667	0.915	0.949	0.935	0.857	0.924	0.818	0.286	0.988	0.364	0.719
		29	RF	0.633	<b>1.000</b>	<b>0.776</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.885	<b>0.825</b>	<b>0.429</b>	0.988	<b>0.500</b>	<b>0.829</b>
		29	BT	0.667	0.979	0.755	<b>0.989</b>	0.975	0.984	0.905	<b>0.893</b>	0.809	0.286	<b>0.994</b>	0.400	0.811

**Tabla 7.** Subconjunto de 9 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
9 JS	NSL	10,017	KNN	0.533	0.902	0.533	0.947	0.899	0.932	0.738	<b>0.931</b>	0.756	0.286	0.976	0.308	0.640
		10,017	RF	0.667	0.993	0.784	<b>0.989</b>	0.987	<b>0.989</b>	<b>0.976</b>	0.901	0.854	0.286	<b>0.988</b>	0.364	0.800
		10,017	BT	0.667	<b>1.000</b>	0.800	0.979	<b>1.000</b>	<b>0.989</b>	<b>0.976</b>	0.908	<b>0.863</b>	<b>0.714</b>	0.982	<b>0.667</b>	<b>0.850</b>
	UVA	2,500	KNN	0.533	0.923	0.561	0.926	0.873	0.911	0.786	0.908	0.759	0.286	<b>0.988</b>	0.364	0.681
		2,500	RF	<b>0.700</b>	<b>1.000</b>	<b>0.824</b>	<b>0.989</b>	0.987	<b>0.989</b>	0.952	0.901	0.842	0.571	0.994	0.667	<b>0.886</b>
		2,500	BT	0.667	<b>1.000</b>	0.800	0.979	<b>1.000</b>	<b>0.989</b>	<b>0.976</b>	0.908	<b>0.863</b>	<b>0.714</b>	0.982	<b>0.667</b>	<b>0.850</b>
	MDI	1,386	KNN	0.533	0.923	0.561	0.926	0.873	0.911	0.786	0.908	0.759	0.286	<b>0.988</b>	0.364	0.681
		1,386	RF	0.533	0.923	0.561	0.926	0.873	0.911	0.786	0.908	0.759	0.286	<b>0.988</b>	0.364	0.833
		1,386	BT	0.667	<b>1.000</b>	0.800	0.979	<b>1.000</b>	<b>0.989</b>	<b>0.976</b>	0.908	<b>0.863</b>	<b>0.714</b>	0.982	<b>0.667</b>	<b>0.850</b>
	PCA	29	KNN	0.533	0.923	0.561	0.926	0.873	0.911	0.786	0.908	0.759	0.286	<b>0.988</b>	0.364	0.681
		29	RF	0.667	0.993	0.784	<b>0.989</b>	0.987	<b>0.989</b>	0.952	0.901	0.842	0.429	<b>0.988</b>	0.500	0.824
		29	BT	0.667	<b>1.000</b>	0.800	0.979	<b>1.000</b>	<b>0.989</b>	<b>0.976</b>	0.908	<b>0.863</b>	<b>0.714</b>	0.982	<b>0.667</b>	<b>0.850</b>

### 5.2.2. Algoritmos de aprendizaje profundo

En esta sección se reportan los resultados considerando tanto los datos balanceados como los datos originales (desbalanceados) utilizando los algoritmos de aprendizaje profundo descritos en la Subsección 2.4.2. Solo se considero el conjunto de 21 articulaciones con el objetivo de comparar directamente los resultados con el estado del arte cuyos modelos consideran el mismo número de articulaciones.

### 5.2.2.1. Datos balanceados

En la Tabla 8 se observa que el mejor algoritmo para clasificar emociones con el número de instancias equilibradas se basó en el algoritmo CNN-1 con el subconjunto de 21 articulaciones, el mAP es de 0.658. Inspeccionando de modo particular por emociones con el F1-Score se puede observar que: para feliz se alcanzó 0.870 con la CNN-3; triste con 0.545 para CNN-1; enojado con 0.654 con la CNN-3 y neutral con 0.333 para CNN-3. En particular, la sensibilidad más alta fue para feliz con 0.833 (sd = 0), y la más baja fue 0.286 (sd = 0.143) para enojado; mientras que la especificidad más alta fue para enojado con 0.936 (sd = 0.022), y la más baja para triste con 0.821 (sd = 0.071); esto indica que las instancias mal clasificadas fueron confundidas principalmente con triste.

**Tabla 8.** Subconjunto de 21 articulaciones para datos balanceados para algoritmos de aprendizaje profundo.

Conjunto de articulaciones	Clasificadores	Feliz			Triste			Enojado			Neutral			mAP
		Spe	Sen	F1	Sen	Spe	F1	Sen	Spe	F1	Sen	Spe	F1	
21JS	CNN-1	<b>0.833</b>	0.905	0.833	<b>0.600</b>	<b>0.893</b>	<b>0.545</b>	<b>0.429</b>	0.923	<b>0.500</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.658</b>
	CNN-2	<b>0.833</b>	0.857	0.800	<b>0.600</b>	0.821	0.462	0.143	0.923	0.200	0.667	<b>0.875</b>	0.667	0.535
	CNN-3	<b>0.833</b>	<b>0.952</b>	<b>0.870</b>	<b>0.600</b>	0.750	0.400	0.286	<b>0.962</b>	0.400	0.444	0.792	0.444	0.577

### 5.2.2.2. Datos desbalanceados

En la Tabla 9 se observa que el mejor modelo para clasificar emociones con el número de instancias desequilibradas se basó en el algoritmo CNN-3 con el conjunto de 21 articulaciones, el mAP es de 0.704. Inspeccionando de modo particular por emociones con el F1-Score: se puede ver que para feliz se alcanzó 0.981 con la CNN-1 y CNN-3; triste con 0.762 para la CNN-1; enojado con 0.654 para la CNN-3 y neutral con 0.333 para la CNN-2. En particular, la sensibilidad más alta fue para feliz con 0.946 (sd = 0.031), y la más baja fue 0.300 (sd = 0.173) para neutral; mientras que la especificidad más alta fue para neutral con 0.962 (sd = 0.041), y la más baja para triste con 0.906 (sd = 0.021). Al igual que para los modelos basados en los aprendizaje de máquina tradicional la clase con menor número de instancias (neutral) obtuvo resultados bajos.

**Tabla 9.** Subconjunto de 21 articulaciones para datos desbalanceados para algoritmos de aprendizaje profundo.

Conjunto de articulaciones	Clasificadores	Feliz			Triste			Enojado			Neutral			mAP
		Spe	Sen	F1	Sen	Spe	F1	Sen	Spe	F1	Sen	Spe	F1	
21JS	CNN-1	0.951	<b>1.000</b>	<b>0.981</b>	0.762	<b>0.930</b>	<b>0.762</b>	<b>0.621</b>	0.916	0.600	0.200	<b>0.989</b>	0.286	0.701
	CNN-2	0.914	0.903	0.916	0.619	0.901	0.634	0.345	0.910	0.377	<b>0.500</b>	0.914	<b>0.333</b>	0.561
	CNN-3	<b>0.975</b>	0.981	<b>0.981</b>	<b>0.881</b>	0.887	0.779	0.586	<b>0.961</b>	<b>0.654</b>	0.200	0.983	0.267	<b>0.704</b>

### 5.2.3. Replicación de algoritmos del estado del arte

En esta sección se exponen los resultados reportados en el Estado del Arte (ver Tabla 10) y los resultados obtenidos de la replicación de los mismos algoritmos encontrados en la literatura (Tabla 11).

La Tabla 10 muestra los resultados reportados en el estado del arte, se observa que el mejor modelo para clasificar emociones cuando el número de marchas es diferente para cada emoción se basó en el algoritmo TAEW con el subconjunto de 21 articulaciones, el mAP es 0.840. Si nos enfocamos en las emociones de manera individual con AP para el conjunto de 21 articulaciones: feliz alcanzó 0.980 con STGCN, DGNN y TAEW; triste tuvo un AP de 0.890 con el algoritmo TAEW; enojado tuvo un AP de 0.810 con TAEW y el estado neutral tuvo un AP 0.710 con el algoritmo TAEW.

**Tabla 10.** Resultados reportados en el Estado del Arte.

Joint Subset	Classifiers	AP				mAP
		Happy	Sad	Angry	Neutral	
21JS	STGCN	<b>0.980</b>	0.830	0.420	0.180	0.610
	DGNN	<b>0.980</b>	0.880	0.730	0.370	0.740
	STEP	0.970	0.880	0.720	0.520	0.770
	LSTM	0.960	0.840	0.620	0.620	0.730
	TAEW	<b>0.980</b>	<b>0.890</b>	<b>0.810</b>	<b>0.710</b>	<b>0.840</b>

Los resultados mostrados en la Tabla 11 son producto de la replicación de los algoritmos previos. Una de las peculiaridades al tratar de replicar los resultados, es que no se pudieron llegar a los mismos valores reportados. Se observa que el mejor modelo para clasificar emociones es el algoritmo DGNN, con el mAP de 0.784. Si nos centramos en las emociones de manera individual con AP para el conjunto de 21 articulaciones: feliz alcanzó 0.970 con STGCN; triste tuvo un AP de 0.956 con DGNN; enojado alcanzó 0.943 con DGNN y el estado neutral tuvo un AP 0.250 con el algoritmo DGNN. Es importante hacer notar que el mAP de el clasificador DGNN resultó mayor que el reportado por los autores mientras que el clasificador TAEW, que fue el valor más alto reportado, disminuyo.

**Tabla 11.** Replicación de algoritmos y parámetros del Estado del Arte para 21 articulaciones.

Conjunto de articulaciones	Classifiers	AP				mAP
		Happy	Sad	Angry	Neutral	
21 JS	STGCN	<b>0.970</b>	0.224	0.106	0.091	0.348
	DGNN	0.985	<b>0.956</b>	<b>0.943</b>	<b>0.250</b>	<b>0.784</b>
	STEP	0.520	0.301	0.217	0.123	0.279
	LSTM	0.636	0.283	0.234	0.125	0.280
	TAEW	0.790	0.890	0.668	0.124	0.770

### 5.3. Comparativa de los mejores clasificadores

#### 5.3.1. Mejores clasificadores para datos balanceados

A continuación se analizan los resultados de los mejores clasificadores propuestos al utilizar los datos balanceados.

En la Tabla 12 se presentan los mejores modelos por subconjunto de articulaciones para los datos que están balanceados. En general, se puede observar que los mejores clasificadores no utilizan el modelo de articulaciones completo y además utilizan un método de selección de características, principalmente PCA y MDI, basándose principalmente en los modelos de *boosting* y *random forest*. El mejor modelo combinó el algoritmo *boosting* y el método de selección PCA utilizando en subconjunto de 21 articulaciones obteniendo un mAP de 0.840, clasificando correctamente todas las instancias de triste y siendo neutral la clase con menor tasa de clasificación. Si analizamos las emociones de forma individual con la métrica AP, feliz tiene 0.857 con el algoritmo *boosting* usando el método de selección PCA (número de características = 29); triste con 1 para todos los algoritmos que emplean *boosting* y el método de selección PCA (número de características = 29); enojado en particular alcanzó un valor de 1 en su mayoría con el algoritmo RF y neutral tiene 0.727 con los clasificadores RF empleando los métodos de selección MDI (número de características = 1,386) y PCA (número de características = 29).

**Tabla 12.** Comparación de los mejores clasificadores para datos balanceados.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
9 JS	BT-PCA	0.750	<b>1.000</b>	0.750	0.636	<b>0.840</b>
12 JS	RF-MDI	0.667	0.909	<b>1.000</b>	<b>0.727</b>	0.826
6 JS	BT-PCA	0.600	<b>1.000</b>	<b>1.000</b>	0.636	0.809
14 JS	RF-MDI	0.600	0.900	<b>1.000</b>	<b>0.727</b>	0.807
10 JS	RF-PCA	0.545	0.900	<b>1.000</b>	<b>0.727</b>	0.793
4 JS	BT-PCA	<b>0.857</b>	0.909	0.750	0.636	0.788
21 JS	KNN-UVA	0.500	<b>1.000</b>	<b>1.000</b>	0.636	0.784
5 JS	BT-PCA	0.800	<b>1.000</b>	0.500	0.700	0.750
21 JS	CNN-1	0.833	0.500	0.600	0.700	0.697

#### 5.3.2. Mejores clasificadores para datos desbalanceados

A continuación se analizan los resultados de los mejores clasificadores propuestos al utilizar los datos originales, en comparación con el Estado del Arte.

En general, se puede observar que los mejores clasificadores no utilizan el modelo de articulaciones completo y además utilizan un método de selección de características, principalmente PCA y MDI, basándose principalmente en los modelos de *boosting* y *random forest*. El mejor modelo combinó el algoritmo *boosting* y el método de selección PCA utilizando un subconjunto de 21 articulaciones obteniendo un mAP de 0.840, clasificando correctamente todas las instancias de triste y siendo neutral la clase con menor tasa de clasificación.

En la Tabla 13 se muestran los resultados de los mejores modelos por subconjunto de articulaciones para los datos que están desbalanceados. En general, se puede observar que los tres mejores modelos no utilizan el conjunto de articulaciones completo e inclusive utilizan un método de selección que reduce de manera notoria el número de características. El mejor modelo combinó el algoritmo *random forest* y el método de selección UVA utilizando el subconjunto de 9 articulaciones obteniendo un mAP de 0.886, clasificando correctamente todas las instancias de feliz y siendo neutral la clase con menor tasa de clasificación. Si analizamos las emociones de forma individual con la métrica AP. Feliz tiene 1 en todos los subconjuntos que emplean el algoritmo RF usando todos los métodos de selección; triste con 0.989 tiende a tener el valor más alto con los subconjuntos que utilizan RF con cualquier método de selección e inclusive sin método de selección y con el método de selección PCA; enojado con 0.810 con el algoritmo TAEW y neutral con 0.800 con el algoritmo *boosting* y el método de selección PCA. Se resalta en particular el tercer mejor clasificador el cual en comparación al mejor resultado en la literatura (TAEW), mejora en la clasificación de las instancias de feliz, triste y neutral. Es importante señalar que los clasificadores construidos con algoritmos de aprendizajes de máquina tradicionales son competitivos con respecto a los modelos basados en aprendizaje profundos.

**Tabla 13.** Comparación de los mejores clasificadores para datos desbalanceados.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
9 JS	RF-UVA	<b>1.000</b>	0.979	0.755	0.600	<b>0.886</b>
4 JS	BT-PCA	0.909	0.979	0.788	<b>0.800</b>	0.869
6 JS	RF-MDI	<b>1.000</b>	<b>0.989</b>	0.702	0.750	0.860
21 JS	TAEW*	0.980	0.890	<b>0.810</b>	0.710	0.840
5 JS	BT-PCA	0.917	<b>0.989</b>	0.750	0.667	0.831
21 JS	RF-PCA	<b>1.000</b>	<b>0.989</b>	0.727	0.600	0.829
10 JS	RF-NSL	0.944	<b>0.989</b>	0.714	0.600	0.812
14 JS	RF-MDI	0.947	<b>0.989</b>	0.714	0.500	0.788
21 JS	DGNN	0.985	0.956	0.943	0.250	0.784
12 JS	RF-NSL	0.947	<b>0.989</b>	0.727	0.400	0.766
21 JS	CNN-3	0.981	0.779	0.654	0.267	0.704

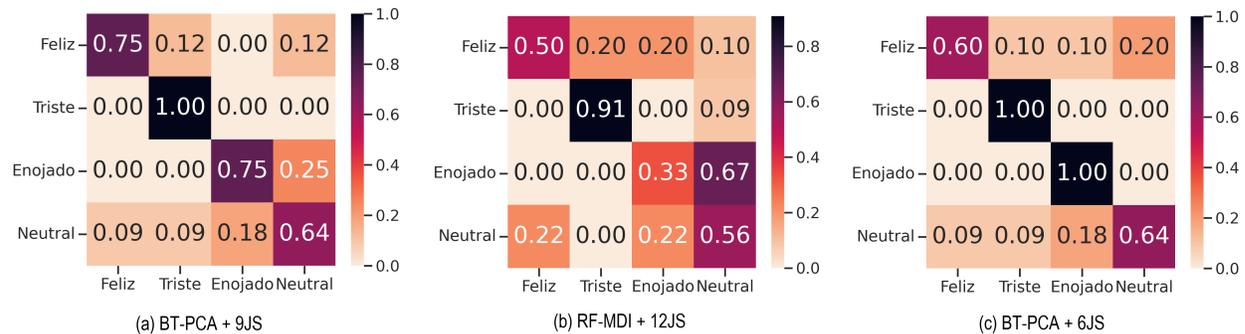
### 5.3.3. Discusión

Posteriormente, se seleccionaron los 3 mejores clasificadores por tipo de datos y se analizó en detalle la proporción de marchas correctamente clasificadas o mal clasificadas.

Las matrices de confusión de la Figura 30 representan el conjunto de datos balanceados. Como se puede observar, los modelos basados en *boosting* (9 y 6 articulaciones) con el método de selección PCA clasifican correctamente triste, lo que confirma su alta sensibilidad y especificidad; mientras que las instancias mal clasificadas de las otras emociones no se confunden con triste.

Este mismo patrón ocurre en el modelo basado con RF (12 articulaciones) con el método de selección MDI, sin embargo, una proporción mínima de las marchas de triste se clasificaron erróneamente como neutral.

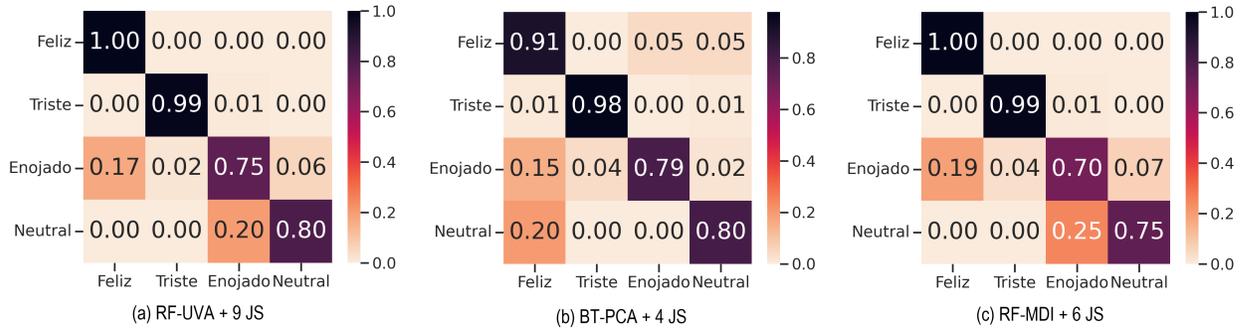
Además se puede notar que la emoción que presenta mayor confusión es enojado; existe una proporción significativa donde las marchas de enojado se clasificaron erróneamente como neutral y en menor medida como feliz.



**Figura 30.** Matrices de confusión para los mejores clasificadores usando datos balanceados.

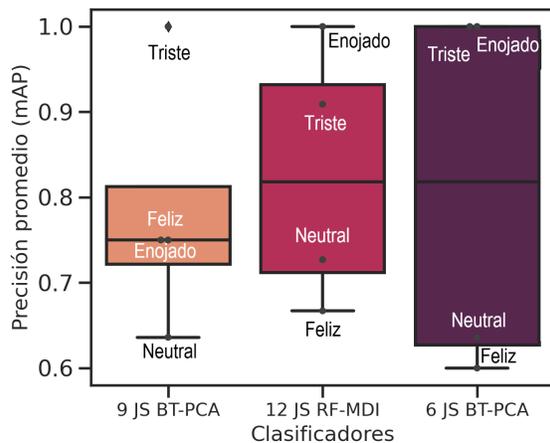
La Figura 31 presenta las matrices de confusión cuando se utiliza el conjunto de datos original. Al igual que con el conjunto de datos balanceado, los modelos basados en RF (9 y 21 articulaciones) clasifican correctamente a triste, incluso con mayor sensibilidad y especificidad; mientras que hay confusión en una ligera proporción con enojo.

Una vez más, la tasa de confusión más alta para todas las emociones se observó en la matriz de confusión del modelo basado en RF (9 articulaciones).

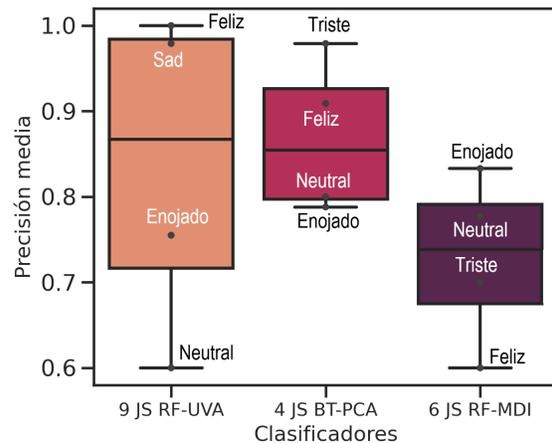


**Figura 31.** Matrices de confusión para los mejores clasificadores usando datos desbalanceados.

Finalmente, las Figuras 32 y 33 son una comparación de los algoritmos *random forest* y *boosting*, usando el promedio de precisión de todas las clases (mAP) para cada conjunto de datos. Como se puede observar, la mediana para los clasificadores construidos con datos no balanceados es levemente mayor que la que utiliza datos balanceados, además, presentan menor dispersión entre emociones. En todos los casos triste es la emoción que mejor se puede clasificar, mientras que en general feliz es la peor para datos balanceados y neutral para datos no balanceados. Una de las peculiaridades que se puede observar es que mientras el modelo contenga un número reducido de articulaciones, la emoción que peor se desempeña es neutral, mientras feliz es la mejor.



**Figura 32.** Datos balanceados.



**Figura 33.** Datos desbalanceados.

**Figura 34.** Comparación de clasificadores de cada emoción.

Estos resultados indican que la marcha de sujetos jóvenes que tienen implícita la emoción triste es posible identificarlas utilizando modelos de clasificación basados en modelos de aprendizaje de máquina, en particular reduciendo el conjunto de características temporales, estadísticas y espectrales con un método

de selección automático. También se confirma que no es necesario utilizar todos los datos extraídos del esqueleto humano ya que con el análisis descriptivo realizado se encontraron subconjuntos que mejoran los resultados de clasificación que los reportados en el estado del arte y que utilizan el modelo completo de 21 articulaciones.

Se plantea también una relación con los hallazgos presentados en el estado del arte. La mayoría de los enfoques está basado ya sea en extraer características automáticamente a partir de modelos profundos, o extraer características de alto nivel, como ángulos, áreas o distancias entre la posición de las articulaciones. En este trabajo se utilizó solamente características temporales, estadísticas y espectrales, para entrenar modelos de aprendizaje automático, obteniendo, en algunos casos, mejores resultados. Se resalta esta aportación ya que sin usar aún representaciones de los datos sofisticadas, el rendimiento de los modelos de clasificación es prometedor. A partir de esto, se plantea un enfoque híbrido en el trabajo a futuro.

Respecto a la relación de las emociones respecto a la marcha reportada en la literatura. Se han identificado relaciones entre patrones de movilidad de las articulaciones que corresponden al tronco y la emoción de tristeza (Randhavane et al., 2019), y justamente los subconjuntos de nueve y cuatro articulaciones.  $S_4$  y  $S_7$ , respectivamente, permiten identificar esta emoción con muy buena precisión ( $AP = 0.979$ ). A diferencia de los autores, no hay confusión con instancias de la clase neutral.

## Capítulo 6. Conclusiones y trabajo a futuro

---

El gran reto de reconocer emociones es la subjetividad que existe entre ellas, existiendo mucha variación entre cómo representarlas, es decir, no hay modo de cuantificar, con plena objetividad, qué emoción se produce y en qué grado. De igual modo, la variabilidad fisiológica del ser humano agrega mucha complejidad sobre cómo reconocer las emociones a partir del análisis cuantitativo de la marcha humana. El presente trabajo ha pretendido desarrollar algoritmos que permitan detectar un conjunto de emociones sin importar las dificultades anteriores, pero con algunas limitaciones.

El objetivo general de esta investigación fue estudiar las emociones humanas y su relación con la marcha humana de sujetos jóvenes a partir del seguimiento de sus articulaciones.

### 6.1. Conclusiones

Para lograr este objetivo se realizó un análisis en profundidad de la importancia que tiene cada articulación del esqueleto para clasificar el conjunto de emociones propuestas utilizando algoritmos de aprendizaje automático (ver Capítulo 5, Sección 5.1). En este proceso se utilizaron los algoritmos *k-nearest neighbors* (*KNN*), *boosting* (*base support vector machines*) y *random forest*, sobre series de tiempo de datos extraídos de la marcha humana.

Durante la realización del trabajo se redujo la cantidad de articulaciones con respecto al modelo completo de 21 articulaciones. Como resultado se observó que el rendimiento de los clasificadores disminuía significativamente para los modelos entrenados con menos articulaciones a excepción de los modelos basados en 9 y 4 articulaciones (ver Capítulo 5, Sección 5.1.3).

Los resultados mostraron que a nivel emociones la mejor clasificada es tristeza y además se notó que el estado neutral debe abordarse con especial atención cuando los datos están desbalanceados (ver Capítulo 5, Sección 5.2.1.2).

Los resultados obtenidos al utilizar los datos con clases desbalanceadas fueron comparados directamente con los datos reportados en el Estado del Arte, obteniendo 0.886 de precisión promedio de todas las clases para nuestro clasificador *random forest* con el método de selección UVA que es el más prometedor, obteniendo una mejora de 0.046 con respecto al trabajo reportado del estado del arte (ver Capítulo 5, Sección 5.3.2).

A diferencia de los trabajos reportados en el Estado del Arte se decidió abordar el problema con una

estrategia que aprovechara al máximo la relación entre las emociones y la marcha. En otros trabajos se proponen nuevas métricas o se deja la tarea de extracción de características a las arquitecturas profundas y se espera que se ajusten a los datos, en este estudio, se realizó un análisis a profundidad de la cantidad de información que proporciona cada articulación para diferenciar el conjunto de emociones abordadas.

## **6.2. Contribuciones**

Tras la realización de este trabajo se realizaron contribuciones al estado actual de la clasificación de emociones a partir de patrones de la marcha humana.

La generación de modelos de clasificación explicativos a partir de la formulación y reducción de varios subconjuntos de articulaciones mediante un análisis descriptivo de los datos.

Se mejoraron las métricas de clasificación de la clase minoritaria neutral (81 marchas), a pesar del desequilibrio existente en comparación de la clase mayoritaria feliz (1019 marchas).

Se generaron diferentes características a las reportadas en el trabajo relacionado, aquellas pertenecientes a los grupos de familia temporales, espectrales y estadísticas.

A partir del análisis descriptivo y de la reducción de dimensiones de las características se superaron los resultados reportados en el estado del arte.

Publicación de un artículo con los resultados preliminares en la 14va Conferencia Internacional sobre Computación Ubicua e Inteligencia Ambiental (UCAmI2022).

## **6.3. Limitaciones**

Los experimentos realizados en este trabajo revelan que es necesario tener las clases balanceadas para tener una comparación justa entre las clases. Además que para entrenar modelos de aprendizaje profundo se necesitan un mayor numero de instancias de las que se utilizaron. Los datos de marcha utilizados fueron capturados en entornos mayormente controlados, lo que reduce la generalización de los modelos aprendidos.

## **6.4. Trabajo a futuro**

Conforme se desarrolló este trabajo surgieron nuevas vertientes que podrían enriquecerlo, a pesar de que no se desarrollaron quedan como propuesta para trabajo a futuro y se mencionan a continuación:

Para mejorar el rendimiento de los modelos se requiere aumentar el número de datos. El propósito de aumentar los datos es tener balanceadas todas las clases, evitar que exista confusión entre las clases minoritarias y tener más datos para reentrenar los modelos de aprendizaje profundo y clasifiquen mejor las series temporales de las marchas. Esto se puede obtener ya sea sintetizando datos o adquiriendo nuevos conjuntos de datos con las características estudiadas, incluso con otras fuentes de adquisición de datos, por ejemplo, cámaras estereoscópicas, dispositivos LIDAR, ToF cámaras, entre otros.

Otro enfoque que se le podría dar a este trabajo de tesis es implementar nuevas clases que reflejen emociones que compartan similitudes utilizando un enfoque multimodal, es decir, aparte de combinar solo datos de la marcha humana, combinarlo con datos de gestos, audio de diversas personas repitiendo diversas frases, datos fisiopsicologicos (Heart rate, tensión muscular, entre otros) y parámetros espacio temporales.

De los algoritmos trabajados solo uno pertenecían a la rama de aprendizaje profundo y el resto de aprendizaje de máquina tradicional, entonces se puede enfocar este tema a explorar más a profundidad este primer grupo, por ejemplo, autoencoders, LSTM o transformers.

Estudiar la relación entre los parámetros espacio temporales de la marcha humana, como, la cadencia, la velocidad, las fases del ciclo de la marcha humana respecto a cada emoción.

## Literatura citada

- Acredolo, L. P. y Goodwyn, S. W.. 1985. Symbolic gesturing in language development. *Human development*, 28(1).
- Ahmed, F. y Gavrilova, M. L.. 2019. Two-layer feature selection algorithm for recognizing human emotions from 3d motion analysis. En: *Computer Graphics International Conference*..Springer.
- Ahmed, F., Bari, A. H., y Gavrilova, M. L.. 2019. Emotion recognition from body movement. *IEEE Access*, 8.
- Alm, C. O., Roth, D., y Sproat, R.. 2005. Emotions from text: machine learning for text-based emotion prediction. En: *Proceedings of human language technology conference and conference on empirical methods in natural language processing*.
- Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., y Gamboa, H.. 2020. TSFEL: Time series feature extraction library. *SoftwareX*, 11.
- Beauchet, O., Allali, G., Berrut, G., Hommet, C., Dubost, V., y Assal, F.. 2008. Gait analysis in demented subjects: Interests and perspectives. *Neuropsychiatric disease and treatment*, 4(1).
- Benitez-Quiroz, C. F., Srinivasan, R., Feng, Q., Wang, Y., y Martinez, A. M.. 2017. Emotionet challenge: Recognition of facial expressions of emotion in the wild. *arXiv:1703.01210*.
- Bhattacharya, U., Mittal, T., Chandra, R., Randhavane, T., Bera, A., y Manocha, D.. 2020a. Step: Spatial temporal graph convolutional networks for emotion perception from gaits. En: *Proceedings of the AAAI Conference on Artificial Intelligence*..Vol. 34.
- Bhattacharya, U., Roncal, C., Mittal, T., Chandra, R., Kapsaskis, K., Gray, K., Bera, A., y Manocha, D.. 2020b. Take an emotion walk: Perceiving emotions from gaits using hierarchical attention pooling and affective mapping. En: *European Conference on Computer Vision*..Springer.
- Cao, Z., Simon, T., Wei, S.-E., y Sheikh, Y.. 2017. Realtime multi-person 2d pose estimation using part affinity fields. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Carnegie, M.. 2001. Cmu graphics lab motion capture database. Consultado el 01 de julio de 2022. de <https://mocap.cs.cmu.edu>.
- Chiu, M., Shu, J., y Hui, P.. 2018. Emotion recognition through gait on mobile devices. En: *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*..IEEE.
- Cortes, C. y Vapnik, V.. 1995. Support-vector networks. *Machine learning*, 20(3).
- Crane, E. y Gross, M.. 2007. Motion capture and emotion: Affect detection in whole body movement. En: *International Conference on Affective Computing and Intelligent Interaction*..Springer.
- Dellaert, F., Polzin, T., y Waibel, A.. 1996. Recognizing emotion in speech. En: *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*..IEEE, Vol. 3.
- Ekman, P.. 1999. Basic emotions. *Handbook of cognition and emotion*, 98(45-60).
- Ekman, P. y Friesen, W. V.. 1967. Head and body cues in the judgment of emotion: A reformulation. *Perceptual and motor skills*, 24(3 PT 1).
- Fix, E. y Hodges, J. L.. 1989. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 57(3).

- Frank, M. G. y Stennett, J.. 2001. The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of personality and social psychology*, 80(1).
- Freund, Y. y Schapire, R. E.. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1).
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., y De, D.. 2020. Fundamental concepts of convolutional neural network. En: *Recent trends and advances in artificial intelligence and Internet of Things*. Springer.
- Glowinski, D., Camurri, A., Volpe, G., Dael, N., y Scherer, K.. 2008. Technique for automatic emotion recognition by body gesture analysis. En: *2008 IEEE Computer society conference on computer vision and pattern recognition workshops*. IEEE.
- Gonzalez, L.. 2020. Métodos de selección de características. Consultado el 31 de Agosto de 2022. de <https://aprendeia.com/metodos-de-seleccion-de-caracteristicas-machine-learning/>.
- Ho, T. K.. 1995. Random decision forests. En: *Proceedings of 3rd international conference on document analysis and recognition*. IEEE, Vol. 1.
- Hockenbury, D. y Hockenbury, S.. 2007. *Discovering Psychology*. Worth Publishers.
- Ionescu, C., Papava, D., Olaru, V., y Sminchisescu, C.. 2013. Human3.6m: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36(7).
- Janssen, D., Schöllhorn, W. I., Lubienetzki, J., Fölling, K., Kokenge, H., y Davids, K.. 2008. Recognition of emotions in gait patterns by means of artificial neural nets. *Journal of Nonverbal Behavior*, 32(2).
- Kang, G. E. y Gross, M. M.. 2016. The effect of emotion on movement smoothness during gait in healthy young adults. *Journal of biomechanics*, 49(16).
- Kleinsmith, A. y Bianchi-Berthouze, N.. 2012. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1).
- Kozłowska, K., Walker, P., McLean, L., y Carrive, P.. 2015. Fear and the defense cascade: clinical implications and management. *Harvard review of psychiatry*.
- Lee, C.. 2020. Feature importance measures for tree models. Consultado el 31 de Agosto de 2022. de <https://medium.com/the-artificial-impostor/feature-importance-measures-for-tree-models-part-i-47f187c1a2c3>.
- Lee, S.-S., Choi, S. T., y Choi, S.-I.. 2019. Classification of gait type based on deep learning using various sensors with smart insole. *Sensors*, 19(8).
- Li, S., Cui, L., Zhu, C., Li, B., Zhao, N., y Zhu, T.. 2016. Emotion recognition using kinect motion capture data of human gaits. *PeerJ*, 4.
- Lippert, L. et al.. 2006. *Clinical kinesiology and anatomy*. FA Davis.
- Liu, Z., Zhang, H., Chen, Z., Wang, Z., y Ouyang, W.. 2020. Disentangling and unifying graph convolutions for skeleton-based action recognition. En: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.
- Ma, Y., Paterson, H. M., y Pollick, F. E.. 2006. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior research methods*, 38(1).

- Martin, R. C.. 2020. Including maladaptive anger in psychology courses. *Teaching of Psychology*, 47(1).
- Mase, K.. 1991. Recognition of facial expression from optical flow. *IEICE Transactions on Information and Systems*, 74(10).
- Matthews, T., Danese, A., Wertz, J., Odgers, C. L., Ambler, A., Moffitt, T. E., y Arseneault, L.. 2016. Social isolation, loneliness and depression in young adulthood: a behavioural genetic analysis. *Social psychiatry and psychiatric epidemiology*, 51(3).
- Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., y Schulte, D.. 2009. Embodiment of sadness and depression—gait patterns associated with dysphoric mood. *Psychosomatic medicine*, 71(5).
- Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., y Reuter-Lorenz, P. A.. 2005. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4).
- Muro-De-La-Herran, A., Garcia-Zapirain, B., y Mendez-Zorrilla, A.. 2014. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2).
- Narang, S., Best, A., Feng, A., Kang, S.-h., Manocha, D., y Shapiro, A.. 2017. Motion recognition of self and others on realistic 3D avatars. *Computer Animation and Virtual Worlds*, 28(3-4).
- Oaten, M., Stevenson, R. J., Williams, M. A., Rich, A. N., Butko, M., y Case, T. I.. 2018. Moral violations and the experience of disgust and anger. *Frontiers in Behavioral Neuroscience*, 12.
- Omuya, E. O., Okeyo, G. O., y Kimwele, M. W.. 2021. Feature selection for classification using principal component analysis and information gain. *Expert Systems with Applications*, 174.
- Pareek, P. y Thakkar, A.. 2021. A survey on video-based human action recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Review*, 54(3).
- Physiopedia. 2022. Joint Range of Motion During Gait. Consultado el 12 de octubre de 2022. de [https://www.physio-pedia.com/Joint\\_Range\\_of\\_Motion\\_During\\_Gait](https://www.physio-pedia.com/Joint_Range_of_Motion_During_Gait).
- Picard, R. W. y Klein, J.. 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. *Interacting with computers*, 14(2).
- Plutchik, R.. 1984. In search of the basic emotions. *PsycCRITIQUES*, 29(6).
- Prakash, C., Kumar, R., y Mittal, N.. 2018. Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review*, 49(1).
- Randhavane, T., Bhattacharya, U., Kapsaskis, K., Gray, K., Bera, A., y Manocha, D.. 2019. Identifying emotions from walking using affective and deep features. *arXiv:1906.11884*.
- Reddy, S.. 2014. Walk this way: Acting happy can make it so. Consultado el 25 de octubre de 2022. de <https://www.wsj.com/articles/walk-this-way-acting-happy-can-make-it-so-1416263697>.
- Robinson, M. D. y Clore, G. L.. 2002. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin*, 128(6).
- Rodrigo, J. A.. 2017. Análisis de componentes principales (principal component analysis, pca) y t-sne. Consultado el 31 de Agosto de 2022.

- Roether, C. L., Omlor, L., Christensen, A., y Giese, M. A.. 2009a. Critical features for the perception of emotion from gait. *Journal of vision*, 9(6).
- Roether, C. L., Omlor, L., y Giese, M. A.. 2009b. Features in the recognition of emotions from dynamic bodily expression. En: *Dynamics of visual motion processing*..Springer.
- Sapiński, T., Kamińska, D., Pelikant, A., y Anbarjafari, G.. 2019. Emotion recognition from skeletal movements. *Entropy*, 21(7).
- Schmelzer, R.. 2020. Can ai detect your emotion just by how you walk? de <https://www.forbes.com/sites/cognitiveworld/2020/03/29/can-ai-detect-your-emotion-just-by-how-you-walk/?sh=2fc4dafd69de>.
- Sedgley, C.. 2013. The responsibilities of being a physiotherapist. *Tidy's Physiotherapy*.
- Seidenberg, P. H. y Beutler, A. I.. 2008. *The sports medicine resource manual*. Elsevier.
- Shi, L., Zhang, Y., Cheng, J., y Lu, H.. 2019. Skeleton-based action recognition with directed graph neural networks. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Shultz, S. J., Houglum, P. A., y Perrin, D. H.. 2015. *Examination of musculoskeletal injuries*. Human Kinetics.
- Srivastava, S., Rastogi, V., Prakash, C., y Sethi, D.. 2021. Robust approach for emotion classification using gait. En: *International Conference on Innovative Computing and Communications*..Springer.
- Tekscan. 2020. Why is gait analysis important? Consultado el 30 de Agosto de 2022. de <https://www.tekscan.com/blog/medical/why-gait-analysis-important>.
- Tekscan. 2021. The gait cycle: Phases, parameters to evaluate & technology. Consultado el 30 de Agosto de 2022. de <https://www.tekscan.com/blog/medical/gait-cycle-phases-parameters-evaluate-technology>.
- Tilmanne, J. y Dutoit, T.. 2010. Expressive gait synthesis using pca and gaussian modeling. En: *International Conference on Motion in Games*..Springer.
- Trampe, D., Quidbach, J., y Taquet, M.. 2015. Emotions in everyday life. *PloS one*, 10(12).
- Trew, M. y Everett, T.. 2005. *Human movement: an introductory text*. Churchill Livingstone.
- Venture, G., Kadone, H., Zhang, T., Grèzes, J., Berthoz, A., y Hicheur, H.. 2014. Recognizing emotions conveyed by human gait. *International Journal of Social Robotics*, 6(4).
- Wang, K., Peng, X., Yang, J., Meng, D., y Qiao, Y.. 2020. Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Transactions on Image Processing*, 29.
- Xia, S., Wang, C., Chai, J., y Hodgins, J.. 2015. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. on Graphics (TOG)*, 34(4).
- Xu, S., Fang, J., Hu, X., Ngai, E., Guo, Y., Leung, V., Cheng, J., y Hu, B.. 2020. Emotion recognition from gait analyses: Current research and future directions. arXiv:2003.11461.
- Yan, S., Xiong, Y., y Lin, D.. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. En: *Thirty-second AAAI conference on artificial intelligence*.

- Zacharatos, H., Gatzoulis, C., y Chrysanthou, Y. L.. 2014. Automatic emotion recognition based on body movement analysis: a survey. *IEEE computer graphics and applications*, 34(6).
- Zhu, C., Li, P., Zhang, Z., Liu, D., y Luo, W.. 2019. Characteristics of the regulation of the surprise emotion. *Scientific Reports*, 9(1).

## Apéndice A. Listado de características extraídas

### .1. Características temporales

- *abs energy*: Calcula la energía absoluta de la señal.
- *Area under the curve*: Calcula el área bajo la curva de la señal calculada con regla trapezoidal.
- *Autocorrelation*: Calcula la autocorrelación de la señal.
- *Centroid*: Calcula el centroide a lo largo del eje de tiempo.
- *Entropy*: Calcula la entropía de la señal utilizando la entropía de Shannon.
- *Mean absolute diff*: Calcula las diferencias medias absolutas de la señal.
- *Mean diff*: Calcula la media de las diferencias de la señal.
- *Median absolute diff*: Calcula las diferencias absolutas medianas de la señal.
- *Median diff*: Calcula la mediana de las diferencias de la señal.
- *Negative turning points*: Calcula el número de puntos de inflexión negativos de la señal.
- *Neighbourhood peaks*: Calcula el número de picos de una vecindad definida de la señal.
- *Peak to peak distance*: Calcula la distancia pico a pico.
- *Positive turning points*: Calcula el número de puntos de inflexión positivos de la señal.
- *Signal distance*: Calcula la distancia recorrida por la señal.
- *Slope*: Calcula la pendiente de la señal ajustando una ecuación lineal a los datos observados.
- *Sum absolute diff*: Calcula la suma de las diferencias absolutas de la señal.
- *Total energy*: Calcula la energía total de la señal.
- *Zero crossing rate*: Calcula la tasa de cruce por cero de la señal.

### .2. Características estadísticas

- *ECDF (1,...,10)*: Calcula los valores de ECDF (función de distribución acumulada empírica) a lo largo del eje de tiempo.
- *ECDF Percentile (1,2)*: Determina el valor percentil del ECDF.
- *ECDF Percentile Count (1,2)*: Determina la suma acumulativa de muestras que son menores que el percentil.
- *Histogram (1,...,10)*: Calcula el histograma de la señal.
- *Interquartile range*: Calcula el rango intercuartílico de la señal.
- *Kurtosis*: Calcula la curtosis de la señal.
- *Max*: Calcula el valor máximo de la señal.

- *Mean*: Calcula el valor medio de la señal.
- *Mean absolute deviation*: Calcula la desviación absoluta media de la señal.
- *Median*: Calcula la mediana de la señal.
- *Median absolute deviation*: Calcula la desviación absoluta media de la señal.
- *Min*: Calcula el valor mínimo de la señal.
- *Root mean square*: Calcula la raíz cuadrada media de la señal.
- *Skewness*: Calcula la asimetría de la señal.
- *Standard deviation*: Calcula la desviación estándar de la señal.
- *Variance*: Calcula la varianza de la señal.

### .3. Características espectrales

- *spectral (1,...,14)*: Calcula el valor medio de cada frecuencia de espectrograma.
- *Fundamental frequency (1,...,10)*: Calcula la frecuencia fundamental.
- *Human range energy*: Calcula la relación de energía del rango humano dada por la relación entre la energía en la frecuencia de 0,6-2,5 Hz y toda la banda de energía.
- *LPCC (1,...,12)*: Calcula los coeficientes cepstrales de predicción lineal.
- *MFCC (1,...,12)*: Calcula los coeficientes cepstrales de MEL.
- *Max power spectrum*: Calcula la densidad de espectro de potencia máxima.
- *Maximum frequency*: Calcula la frecuencia máxima.
- *Median frequency*: Calcula la frecuencia media.
- *Power bandwidth*: Calcula el ancho de banda de la densidad del espectro de potencia de la señal.
- *Spectral centroid*: Calcula el baricentro del espectro.
- *Spectral decrease*: Calcula la cantidad de disminución de la amplitud del espectro.
- *Spectral distance*: Calcula la distancia espectral de la señal.
- *Spectral entropy*: Calcula la entropía espectral de la señal en función de la transformada de Fourier.
- *Spectral kurtosis*: Calcula la planitud de una distribución alrededor de su valor medio.
- *Spectral positive turning points*: Calcula el número de puntos de inflexión positivos de la señal de magnitud fft
- *Spectral roll-off*: Calcula la frecuencia donde el 95 % de la magnitud de la señal está contenida por debajo de este valor.
- *Spectral roll-on*: Calcula la frecuencia donde el 5 % de la magnitud de la señal está contenida por debajo de este valor.
- *Spectral skewness*: Calcula la asimetría de una distribución alrededor de su valor medio.

- *Spectral slope*: Calcula la pendiente espectral, obtenida por regresión lineal de la amplitud espectral.
- *Spectral spread*: Calcula la dispersión del espectro alrededor de su valor medio.
- *Spectral variation*: Calcula la cantidad de variación del espectro a lo largo del tiempo.
- *Wavelet absolute mean (1,...,10)*: Calcula el valor medio absoluto CWT de cada escala de wavelet.
- *Wavelet energy (1,...,10)*: Calcula la energía CWT de cada escala de wavelet.
- *Wavelet entropy*: Calcula la entropía CWT de la señal.
- *Wavelet standard deviation (1,...,10)*: Calcula el valor estándar de CWT de cada escala de wavelet.
- *Wavelet variance (1,...,10)*: Calcula el valor de varianza de CWT de cada escala de wavelet.

## Apéndice B. Detalle de resultados: Tablas de los diferentes subconjunto de articulaciones

Tabla 12. Subconjunto de 14 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
14 JS	NSL	15,582	KNN	0.571	0.615	0.381	0.667	<b>1.000</b>	0.800	<b>0.600</b>	0.929	<b>0.600</b>	0.333	<b>0.875</b>	0.400	0.596
		15,582	RF	0.714	0.846	0.625	<b>0.833</b>	0.952	<b>0.870</b>	0.400	<b>0.964</b>	0.500	0.778	<b>0.875</b>	0.737	<b>0.708</b>
		15,582	BT	0.714	0.846	0.625	<b>0.833</b>	0.952	<b>0.870</b>	0.200	0.893	0.222	0.556	0.833	0.556	0.568
	UVA	2,500	KNN	0.571	0.615	0.381	0.667	<b>1.000</b>	<b>0.800</b>	<b>0.600</b>	0.929	<b>0.600</b>	0.333	<b>0.875</b>	0.400	0.596
		2,500	RF	<b>0.857</b>	0.923	<b>0.800</b>	<b>0.833</b>	0.952	<b>0.870</b>	0.400	<b>0.964</b>	0.500	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.656</b>
		2,500	BT	0.714	0.846	0.625	<b>0.833</b>	0.952	<b>0.870</b>	0.200	0.893	0.222	0.556	0.833	0.556	0.568
	MDI	1,386	KNN	0.571	0.615	0.381	0.667	<b>1.000</b>	0.800	<b>0.600</b>	0.929	<b>0.600</b>	0.333	<b>0.875</b>	0.400	0.596
		1,386	RF	0.714	0.846	0.625	0.750	0.952	0.818	0.200	<b>0.964</b>	0.286	<b>0.889</b>	0.833	0.762	<b>0.807</b>
		1,386	BT	0.714	0.846	0.625	<b>0.833</b>	0.952	<b>0.870</b>	0.200	0.893	0.222	0.556	0.833	0.556	0.568
	PCA	29	KNN	0.571	0.615	0.381	0.667	<b>1.000</b>	0.800	<b>0.600</b>	0.929	<b>0.600</b>	0.333	<b>0.875</b>	0.400	0.596
		29	RF	0.857	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	0.952	<b>0.870</b>	0.400	<b>0.964</b>	0.500	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.763</b>
		29	BT	0.714	0.846	0.625	<b>0.833</b>	0.952	<b>0.870</b>	0.200	0.893	0.222	0.556	0.833	0.556	0.568

Tabla 13. Subconjunto de 12 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
12 JS	NSL	13,356	KNN	0.429	<b>0.923</b>	0.500	<b>0.917</b>	0.905	<b>0.880</b>	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.567
		13,356	RF	0.714	0.846	0.625	0.750	<b>0.952</b>	0.818	<b>0.400</b>	0.964	0.500	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.712</b>
		13,356	BT	0.714	0.808	0.588	0.833	<b>0.952</b>	0.870	0.200	0.929	0.250	0.556	0.833	0.556	0.574
	UVA	2,500	KNN	0.429	<b>0.923</b>	0.500	<b>0.917</b>	0.905	<b>0.880</b>	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.567
		2,500	RF	<b>0.857</b>	0.808	0.667	0.750	<b>0.952</b>	0.818	0.200	0.964	0.286	0.778	<b>0.875</b>	0.737	<b>0.661</b>
		2,500	BT	0.714	0.808	0.588	0.833	<b>0.952</b>	0.870	0.200	0.929	0.250	0.556	0.833	0.556	0.574
	MDI	1,386	KNN	0.429	<b>0.923</b>	0.500	<b>0.917</b>	0.905	<b>0.880</b>	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.567
		1,386	RF	<b>0.857</b>	0.885	<b>0.750</b>	0.833	<b>0.952</b>	0.870	<b>0.400</b>	<b>1.000</b>	<b>0.571</b>	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.826</b>
		1,386	BT	0.714	0.808	0.588	0.833	<b>0.952</b>	0.870	0.200	0.929	0.250	0.556	0.833	0.556	0.574
	PCA	29	KNN	0.429	<b>0.923</b>	0.500	<b>0.917</b>	0.905	<b>0.880</b>	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.567
		29	RF	0.714	0.808	0.588	0.750	<b>0.952</b>	0.818	0.200	0.964	0.286	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.657</b>
		29	BT	0.714	0.808	0.588	0.833	<b>0.952</b>	<b>0.870</b>	0.200	0.929	0.250	0.556	0.833	0.556	0.574

Tabla 14. Subconjunto de 10 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
10 JS	NSL	11,130	KNN	0.429	<b>0.885</b>	0.462	<b>0.833</b>	0.905	0.833	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.539
		11,130	RF	0.714	<b>0.885</b>	0.667	<b>0.833</b>	0.952	0.870	<b>0.400</b>	0.929	<b>0.444</b>	0.778	<b>0.875</b>	0.737	<b>0.684</b>
		11,130	BT	<b>0.857</b>	0.808	0.667	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.556	<b>0.875</b>	0.588	0.605
	UVA	2,500	KNN	0.429	0.885	0.462	<b>0.833</b>	0.905	0.833	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.539
		2,500	RF	<b>0.857</b>	0.769	0.632	0.750	0.952	0.818	0.200	<b>1.000</b>	0.333	0.778	<b>0.875</b>	0.737	<b>0.775</b>
		2,500	BT	<b>0.857</b>	0.808	0.667	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.556	<b>0.875</b>	0.588	0.605
	MDI	1,386	KNN	0.429	<b>0.885</b>	0.462	<b>0.833</b>	0.905	0.833	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.539
		1,386	RF	<b>0.857</b>	<b>0.885</b>	<b>0.750</b>	0.750	0.952	0.818	<b>0.400</b>	0.964	0.500	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.740</b>
		1,386	BT	0.857	0.808	0.667	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.556	<b>0.875</b>	0.588	0.605
	PCA	29	KNN	0.429	<b>0.885</b>	0.462	<b>0.833</b>	0.905	0.833	<b>0.400</b>	0.786	0.308	0.444	<b>0.875</b>	0.500	0.539
29		RF	<b>0.857</b>	<b>0.808</b>	0.667	0.750	0.952	0.818	0.200	<b>1.000</b>	0.333	<b>0.889</b>	<b>0.875</b>	<b>0.800</b>	<b>0.793</b>	
29		BT	<b>0.857</b>	0.808	0.667	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.200	0.893	0.222	0.556	<b>0.875</b>	0.588	0.605	

Tabla 15. Subconjunto de 6 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
6 JS	NSL	6,678	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.763
		6,678	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	0.833	0.700	0.784
		6,678	BT	<b>0.857</b>	0.846	0.706	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.400	<b>1.000</b>	0.571	<b>0.778</b>	0.833	0.700	<b>0.809</b>
	UVA	2,500	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.763
		2,500	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	0.833	0.700	0.784
		2,500	BT	<b>0.857</b>	0.846	0.706	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.400	<b>1.000</b>	0.571	<b>0.778</b>	0.833	0.700	<b>0.809</b>
MDI	1,386	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.763	
	1,386	RF	<b>0.857</b>	<b>0.923</b>	<b>0.800</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	0.833	0.700	0.784	
	1,386	BT	<b>0.857</b>	0.846	0.706	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.400	<b>1.000</b>	0.571	<b>0.778</b>	0.833	0.700	<b>0.809</b>	
PCA	29	KNN	<b>0.857</b>	0.846	0.706	0.750	<b>1.000</b>	0.857	<b>0.600</b>	0.964	<b>0.667</b>	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	0.763	
	29	RF	0.714	<b>0.923</b>	0.714	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.600</b>	0.929	0.600	<b>0.778</b>	0.833	0.700	0.738	
	29	BT	<b>0.857</b>	0.846	0.706	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	0.400	<b>1.000</b>	0.571	<b>0.778</b>	0.833	0.700	<b>0.809</b>	

Tabla 16. Subconjunto de 5 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
5 JS	NSL	5,565	KNN	0.571	0.808	0.500	0.833	0.952	0.870	0.200	0.821	0.182	0.444	<b>0.875</b>	0.500	0.523
		5,565	RF	<b>0.714</b>	0.846	0.625	0.750	0.952	0.818	0.600	<b>0.964</b>	<b>0.667</b>	0.667	0.833	0.632	0.701
		5,565	BT	0.571	<b>0.962</b>	<b>0.667</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.800</b>	0.857	0.615	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.750</b>
	UVA	2,500	KNN	0.571	0.808	0.500	<b>0.833</b>	0.952	0.870	0.200	0.821	0.182	0.444	<b>0.875</b>	0.500	0.523
		2,500	RF	<b>0.714</b>	0.846	0.625	0.750	0.952	0.818	0.600	0.929	0.600	0.667	<b>0.875</b>	0.667	0.681
		2,500	BT	0.571	0.808	0.500	<b>0.833</b>	0.952	0.870	0.200	0.821	0.182	0.444	<b>0.875</b>	0.500	<b>0.750</b>
	MDI	1,386	KNN	0.571	0.808	0.500	0.833	0.952	0.870	0.200	0.821	0.182	0.444	<b>0.875</b>	0.500	0.523
		1,386	RF	<b>0.714</b>	0.846	0.625	0.750	0.952	0.818	0.600	0.929	0.600	0.667	<b>0.875</b>	0.667	0.681
		1,386	BT	0.571	<b>0.962</b>	<b>0.667</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.800</b>	0.857	0.615	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.750</b>
	PCA	29	KNN	0.571	0.808	0.500	<b>0.833</b>	0.952	0.870	0.200	0.821	0.182	0.444	<b>0.875</b>	0.500	0.523
		29	RF	<b>0.714</b>	0.846	0.625	0.750	0.952	0.818	0.600	0.929	0.600	0.667	<b>0.875</b>	0.667	0.681
		29	BT	0.571	<b>0.962</b>	<b>0.667</b>	<b>0.833</b>	<b>1.000</b>	<b>0.909</b>	<b>0.800</b>	0.857	0.615	<b>0.778</b>	<b>0.875</b>	<b>0.737</b>	<b>0.750</b>

Tabla 17. Subconjunto de 4 articulaciones para datos balanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
4 JS	NSL	4,452	KNN	0.857	0.731	0.600	0.667	<b>1.000</b>	0.800	0.200	1.000	0.333	0.778	0.833	0.700	0.774
		4,452	RF	0.714	0.923	0.714	0.750	0.952	0.818	0.600	0.929	0.600	0.778	0.833	0.700	0.713
		4,452	BT	<b>0.857</b>	<b>0.962</b>	<b>0.857</b>	<b>0.833</b>	<b>0.952</b>	<b>0.870</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.833</b>	<b>0.700</b>	<b>0.788</b>
	UVA	2,500	KNN	0.857	0.731	0.600	0.667	<b>1.000</b>	0.800	0.200	1.000	0.333	0.778	0.833	0.700	0.774
		2,500	RF	0.714	0.885	0.667	0.750	0.952	0.818	0.600	0.964	0.667	0.778	0.833	0.700	0.745
		2,500	BT	<b>0.857</b>	<b>0.962</b>	<b>0.857</b>	<b>0.833</b>	<b>0.952</b>	<b>0.870</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.833</b>	<b>0.700</b>	<b>0.788</b>
	MDI	1,386	KNN	0.857	0.731	0.600	0.667	<b>1.000</b>	0.800	0.200	1.000	0.333	0.778	0.833	0.700	0.774
		1,386	RF	0.714	0.962	0.769	0.833	0.952	0.870	0.600	0.929	0.600	0.778	0.833	0.700	0.728
		1,386	BT	<b>0.857</b>	<b>0.962</b>	<b>0.857</b>	<b>0.833</b>	<b>0.952</b>	<b>0.870</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.833</b>	<b>0.700</b>	<b>0.788</b>
	PCA	29	KNN	0.857	0.731	0.600	0.667	<b>1.000</b>	0.800	0.200	1.000	0.333	0.778	0.833	0.700	0.774
		29	RF	0.714	0.962	0.769	0.833	0.952	0.870	0.600	0.929	0.600	0.778	0.833	0.700	0.745
		29	BT	<b>0.857</b>	<b>0.962</b>	<b>0.857</b>	<b>0.833</b>	<b>0.952</b>	<b>0.870</b>	<b>0.600</b>	<b>0.964</b>	<b>0.667</b>	<b>0.778</b>	<b>0.833</b>	<b>0.700</b>	<b>0.788</b>

Tabla 18. Subconjunto de 14 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
14 JS	NSL	15,582	KNN	0.567	0.888	0.540	0.936	0.937	0.941	0.667	0.870	0.644	0.000	0.988	0.000	0.521
		15,582	RF	0.633	<b>0.993</b>	0.760	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.878	0.804	0.143	0.982	0.182	0.725
		15,582	BT	<b>0.700</b>	0.972	<b>0.764</b>	<b>0.989</b>	<b>0.975</b>	0.984	0.881	<b>0.908</b>	0.813	<b>0.286</b>	0.988	<b>0.364</b>	<b>0.769</b>
	UVA	2,500	KNN	0.500	0.923	0.536	0.947	0.911	0.937	0.714	0.847	0.652	0.000	<b>0.994</b>	0.000	0.526
		2,500	RF	0.633	0.972	0.717	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.881	0.885	0.787	0.286	0.988	0.364	0.757
		2,500	BT	<b>0.700</b>	0.972	<b>0.764</b>	<b>0.989</b>	<b>0.975</b>	0.984	0.881	<b>0.908</b>	0.813	<b>0.286</b>	0.988	<b>0.364</b>	<b>0.769</b>
	MDI	1,386	KNN	0.500	0.923	0.536	0.947	0.911	0.937	0.714	0.847	0.652	0.000	<b>0.994</b>	0.000	0.526
		1,386	RF	0.600	<b>0.993</b>	0.735	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.878	0.816	0.286	0.988	0.364	<b>0.788</b>
		1,386	BT	<b>0.700</b>	0.972	<b>0.764</b>	<b>0.989</b>	<b>0.975</b>	0.984	0.881	<b>0.908</b>	0.813	<b>0.286</b>	0.988	<b>0.364</b>	<b>0.769</b>
	PCA	29	KNN	0.500	0.923	0.536	0.947	0.911	0.937	0.714	0.847	0.652	0.000	<b>0.994</b>	0.000	0.526
		29	RF	0.600	<b>0.993</b>	0.735	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.885	<b>0.825</b>	<b>0.286</b>	0.982	0.333	0.766
		29	BT	<b>0.700</b>	0.972	<b>0.764</b>	<b>0.989</b>	<b>0.975</b>	0.984	0.881	<b>0.908</b>	0.813	<b>0.286</b>	0.988	<b>0.364</b>	<b>0.769</b>

Tabla 19. Subconjunto de 12 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
12 JS	NSL	1,386	KNN	0.656	0.923	0.667	0.968	0.949	0.979	0.714	<b>0.908</b>	0.714	<b>0.364</b>	<b>0.988</b>	<b>0.286</b>	0.704
		1,386	RF	0.735	0.993	0.600	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.825</b>	0.885	<b>0.952</b>	0.333	0.982	<b>0.286</b>	<b>0.766</b>
		1,386	BT	0.764	0.972	<b>0.700</b>	0.979	0.975	0.979	0.804	0.901	0.881	<b>0.364</b>	<b>0.988</b>	<b>0.286</b>	0.765
	UVA	2,500	KNN	0.630	0.951	0.567	0.974	0.949	<b>0.989</b>	0.697	0.878	0.738	0.167	0.976	0.143	0.632
		2,500	RF	0.694	0.986	0.567	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.796	0.870	0.929	0.182	0.982	0.143	0.708
		2,500	BT	0.764	0.972	<b>0.700</b>	0.979	0.975	0.979	0.804	0.901	0.881	<b>0.364</b>	<b>0.988</b>	<b>0.286</b>	<b>0.765</b>
	MDI	1,386	KNN	0.630	0.951	0.567	0.974	0.949	<b>0.989</b>	0.697	0.878	0.738	0.167	0.976	0.143	0.632
		1,386	RF	0.760	0.993	0.633	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.813	0.885	0.929	0.333	0.982	0.286	<b>0.765</b>
		1,386	BT	0.764	0.972	<b>0.700</b>	0.979	0.975	0.979	0.804	0.901	0.881	<b>0.364</b>	<b>0.988</b>	<b>0.286</b>	<b>0.765</b>
	PCA	29	KNN	0.630	0.951	0.567	0.974	0.949	<b>0.989</b>	0.697	0.878	0.738	0.167	0.976	0.143	0.632
		29	RF	<b>0.776</b>	<b>1.000</b>	0.633	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.792	0.878	0.905	0.308	0.976	<b>0.286</b>	0.757
		29	BT	0.764	0.972	<b>0.700</b>	0.979	0.975	0.979	0.804	0.901	0.881	<b>0.364</b>	<b>0.988</b>	<b>0.286</b>	<b>0.765</b>

Tabla 20. Subconjunto de 10 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
10 JS	NSL	11,130	KNN	<b>0.733</b>	0.916	0.688	0.979	0.949	0.968	0.690	<b>0.924</b>	0.716	0.286	<b>0.988</b>	0.364	0.712
		11,130	RF	0.567	<b>0.993</b>	0.708	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.878	0.816	<b>0.429</b>	<b>0.988</b>	<b>0.500</b>	<b>0.812</b>
		11,130	BT	0.667	0.972	0.741	0.979	0.975	0.979	0.857	0.893	0.783	0.286	<b>0.982</b>	0.333	0.733
	UVA	2,500	KNN	0.533	0.944	0.593	<b>0.989</b>	0.949	0.974	0.786	0.878	0.725	0.143	<b>0.988</b>	0.200	0.658
		2,500	RF	0.633	<b>0.993</b>	<b>0.760</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.878	0.816	0.143	<b>0.988</b>	0.200	<b>0.747</b>
		2,500	BT	0.667	0.972	0.741	0.979	0.975	0.979	0.857	0.893	0.783	0.286	<b>0.982</b>	0.333	0.733
	MDI	1,386	KNN	0.533	0.944	0.593	<b>0.989</b>	0.949	0.974	0.786	0.878	0.725	0.143	<b>0.988</b>	0.200	0.658
		1,386	RF	0.600	<b>0.993</b>	0.735	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.885	<b>0.825</b>	0.286	<b>0.982</b>	0.333	<b>0.766</b>
		1,386	BT	0.667	0.972	0.741	0.979	0.975	0.979	0.857	0.893	0.783	0.286	<b>0.982</b>	0.333	0.733
	PCA	29	KNN	0.533	0.944	0.593	<b>0.989</b>	0.949	0.974	0.786	0.878	0.725	0.143	<b>0.988</b>	0.200	0.658
		29	RF	0.600	<b>0.993</b>	0.735	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.870	0.796	0.143	<b>0.982</b>	0.182	0.721
		29	BT	0.667	0.972	0.741	0.979	0.975	0.979	0.857	0.893	0.783	0.286	<b>0.982</b>	0.333	<b>0.733</b>

Tabla 21. Subconjunto de 6 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
6 JS	NSL	6,678	KNN	0.533	0.888	0.516	0.894	0.937	0.918	0.690	0.863	0.652	0.286	0.982	0.333	0.615
		6,678	RF	0.600	0.993	0.735	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.929</b>	0.870	0.796	<b>0.429</b>	<b>0.994</b>	<b>0.545</b>	<b>0.846</b>
		6,678	BT	<b>0.667</b>	0.972	0.741	0.979	<b>0.987</b>	0.984	<b>0.929</b>	<b>0.908</b>	<b>0.839</b>	<b>0.429</b>	0.988	0.500	0.797
	UVA	2,500	KNN	0.433	0.916	0.473	0.883	0.899	0.897	0.786	0.870	0.717	0.143	0.964	0.143	0.559
		2,500	RF	0.567	0.993	0.708	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.855	0.780	0.286	<b>0.994</b>	0.400	<b>0.818</b>
		2,500	BT	<b>0.667</b>	0.972	0.741	0.979	<b>0.987</b>	0.984	<b>0.929</b>	<b>0.908</b>	<b>0.839</b>	<b>0.429</b>	0.988	0.500	0.797
	MDI	1,386	KNN	0.433	0.916	0.473	0.883	0.899	0.897	0.786	0.870	0.717	0.143	0.964	0.143	0.559
		1,386	RF	0.633	<b>1.000</b>	<b>0.776</b>	0.979	<b>0.987</b>	0.984	0.952	0.870	0.808	<b>0.429</b>	0.994	0.545	<b>0.860</b>
		1,386	BT	<b>0.667</b>	0.972	0.741	0.979	<b>0.987</b>	0.984	<b>0.929</b>	<b>0.908</b>	<b>0.839</b>	<b>0.429</b>	0.988	0.500	0.797
	PCA	29	KNN	0.433	0.916	0.473	0.883	0.899	0.897	0.786	0.870	0.717	0.143	0.964	0.143	0.559
		29	RF	0.567	0.993	0.708	0.979	<b>0.987</b>	0.984	<b>0.929</b>	0.855	0.780	0.286	0.988	0.364	0.777
		29	BT	<b>0.667</b>	0.972	0.741	0.979	<b>0.987</b>	0.984	<b>0.929</b>	<b>0.908</b>	<b>0.839</b>	<b>0.429</b>	0.988	0.500	<b>0.797</b>

Tabla 22. Subconjunto de 5 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
5 JS	NSL	5,565	KNN	0.600	0.937	0.632	0.926	0.937	0.935	0.833	0.893	0.769	0.429	0.988	0.500	0.732
		5,565	RF	0.633	0.993	0.760	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	<b>0.952</b>	0.893	<b>0.833</b>	0.286	0.982	0.333	0.770
		5,565	BT	<b>0.733</b>	0.986	<b>0.815</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.901	0.830	0.286	<b>0.994</b>	0.400	<b>0.831</b>
	UVA	2,500	KNN	0.633	0.916	0.623	0.915	0.911	0.920	0.738	<b>0.908</b>	0.729	<b>0.571</b>	0.988	<b>0.615</b>	0.731
		2,500	RF	0.633	<b>1.000</b>	0.776	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.893	0.821	0.429	0.976	0.429	0.788
		2,500	BT	<b>0.733</b>	0.986	<b>0.815</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.901	0.830	0.286	<b>0.994</b>	0.400	<b>0.831</b>
	MDI	1,386	KNN	0.633	0.916	0.623	0.915	0.911	0.920	0.738	<b>0.908</b>	0.729	<b>0.571</b>	0.988	<b>0.615</b>	0.731
		1,386	RF	0.567	0.986	0.694	<b>0.989</b>	0.987	<b>0.989</b>	0.905	0.885	0.800	0.286	0.970	0.286	0.722
		1,386	BT	<b>0.733</b>	0.986	<b>0.815</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.901	0.830	0.286	<b>0.994</b>	0.400	<b>0.831</b>
	PCA	29	KNN	0.633	0.916	0.623	0.915	0.911	0.920	0.738	<b>0.908</b>	0.729	<b>0.571</b>	0.988	<b>0.615</b>	0.731
		29	RF	0.633	0.993	0.760	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.893	0.821	0.286	0.976	0.308	0.752
		29	BT	<b>0.733</b>	0.986	<b>0.815</b>	<b>0.989</b>	<b>0.987</b>	<b>0.989</b>	0.929	0.901	0.830	0.286	<b>0.994</b>	0.400	<b>0.831</b>

Tabla 23. Subconjunto de 4 articulaciones para datos desbalanceados.

Subconjunto de articulaciones	AF-SEL	# de características	CLF	Feliz			Triste			Enojado			Neutral			mAP
				Sen	Spe	F1										
4 JS	NSL	4,452	KNN	0.567	0.944	0.618	0.915	0.924	0.925	0.714	<b>0.939</b>	0.750	<b>0.714</b>	0.922	0.400	0.671
		4,452	RF	0.633	<b>0.986</b>	0.745	<b>0.979</b>	0.962	0.974	0.952	0.908	0.851	0.571	<b>0.994</b>	<b>0.667</b>	0.861
		4,452	BT	<b>0.667</b>	<b>0.986</b>	<b>0.769</b>	<b>0.979</b>	<b>0.975</b>	<b>0.979</b>	<b>0.976</b>	0.916	<b>0.872</b>	0.571	<b>0.994</b>	<b>0.667</b>	<b>0.869</b>
	UVA	2,500	KNN	0.633	0.909	0.613	0.926	0.962	0.946	0.667	0.924	0.700	<b>0.714</b>	0.952	0.500	0.670
		2,500	RF	0.633	0.909	0.613	0.926	0.962	0.946	0.667	0.924	0.700	<b>0.714</b>	0.952	0.500	0.670
		2,500	BT	<b>0.667</b>	<b>0.986</b>	<b>0.769</b>	<b>0.979</b>	<b>0.975</b>	<b>0.979</b>	<b>0.976</b>	0.916	<b>0.872</b>	0.571	<b>0.994</b>	<b>0.667</b>	<b>0.869</b>
	MDI	1,386	KNN	0.633	0.979	0.731	<b>0.979</b>	0.962	0.974	0.929	0.908	0.839	0.571	<b>0.994</b>	0.667	0.670
		1,386	RF	0.633	0.979	0.731	<b>0.979</b>	0.962	0.974	0.929	0.908	0.839	0.571	<b>0.994</b>	<b>0.667</b>	0.849
		1,386	BT	<b>0.667</b>	<b>0.986</b>	<b>0.769</b>	<b>0.979</b>	<b>0.975</b>	<b>0.979</b>	<b>0.976</b>	0.916	<b>0.872</b>	0.571	<b>0.994</b>	<b>0.667</b>	<b>0.869</b>
	PCA	29	KNN	0.633	0.909	0.613	0.926	0.962	0.946	0.667	0.924	0.700	0.714	0.952	0.500	0.670
		29	RF	0.600	<b>0.986</b>	0.720	<b>0.979</b>	0.962	0.974	0.929	0.901	0.830	0.571	0.988	0.615	0.821
		29	BT	<b>0.667</b>	<b>0.986</b>	<b>0.769</b>	<b>0.979</b>	<b>0.975</b>	<b>0.979</b>	<b>0.976</b>	0.916	<b>0.872</b>	0.571	<b>0.994</b>	<b>0.667</b>	<b>0.869</b>

**Tabla 24.** Subconjunto de 14 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
14 JS	STGCN	0.897	0.146	0.103	0.017	0.291
	DGNN	<b>0.966</b>	0.570	0.563	<b>1.000</b>	0.775
	STEP	0.531	0.280	0.268	0.120	0.264
	LSTM	0.587	0.310	0.234	0.152	0.282
	TAEW	0.801	<b>0.833</b>	<b>0.632</b>	0.168	<b>0.779</b>

**Tabla 25.** Subconjunto de 12 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
12 JS	STGCN	0.946	0.198	0.099	0.086	0.339
	DGNN	<b>0.959</b>	0.559	0.368	<b>0.842</b>	0.721
	STEP	0.550	0.281	0.177	0.101	0.292
	LSTM	0.533	0.359	0.288	0.120	0.287
	TAEW	0.785	<b>0.922</b>	<b>0.734</b>	0.178	<b>0.805</b>

**Tabla 26.** Subconjunto de 10 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
10 JS	STGCN	0.905	0.140	0.096	0.045	0.297
	DGNN	<b>0.951</b>	0.399	0.451	<b>1.000</b>	0.700
	STEP	0.448	0.237	0.136	0.117	0.243
	LSTM	0.527	0.386	0.288	0.114	0.290
	TAEW	0.755	<b>0.850</b>	<b>0.726</b>	0.242	<b>0.797</b>

**Tabla 27.** Subconjunto de 9 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
9 JS	STGCN	0.7599	0.0635	0.0621	0.0097	0.2238
	DGNN	<b>0.9457</b>	0.4187	0.4509	<b>1</b>	0.7038
	STEP	0.4311	0.2241	0.1159	0.1167	0.2116
	LSTM	0.5271	0.3858	0.2881	0.1141	0.2897
	TAEW	0.7604	<b>0.8426</b>	<b>0.5484</b>	0.1191	<b>0.7473</b>

**Tabla 28.** Subconjunto de 6 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
6 JS	STGCN	0.766	0.086	0.135	0.016	0.251
	DGNN	<b>0.943</b>	0.362	0.419	<b>0.853</b>	0.681
	STEP	0.424	0.214	0.102	0.114	0.197
	LSTM	0.641	0.277	0.185	0.109	0.265
	TAEW	0.812	<b>0.882</b>	<b>0.717</b>	0.151	<b>0.790</b>

**Tabla 29.** Subconjunto de 5 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
5 JS	STGCN	0.742	0.064	0.059	0.009	0.219
	DGNN	<b>0.901</b>	0.358	0.444	<b>1.000</b>	0.676
	STEP	0.395	0.203	0.151	0.091	0.179
	LSTM	0.598	0.359	0.228	0.125	0.287
	TAEW	0.753	<b>0.903</b>	<b>0.713</b>	0.283	<b>0.783</b>

**Tabla 30.** Subconjunto de 4 articulaciones estado del arte.

Subconjunto de articulaciones	Clasificadores	AP				mAP
		Feliz	Triste	Enojado	Neutral	
4 JS	STGCN	0.7221	0.0644	0.0498	0.0181	0.2187
	DGNN	<b>0.924</b>	0.315	0.230	<b>1.000</b>	0.617
	STEP	0.231	0.142	0.144	0.053	0.152
	LSTM	0.554	0.315	0.245	0.147	0.278
	TAEW	0.817	<b>0.899</b>	<b>0.701</b>	0.279	<b>0.806</b>