



Escuela Técnica Superior de Ingeniería – ICAI

APLICACIÓN DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO PARA LA CARACTERIZACIÓN DE LA SEÑAL DEL RITMO CARDÍACO

Autor: Alfonso Serrano-Suñer García de Leyaristy
Director: Antonio Muñoz San Roque

Índice de la memoria

Capítulo 1. Introducción	1
1.1 Motivación del proyecto	1
1.2 Enfoque del proyecto	2
1.3 Objetivos específicos del estudio	3
Capítulo 2. Estado de la Cuestión	6
2.1 Estado actual de la tecnología portátil	6
2.2 Aplicaciones de la tecnología wearable al campo de la salud	7
2.3 Revisión de las técnicas de aprendizaje automático más empleadas para el análisis de variables biométricas	11
2.4 Modelos de aprendizaje automático empleados en el proyecto	13
2.4.1 Modelos ocultos de Markov	13
2.4.2 Clustering de series temporales	18
2.4.3 Autoencoders	21
Capítulo 3. Metodología	25
3.1 Reclutamiento de voluntarios y toma de medidas	25
3.1.1 Selección de voluntarios	25
3.1.2 Procedimiento para la extracción de datos	26
3.2 Preprocesamiento básico y Análisis Exploratorio	27
3.3 Ajuste de modelos y análisis de resultados	28
Capítulo 4. Preprocesamiento y EDA	30
4.1 Preprocesamiento básico de los datos en bruto	30
4.2 Análisis Exploratorio de los Datos (EDA)	33
Capítulo 5. Desarrollo de modelos	38
5.1 Modelos ocultos de Markov	39
5.1.1 Caracterización HMM de la serie temporal en bruto	40
5.2 Autoencoders	53
5.3 Clustering	66
Capítulo 6. Conclusiones	80
6.1 Limitaciones en la metodología y alcance del estudio	80

6.1.1	<i>Discusión de limitaciones en la muestra de voluntarios.....</i>	80
6.1.2	<i>Discusión de limitaciones en el procedimiento de adquisición de datos.....</i>	81
6.2	Valoración crítica de los resultados	82
6.3	Conclusión.....	83
Capítulo 7.	<i>Bibliografía.....</i>	85
ANEXO I:	<i>CARACTERÍSTICAS TÉCNICAS DE LAS PULSERAS EMPLEADAS....</i>	93
ANEXO II:	<i>CÓDIGO FUENTE.....</i>	95

Índice de figuras

Figura 1: Diagrama genérico de un HMM (Mariano et al., 2015)	14
Figura 2: Ejemplo de un autoencoder básico de arquitectura simétrica.	21
Figura 3: Metodología implementada.....	25
Figura 4: Diagrama de modelización.....	28
Figura 5: Flujo de preprocesamiento básico	31
Figura 6: Fragmento de la serie temporal de ritmo cardiaco para el sujeto ID1	33
Figura 7: Fragmento con regiones correspondientes al sueño sombreadas.....	34
Figura 8: fragmentos de la serie de ritmo cardiaco para otros sujetos	35
Figura 9: Comparativa de distribuciones de ritmo cardiaco dentro y fuera del periodo de sueño.....	35
Figura 10: Flujo de modelización y análisis ampliado	38
Figura 11: Metodología específica HMM	41
Figura 12: evolución de la verosimilitud logarítmica en función del número de estados. ..	41
Figura 13: Evolución de probabilidades a posterior de cada uno de los cinco estados (noche del 16/9/22).	42
Figura 14: Frecuencia cardiaca para la noche del 16 de septiembre.	42
Figura 15: Distribución de los valores de ritmo cardiaco para cada estado (sujeto ID1)...	43
Figura 16: Verosimilitud logarítmica en ID1 para segmentos no deslizantes de 4 horas de duración.	45
Figura 17: Detalle de verosimilitud logarítmica (ID1) para segmentos no deslizantes de 4 horas.....	45
Figura 18: Verosimilitud logarítmica en ID1 para segmentos no deslizantes de 1 hora de duración.	47
Figura 19: Caracterización de estados ocultos en ID2	49
Figura 20: Caracterización de estados ocultos en ID3	49
Figura 21: Caracterización de estados ocultos en ID6	49

Figura 22: Caracterización de estados ocultos en ID7	50
Figura 23: Caracterización de estados ocultos en ID8	50
Figura 24: caracterización de estados ocultos en ID9	50
Figura 25: Posible cambio tendencial en ID8.....	53
Figura 26: Metodología específica autoencoders.	54
Figura 27: Arquitectura autoencoder 'feedforward'	55
Figura 28: Evolución del error de reconstrucción durante el entrenamiento.	58
Figura 29: Registro de ritmo cardiaco de ID1 limitado al periodo nocturno (fechas atípicas sombreadas).....	59
Figura 30: Evolución del error de reconstrucción durante el entrenamiento.	60
Figura 31: Registro de ID1 limitado al periodo nocturno (fechas atípicas sombreadas). ...	61
Figura 32: Registro para ID1 limitado al periodo nocturno, con segmentos atípicos sombreados.	62
Figura 33: Registro de ID8 limitado al periodo nocturno (fechas atípicas sombreadas). ...	63
Figura 34: Registro de ID8 limitado al periodo nocturno (fechas atípicas sombreadas). ...	64
Figura 35: Metodología completa clustering.	69
Figura 36: Evolución de la suma de cuadrados intra-cluster según el número de clusters. 70	
Figura 37: Resumen de clustering para ID1	72
Figura 38: Resumen de clustering para ID2	75
Figura 39: Resumen de clustering para ID3	76
Figura 40: Resumen de clustering para ID6	77
Figura 41: Resumen de clustering para ID7	77
Figura 42: Resumen de clustering para ID8	78
Figura 43: Resumen de clustering para ID9.....	79

Índice de tablas

Tabla 1: Cuadro resumen de técnicas ML aplicadas a variables biométricas	13
Tabla 2: Anomalías identificadas mediante verosimilitud logarítmica de segmentos.	52
Tabla 3: Identificación de anomalías mediante autoencoders, comparativa con HMM.	65
Tabla 4: Resumen de resultados tras la aplicación de clustering.	70
Tabla 5: Centroides de los clusters	73
Tabla 6: Proporción de observaciones de cada cluster pertenecientes a cada estado oculto HMM.	73
Tabla 7: Proporción de anomalías en cada cluster.	75
Tabla 8: Comparativa de prestaciones técnicas de las pulseras.....	94

Capítulo 1. INTRODUCCIÓN

1.1 MOTIVACIÓN DEL PROYECTO

El envejecimiento demográfico, el aumento de la obesidad, el sedentarismo o el incremento en la incidencia de enfermedades crónicas representan una preocupación creciente en el ámbito de la salud pública a nivel mundial (Arocha Rodulfo, 2019; Eatock et al., 2017; James et al., 2001). Estas tendencias, que se mantienen al alza desde hace algunos años en las regiones desarrolladas del planeta, contribuyen directamente al aumento en la prevalencia de desórdenes cardiovasculares (Ponikowski et al., 2014) y metabólicos; los primeros se erigen además como unas de las causas principales de mortalidad y morbilidad en la población (Lippi and Plebani, 2013). El envejecimiento y la obesidad también se ven asociados a una mayor vulnerabilidad ante enfermedades infecciosas, ya sean estacionales (como la gripe) o vinculadas a episodios pandémicos similares al experimentado recientemente con el Covid-19 (Kang and Jung, 2020; Yanai, 2020), donde la mayor parte de los fallecimientos se han concentrado en estos grupos de población.

En este contexto, la monitorización del estado de salud de los individuos cobra una creciente importancia: la detección temprana y el seguimiento de dichas condiciones es esencial para prevenir complicaciones y mejorar la calidad de vida de las personas afectadas, así como para mitigar su coste económico global e impacto en los sistemas sanitarios. Del mismo modo, una monitorización adecuada podría permitir una identificación más rápida de cualquier indicio de deterioro en el estado global de salud de los pacientes (particularmente aquellos de mayor edad), redundando en un diagnóstico anticipado y un mejor tratamiento.

Siendo esta la situación, y en paralelo a este fenómeno, podemos observar una fuerte tendencia hacia la integración de la tecnología en nuestra vida diaria. Un modo particular en el que se está manifestando dicha tendencia lo encontramos en la aparición y proliferación de dispositivos *wearable*; se trata de dispositivos comerciales orientados a realizar un seguimiento de la salud y forma física de sus usuarios (su nombre en inglés, que se podría

traducir como ‘llevables’ o ‘portátiles’, hace referencia a su diseño orientado a ser portado sobre el cuerpo como una prenda o adorno más). Estos dispositivos, entre los que se incluyen los relojes inteligentes y las pulseras de actividad, integran una multiplicidad de sensores (acelerómetros, giroscopios, sensores basados en fotoplethismografía...) para la monitorización de diversas variables biométricas, como el número de pasos, el nivel de actividad física o el ritmo cardíaco, todo ello en tiempo real y a un coste relativamente bajo.

La tendencia hacia la adopción e incorporación de dispositivos *wearable* en la rutina diaria y su creciente mejora en cuanto a precisión y accesibilidad económica han abierto un gran potencial de aplicación en el ámbito de la salud y el análisis biométrico. Dado que, como comentábamos, permiten el registro continuo de diversos datos fisiológicos, ofrecen una oportunidad única para analizar patrones y detectar anomalías en el comportamiento de diversas variables biométricas, potencialmente correlacionadas con posibles cambios o alteraciones en la salud cardiovascular y metabólica, nivel de estrés o calidad de sueño de los individuos portadores del dispositivo.

Por otro lado, la aplicación de técnicas de aprendizaje automático en el análisis de datos biométricos ha demostrado ser una herramienta efectiva y precisa para el reconocimiento de patrones y la identificación de eventos de salud. La combinación de estos avances tecnológicos con el análisis de datos biométricos proporciona un marco de trabajo prometedor para el desarrollo de soluciones innovadoras en la identificación de eventos de salud, ya sean episodios febriles asociados a enfermedades víricas estacionales, alteraciones conductuales, o signos de alguna condición cardiovascular o desorden metabólico, lo cual podría permitir una detección temprana y mejor seguimiento de enfermedades de diverso tipo, así como una evaluación más precisa del estado global de salud del individuo.

1.2 ENFOQUE DEL PROYECTO

Dentro de este contexto, el presente proyecto buscará caracterizar el registro diario del ritmo cardíaco, mediante el uso de pulseras de actividad y la aplicación de técnicas de aprendizaje automático. Como se verá más adelante, se pondrá el foco exclusivamente en el periodo

nocturno por razones de homogeneidad. La finalidad última del trabajo será el desarrollo de modelos que permitan, mediante esta caracterización, identificar condiciones de comportamiento normal de variables biométricas vinculadas con la salud cardiovascular y metabólica (como se ha dicho, serán fundamentalmente la frecuencia cardiaca y diversas variables derivadas de ésta), así como patrones anómalos en relación con los parámetros habituales de cada individuo.

El marco de análisis propuesto en este trabajo podría potencialmente facilitar o servir de base para la creación de aplicaciones o servicios de seguimiento del estado de salud y detección temprana de enfermedades, basados en la monitorización continua de datos fisiológicos proporcionados por dispositivos *wearable*. Esta aproximación ofrece una solución accesible, no invasiva y de bajo coste económico para el seguimiento y detección de condiciones de salud cardiovascular y metabólica en la población.

1.3 OBJETIVOS ESPECÍFICOS DEL ESTUDIO

El enfoque descrito en el apartado anterior se puede desglosar y concretar en un objetivo principal y dos secundarios. En síntesis, el objetivo principal al que responde este proyecto y hacia el que se ha orientado el trabajo desarrollado a lo largo de estos meses es el siguiente:

1. A partir de la serie temporal de registros de frecuencia cardiaca tomados por pulseras de actividad, se aplicarán diversos algoritmos de aprendizaje automático no supervisado con el fin de caracterizar y modelar el comportamiento normal del ritmo cardíaco, identificar anomalías en la serie temporal y evaluar la potencial correlación de estas anomalías con eventos de salud o metabólicos diversos, como la disrupción del ritmo circadiano, episodios febriles, malestar general, picos de estrés o cambios sostenidos en la salud metabólica/cardiovascular del sujeto. Si bien el análisis partirá de la serie completa, la exploración de los datos (como se detallará más adelante) llevará a la investigación a centrarse exclusivamente en las muestras tomadas durante los periodos de sueño.

Ello requirió de la colaboración con voluntarios y la elección de los dispositivos más adecuados para la toma de mediciones. Aunque no abordaremos la interpretación médica precisa de las posibles anomalías detectadas, la consecución de este objetivo principal (esto es, caracterización exitosa del comportamiento normal, identificación de anomalías, y evaluación de su correlación con eventos de salud) demostrará la viabilidad del enfoque propuesto. Junto a este objetivo principal, el trabajo de investigación buscará cubrir también los siguientes objetivos secundarios:

2. Examinar y comprender el comportamiento del ritmo cardíaco y sus variables asociadas en condiciones normales y durante la aparición de anomalías. Por tanto, no buscaremos solo modelar el comportamiento de la serie, sino también entender en la medida de lo posible qué características exhibe la serie temporal cuando los modelos entrenados identifican una determinada secuencia como anómala, a fin de poder interpretar las predicciones.
3. Recopilar, limpiar y preprocesar las mediciones, lo que conllevará la creación de una base de datos anonimizada que pueda ser utilizada en futuras investigaciones.

En las sucesivas secciones de este informe se irá exponiendo el procedimiento seguido para alcanzar estos objetivos. Antes de avanzar, es preciso realizar una última aclaración. Como se ha dicho, el fin último que persigue este trabajo es fundamentalmente el de verificar la utilidad de estos modelos como base para herramientas que pueda ponerse a disposición de profesionales de la medicina, para facilitar un mejor seguimiento del estado de salud del sujeto y, potencialmente, la detección temprana de enfermedades o trastornos cardiovasculares/metabólicos. Sin embargo, se debe remarcar que el diagnóstico médico preciso de las anomalías identificadas excede el alcance del proyecto: únicamente se buscará constatar si los cambios en la serie temporal de registro de ritmo cardíaco respecto a los patrones de comportamiento caracterizados como típicos se correlacionan con eventos de salud experimentados por el propio sujeto, como (por ejemplo) episodios catarrales o gripales.

Es importante clarificar este punto, a fin de que el alcance del proyecto quede adecuadamente delimitado y se puedan calibrar correctamente las expectativas respecto a los resultados de la investigación.

Capítulo 2. ESTADO DE LA CUESTIÓN

2.1 ESTADO ACTUAL DE LA TECNOLOGÍA PORTÁTIL

La progresiva mejora en la precisión de las mediciones realizadas por dispositivos *wearable* o aplicaciones para móvil está posibilitando el acceso y seguimiento ininterrumpido de parámetros como el ritmo cardíaco, las horas de sueño y calidad del mismo, la actividad diaria, el recuento de pasos o la saturación de oxígeno en sangre, todo ello de forma no invasiva, económicamente accesible y con un nivel de fiabilidad creciente (Altini and Amft, 2016; Altini and Dunne, 2021).

Se ha de señalar, sin embargo, que aún no están exentos de limitaciones. Una revisión sistemática publicada en 2020 (Cosoli et al., 2020) evaluó el rendimiento de algunos de los modelos más empleados en la investigación en función de su precisión a la hora de medir cuatro parámetros fundamentales: recuento de pasos, energía consumida por el usuario, frecuencia cardíaca y sueño. Para evaluar la precisión de estos dispositivos, los autores se apoyaron en resultados de diversos estudios de la literatura primaria, utilizando dos parámetros principales: el error porcentual absoluto (APE) respecto a mediciones con equipos de nivel clínico, y el coeficiente de correlación (r) de las mediciones tomadas por modelos *wearable* con las de nivel clínico.

Bajo estos criterios, los autores hallaron una significativa variabilidad en la precisión de estos dispositivos, constatando que en líneas generales algunos modelos funcionan mejor que otros dependiendo del parámetro en cuestión que se esté midiendo. Para el recuento de pasos, a fecha de publicación del estudio en torno a la mitad de los dispositivos considerados exhibían tasas de error inferiores a 5% (una tasa de error del 5% suele considerarse aceptable), con algunos de los modelos de las marcas Apple o Fitbit alcanzando coeficientes de correlación bastante elevados (de en torno a un 99%).

Respecto a la frecuencia cardíaca, aunque las cintas pectorales alcanzaron las precisiones más altas, los dispositivos de muñeca (pulseras de actividad o relojes inteligentes) mostraron resultados bastante aceptables tanto en términos de correlación como de error porcentual absoluto. Y en cuanto al registro de sueño, los resultados también tendían a ser bastante aceptables. Con todo, los autores lamentaban la ausencia de procedimientos estandarizados para la validación de los dispositivos, lo cual dificulta la comparabilidad entre estudios. Pero aun considerando las limitaciones existentes, estudios posteriores constatan una tendencia hacia la mejora en la precisión, como comentábamos anteriormente (Altini and Dunne, 2021).

En cualquier caso, encontramos ya en el mercado marcas y modelos que han podido emplearse con éxito en la detección o seguimiento de desórdenes y enfermedades vinculados con diversos campos de salud, incluyendo la salud cardiovascular, metabólica, de sueño, mental y neurológica (Dunn et al., 2018); en el siguiente apartado se desglosará con más detalle este punto. También existe un importante desarrollo en la investigación en torno a algoritmos orientados al procesamiento de los datos brutos de los sensores, a fin de extraer indicadores relevantes relacionados con el estado físico y de salud de los usuarios, como puede ser la detección de pasos (Witt et al., 2019).

2.2 APLICACIONES DE LA TECNOLOGÍA WEARABLE AL CAMPO DE LA SALUD

Numerosas publicaciones han puesto de manifiesto la relevancia de dispositivos ‘wearable’ de nivel comercial para abordar problemas dentro de este campo de conocimiento. En términos generales, podemos identificar dos áreas de estudio hacia donde mayoritariamente se ha dirigido la investigación (Shin et al., 2019): por un lado, la monitorización del estado de salud a partir de variables biométricas; y por otro lado, la evaluación del impacto que estos dispositivos ejercen sobre la conducta de sus usuarios, particularmente de cara a determinar si su mero uso incentiva la actividad física. De estas dos ramas, la que resulta de mayor interés para nuestro trabajo es la primera. La investigación desarrollada dentro de este

área ha puesto de manifiesto el potencial de estos dispositivos a la hora de abordar objetivos vinculados con el seguimiento del estado de salud y su potencial deterioro, o la detección de patologías o eventos de salud de diverso tipo.

Así, un estudio publicado en 2018 que analizaba datos de actividad y ritmo cardiaco de 233 voluntarios normales tomados mediante un dispositivo Fitbit Charge durante un periodo de entre 2 y 6 días, integrándolos con información tomada de cuestionarios y exámenes médicos, demostró la relación entre los parámetros biométricos medidos por la pulsera y marcadores clínicos vinculados con el riesgo de padecer desórdenes cardiovasculares y metabólicos, la remodelación cardiaca y la abundancia de esfingolípidos (Lim et al., 2018).

En particular, el estudio evidenció que el ritmo cardiaco en reposo (RHR) medido por el dispositivo estaba asociado de forma estadísticamente significativa con la mayoría de los principales marcadores clínicos asociados al síndrome metabólico y cardiovascular, como la presión arterial diastólica, el nivel de glucosa en sangre en ayunas, o el nivel de triglicéridos. También halló que el número de pasos diarios – derivado de la pulsera – resultaba ser un buen predictor del remodelado cardiaco inducido por el ejercicio, y estaba significativamente asociado con la abundancia de los principales esfingolípidos circulantes en el organismo. Esto concuerda con las conclusiones de una reciente revisión sistemática de la literatura primaria emitida por el Colegio Americano de Medicina del Deporte, según la cual la evidencia disponible apoya la existencia de una relación inversa entre la cifra de pasos diarios (muy fácil de registrar mediante un dispositivo *wearable*) y el riesgo de eventos cardiovasculares, diabetes y mortalidad en general (Kraus et al., 2019).

Un estudio previo a los anteriores (Tison et al., 2017) aporta evidencia del potencial de aplicación de algoritmos de aprendizaje profundo en la detección de trastornos cardiorrespiratorios a partir de parámetros biométricos básicos medibles por la mayoría de dispositivos comerciales; concretamente, los investigadores hallaron que redes neuronales profundas demostraron ser capaces de detectar apnea de sueño e hipertensión con una tasa de acierto del 90% y 82% respectivamente, recibiendo como únicas entradas la frecuencia cardiaca y el recuento de pasos diarios (ambas métricas tomadas por el Apple Watch).

También hay estudios que han explorado la utilidad de estos dispositivos para la monitorización y detección de fibrilación auricular, alcanzándose tasas de acierto del 95 % (en datos *out-of-sample*) al emplearse métricas basadas en la variabilidad entre intervalos de pulsación y en la calidad de la señal de la forma de onda, extraídas a partir de las mediciones del sensor de fotopletimografía (PPG) de la Samsung Simband (Nemati et al., 2016).

Más recientemente y a raíz del creciente uso de estos dispositivos por parte de la población, se han abierto también líneas de investigación para explorar su utilidad en la detección y seguimiento de las ondas epidémicas estacionales vinculadas con la gripe. En concreto, en un estudio reciente se observó una clara correlación entre los datos de incidencia de gripe registrados por los CDC (Centros para el Control y Prevención de Enfermedades) en Estados Unidos y el porcentaje de medidas anómalas de ritmo cardíaco y sueño entre los voluntarios (Radin et al., 2020).

También se ha evaluado su potencial utilidad en el seguimiento de la incidencia de la enfermedad Covid-19 durante la reciente pandemia (Quer et al., 2021), así como el impacto fisiológico de dicha enfermedad, de nuevo usando métricas fácilmente adquiribles con una pulsera inteligente o dispositivos similares (Radin et al., 2021). Concretamente, en este último estudio se pudieron observar cambios reconocibles en el ritmo cardíaco en reposo, la duración del sueño en horas y el recuento de pasos. Estos cambios no son exclusivos al Covid, en general procesos catarrales o gripales suelen tener impactos parecidos fruto del estrés metabólico al que es sometido el organismo (por ejemplo, al elevarse la temperatura durante la fiebre), que se suelen concretar en una subida del ritmo cardíaco en reposo y una mayor duración del sueño (Radin et al., 2020).

La utilidad de métricas específicas como la variabilidad del ritmo cardíaco (HRV, por sus siglas en inglés) para el seguimiento del estado físico y de salud también se ve muy respaldada por la literatura. Dicha métrica, directamente vinculada con la actividad del sistema nervioso autónomo (en especial con la actividad parasimpática), es particularmente relevante como medida de nuestra capacidad de respuesta y adaptación a diversos estresores, así como del impacto de dichos estresores en nuestro organismo, ya sean agudos y de corta

duración (reflejado en los cambios a corto plazo de la HRV), o más prolongados en el tiempo (reflejado en las tendencias a largo plazo de dicha medida), en especial cuando combinamos dicha métrica con la de ritmo cardiaco en reposo (Altini and Plews, 2021). También hay evidencias de que podría emplearse como predictor de la severidad y evolución de enfermedades graves como la sepsis (Arbo et al., 2020). Es importante recalcar que gran parte de la información de interés aportada por esta métrica se deriva sobre todo de los cambios de ésta a lo largo del tiempo respecto a la línea basal de cada sujeto individual, y no tanto de su divergencia o similitud con respecto a la de otros sujetos o la media de la población.

Las medidas biométricas tomadas por dispositivos *wearables* también tienen interés para la estimación de la edad biológica. Varios estudios han puesto particularmente de manifiesto la relación entre la HRV (variabilidad de ritmo cardiaco) y la edad del organismo, así como su potencial utilidad como marcador de la edad biológica (Bashkirtsev et al., 2021; Choi et al., 2020; Garavaglia et al., 2021; Lee and Byun, 2023), si bien estos se apoyan en HRV derivada de ECG clínicos dada la todavía reducida accesibilidad y fiabilidad de las medidas de HRV proporcionadas por dispositivos comerciales. La relevancia de la edad biológica radica en que es un indicador muy superior del estado de salud y envejecimiento del organismo que la edad cronológica, permitiendo mejores estimaciones del riesgo de padecer una amplia variedad de condiciones asociadas con el envejecimiento, desde la demencia (Wu et al., 2021) hasta la mortalidad y la pérdida funcional, posibilitando además una actuación temprana y preventiva antes de que se manifiesten clínicamente (Guerville et al., 2019).

Por último, también se ha verificado (Sano and Picard, 2013) la utilidad de mediciones de sensores *wearable* para la estimación del nivel de estrés (de modo particular si se combina con métricas de tipo más conductual derivadas del uso del móvil por parte de los sujetos), si bien todavía existen dificultades cuando se busca inferir el estrés exclusivamente a partir de dispositivos comerciales (Chalmers et al., 2021).

En definitiva, si bien aún hay limitaciones que conviene tener presente (la fiabilidad de las medidas, pese a haberse incrementado sustancialmente en los últimos años, para ciertas

variables aún no llega en la mayoría de los dispositivos al nivel de precisión que puede alcanzarse con pruebas clínicas), el estado del arte relativo a este campo de la investigación pone de manifiesto un importante potencial tanto a nivel médico-clínico como, más generalmente, para registrar, evaluar y verificar alteraciones en el estado de salud, forma física y estilo de vida de los usuarios, con múltiples aplicaciones tanto preventivas como de seguimiento.

2.3 REVISIÓN DE LAS TÉCNICAS DE APRENDIZAJE AUTOMÁTICO MÁS EMPLEADAS PARA EL ANÁLISIS DE VARIABLES BIOMÉTRICAS

Vista esta revisión, a continuación se expone un cuadro-resumen sintetizando algunos de los problemas abordados en los diversos estudios revisados, y los algoritmos empleados para generar los modelos con los que se ha tratado de dar respuesta a dichos problemas. Se trata únicamente de una exposición de ejemplos, no pretende ser un resumen exhaustivo.

<i>Objetivo</i>	<i>Modelos empleados</i>	<i>Comentario</i>	<i>Referencias</i>
<i>Abundancia de esfingolípidos circulantes</i>	Regresión lineal	Se identificaron 12 esfingolípidos significativamente asociados con la cifra de pasos diarios ($p < 0,05$)	(Lim et al., 2018)
<i>Apnea de sueño</i>	Arquitectura híbrida <i>deep learning</i> (Red convolucional 1D + LSTM)	Tasa de acierto del 90%	(Tison et al., 2017)
<i>Caracterización del nivel de actividad</i>	Clustering <i>k-means</i>	Diferencias estadísticamente significativas en cuanto a periodos de máxima actividad, edad media, y hora de inicio de sueño	(Lim et al., 2018)
<i>Edad biológica</i>	CNN Perceptrón multicapa	Tasas de acierto moderadas para predicciones basadas en intervalos entre pulsaciones, derivados de mediciones ECG.	(Lee and Byun, 2023)

<i>Estrés</i>	SVM KNN PCA	Mejores resultados obtenidos con SVM de kernel RBF, con tasa de acierto del 87.5%	(Sano and Picard, 2013)
<i>Estimación de incidencia de gripe</i>	Regresión lineal, modelos binomiales	Se modelaron los recuentos de casos de gripe y enfermedades similares con un modelo binomial negativo que basados en datos de la tasa de gripe de los CDC con 3 semanas de retraso (modelo nulo) y la proporción de usuarios semanales de Fitbit con ritmo cardiaco en reposo elevada y mayor duración del sueño por encima de un cierto umbral (modelo completo). Mejoras significativas (6.3–32.9%) sobre los modelos base al incorporar datos Fitbit.	(Radin et al., 2020)
<i>Estado general de salud a partir de marcadores cardiovasculares y metabólicos</i>	Modelos ocultos de Markov (HMM)	HMM modelado a partir de parámetros como la presión arterial sistólica, índice de masa corporal y otros. Se logró dar una interpretación médica a cada uno de los estados ocultos, asociándolos a 3 niveles distintos de riesgo.	(Kawamoto et al., 2013)
<i>Fibrilación auricular</i>	Regresión logística	Con validación cruzada <i>10-fold</i> , se alcanzó una precisión del 95% en los datos <i>out-of-sample</i> , con una sensibilidad del 97%, especificidad del 94% y área bajo la curva ROC de 0.99.	(Nemati et al., 2016)
<i>Hipertensión</i>	Arquitectura híbrida <i>deep learning</i> (Conv. 1D + LSTM)	Tasa de acierto del 82%	(Tison et al., 2017)
<i>Reconocimiento de actividad</i>	Autoencoders HMM	Reconocimiento de diversas actividades (caminar, trabajar ante el ordenador, conducir) a partir de datos GPS y acelerómetros.	(Karaman et al., 2014; Li et al., 2015)

<i>Remodelación cardiaca</i>	Regresión lineal	El recuento de pasos diarios resultó ser un predictor significativo de diversos parámetros vinculados con la remodelación cardiaca asociada al ejercicio.	(Lim et al., 2018)
<i>Severidad de sepsis</i>	Regresión lineal	Gravedad de la enfermedad (definida por el subtipo de sepsis) significativamente correlacionada con diversas métricas derivadas de la HRV.	(Arbo et al., 2020)
<i>Síndrome metabólico cardiovascular</i> y	Regresión logística	Tanto el ritmo cardiaco en reposo como el recuento de pasos diarios resultaron ser predictores significativos de varios marcadores de desorden metabólico y cardiovascular.	(Lim et al., 2018)

Tabla 1: Cuadro resumen de técnicas ML aplicadas a variables biométricas

Este revisión resultó de valor para orientar la elección de los modelos de aprendizaje automático a emplear en el proyecto, que se describen en la siguiente sección.

2.4 MODELOS DE APRENDIZAJE AUTOMÁTICO EMPLEADOS EN EL PROYECTO

Teniendo en cuenta que seguiremos en todo momento un enfoque no supervisado de caracterización, los modelos que se emplearan para este trabajo se describen a continuación.

2.4.1 MODELOS OCULTOS DE MARKOV

Los modelos ocultos de Markov (en adelante HMM, siglas de *Hidden Markov Models*) son modelos estadísticos en los que se parte de la suposición de que el sistema modelado se comporta siguiendo un proceso estocástico subyacente e inobservable (proceso de Markov) que genera una secuencia de sucesos aleatorios (estados). La propiedad fundamental de los procesos de Markov es que el siguiente estado de la secuencia sólo depende del estado actual,

y no de los estados anteriores; además, estas probabilidades de transición entre estados no varían con el tiempo (aunque en el caso de los HMM de entrada-salida podemos modelar variables de entrada que pueden afectar a estas probabilidades de transición).

Los estados ocultos, denominados también estados latentes, no se pueden observar directamente, pero los sistemas representados también cuentan con una o varias variables observables emitidas por el propio sistema, donde la distribución de probabilidades de emisión de dichas variables varía en función del estado en el que se encuentre el sistema. Estas probabilidades de emisión pueden ser cualquier distribución, donde los parámetros que la definirían (θ) estarían condicionados por el estado actual del sistema.

Los HMM, por tanto, vendrían definidos en términos paramétricos por el número de estados ocultos del sistema (un hiperparámetro pre-establecido antes de ajustar el modelo), el vector de probabilidades iniciales π , la matriz de probabilidades de transición entre estados A , y los parámetros θ que determinan la distribución de probabilidad de emisión del sistema, que varían en función de su estado actual. Los parámetros $\lambda = (\pi, A, \theta)$ caracterizan completamente el comportamiento y evolución del sistema modelado.

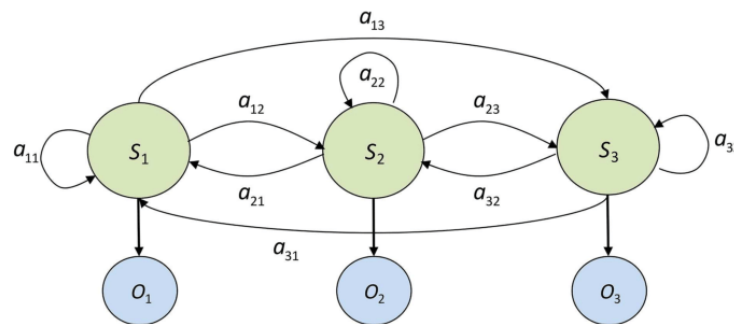


Figura 1: Diagrama genérico de un HMM (Mariano et al., 2015)

El algoritmo Baum-Welch, desarrollado por Leonard E. Baum y otros colaboradores, es un caso especial del algoritmo de maximización de expectativas (EM, siglas de *Expectation-Maximization*). A partir de una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, este

algoritmo permite estimar los parámetros λ que maximizan la probabilidad de la secuencia $P(O|\lambda)$. Consta de los siguientes pasos (Rabiner, 1989):

1. Inicialización aleatoria de los parámetros $\lambda = (\pi, A, \theta)$.
2. Computo de la esperanza (*Expectation*) de la función de verosimilitud, dadas la secuencia observada O y el valor actual de los parámetros λ . Esto se puede desdoblar en dos fases (algoritmo *forward-backward*):

- a. Fase '*forward*': en esta fase se computa recursivamente la probabilidad de la secuencia observada hasta el tiempo t , estando en el estado i para dicho instante t . Si denotamos dicha probabilidad como $\alpha_i(t)$, y N como el número total de posibles estados, tenemos que:

- i. Inicialización: $\alpha_i(1) = \pi_i \cdot \theta_i(o_1)$.
- ii. Paso recursivo: $\alpha_i(t + 1) = \theta_i(o_{t+1}) \cdot \sum_{j=1}^N \alpha_j(t) \cdot a_{ij}$

Donde $1 \leq i \leq N$, y a_{ij} es la probabilidad de transición del estado i al estado j .

- b. Fase '*backward*': se calcula la probabilidad de la secuencia observada desde el tiempo $t+1$ hasta el final, estando en el estado i para el instante temporal t . Si denotamos esta probabilidad como $\beta_i(t)$, ésta puede calcularse recursivamente de la siguiente manera:

- i. Inicialización: $\beta_i(T) = 1$ para $1 \leq i \leq N$
- ii. Paso recursivo: $\beta_i(t) = \sum_{j=1}^N a_{ij} \cdot \beta_j(t + 1) \cdot \theta_j(o_{t+1})$

3. Maximización (*Maximization*, o '*update*'): a partir de los cálculos de esperanza realizados en el paso anterior, se re-calculan los parámetros λ . Si denotamos de la

siguiente forma la probabilidad de que el estado oculto actual X_t sea igual a i para la secuencia observada O y los parámetros λ :

$$\gamma_i(t) = P(X_t = i \mid O, \lambda)$$

$$\xi_{ij}(t) = P(X_t = i, X_{t+1} = j \mid O, \lambda)$$

Los parámetros se pueden actualizar de la siguiente manera:

- Distribución inicial: $\pi_i = \gamma_i(1)$
- Probabilidades de transición: $a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$
- Probabilidades de emisión: $\theta_i(v_k) = \frac{\sum_{t=1}^T 1_{o_t=v_k} \cdot \gamma_i(t)}{\sum_{t=1}^T \gamma_i(t)}$

Los pasos E (*Expectation*) – M (*Maximization*) se repiten hasta alcanzarse la convergencia, que ocurre cuando el cambio en la verosimilitud (o en los propios parámetros) cae por debajo de un cierto umbral.

Una vez ajustados a los datos – siendo estos datos los valores que han tomado las variables observables emitidas por el sistema cuyos estados subyacentes queremos caracterizar –, estos modelos permiten, por un lado, estimar la secuencia óptima de estados que maximiza la verosimilitud de los datos observables (y por ende la probabilidad a posteriori de que, para una secuencia dada de variables observables, el sistema se encuentre en uno u otro estado oculto en el momento de su emisión); y por otro lado, obtener la verosimilitud de que una secuencia dada de variables observables – vista o no en el entrenamiento – haya sido emitida por un sistema caracterizado por los parámetros π , \mathbf{A} , y θ del modelo ajustado.

Dada la naturaleza de estos modelos, pueden servir de base para el desarrollo de técnicas de detección de anomalías, que pueden ser de gran interés dentro de nuestro campo de estudio dado que un comportamiento anómalo en la medición de variables biométricas podría ser

señal de posibles eventos relacionados con la salud. Un artículo publicado en 2003 destaca el alto rendimiento de los modelos ocultos de Markov como herramientas para la detección de intrusiones en infraestructuras de redes informáticas, tanto en términos de precisión global como de minimización de la tasa de falsos positivos (Cho and Park, 2003). En el estudio, se entrenó un HMM para modelar el comportamiento normal, y se utilizó el modelo entrenado para evaluar la probabilidad con la que la secuencia observada del comportamiento actual se genera a partir del modelo, obteniendo una tasa de detección del 100% y una tasa de falsos positivos inferior al 10% al utilizar datos de cambios de privilegios (la investigación se centró en la mejora del tiempo de entrenamiento y del rendimiento al considerar únicamente los flujos de privilegios).

En línea con este estudio, un artículo publicado en 2005 desarrolló un sistema basado en HMM para identificar el tráfico de red TCP como un ataque potencial o como perteneciente al rango de comportamiento normal (Joshi and Phoha, 2005), alcanzando una precisión de clasificación del 79%. Un artículo más reciente también propone un marco para la detección de anomalías en series temporales multivariantes basado en HMM, aunque en su enfoque modelaron el comportamiento anómalo de la serie como uno de los posibles estados ocultos, en lugar de utilizar la verosimilitud de la secuencia de observación (Li et al., 2017). Al combinar el uso de HMM con la aplicación de diferentes métodos de transformación sobre las series temporales, fueron capaces de alcanzar niveles de precisión superiores al 90% en varios ejemplos de datos de la vida real.

En apoyo de la relevancia de la detección de anomalías dentro de nuestro campo de interés específico, estudios recientes aportan pruebas de la existencia de una correlación entre los cambios o anomalías en las mediciones de la frecuencia cardiaca y otras variables biométricas, como la duración del sueño en horas o el recuento de pasos (todas ellas adquiridas mediante dispositivos *wearables*), e infecciones víricas como la COVID-19 o las enfermedades similares a la gripe (Quer et al., 2021; Radin et al., 2020). Esta correlación puede explicarse por el estrés metabólico al que se ve sometido el organismo durante los periodos de enfermedad (por ejemplo, cuando aumenta la temperatura durante la fiebre), que

normalmente se traduce en un aumento de la frecuencia cardiaca en reposo y una mayor duración del sueño (Radin et al., 2020).

También se ha investigado la aplicación directa de los Modelos Ocultos de Markov a datos relacionados con la salud (presión arterial sistólica, índice de masa corporal y otros parámetros medidos durante revisiones médicas anuales), descubriéndose que estos modelos podían identificar tres niveles de riesgo para la salud en función de estas variables (Kawamoto et al., 2013). Al tratarse de modelos paramétricos, los HMM también se han utilizado como medio para el *clustering* de series temporales multivariantes compuestas por variables categóricas y continuas relacionadas con la trayectoria de salud de los sujetos (los HMM se utilizaron para modelar las distintas series temporales, y los modelos se agruparon después de definir una métrica de distancia adecuada) (Ghassempour et al., 2014), y para la diferenciación entre fases REM y NREM en base a cambios en la actividad cerebral durante el sueño (Mendez et al., 2010). En cuanto al uso de HMM con datos adquiridos a partir de dispositivos *wearables*, se han realizado estudios que exploran el uso de estos modelos para el reconocimiento de actividad a partir de datos brutos de sensores (Karaman et al., 2014; Li et al., 2015), probando su relevancia para la modelización de variables de naturaleza biométrica.

2.4.2 CLUSTERING DE SERIES TEMPORALES

En el contexto del análisis de datos y aprendizaje automático, los métodos de *clustering* (que podríamos traducir como ‘agrupación’) son un conjunto de técnicas pertenecientes a la sub-rama de algoritmos de aprendizaje no supervisado que buscan agregar los datos en distintos grupos (clusters) de forma que se maximice la similitud entre las observaciones u objetos dentro de un mismo clúster, y la disimilitud entre observaciones pertenecientes a diferentes clusters, de acuerdo con una métrica de similitud o distancia que deberá estar previamente definida. Es importante recalcar que los algoritmos no se apoyan en ningún conocimiento previo sobre las definiciones de los distintos grupos, sino que el propio proceso de agregación servirá para identificar de forma no supervisada la potencial estructura que pueda o no existir en los datos.

El *clustering* de series temporales es una modalidad particular de *clustering* en la que los objetos a agrupar son series temporales, esto es, secuencias de observaciones ordenadas en el tiempo que reflejan la evolución temporal de uno o varios atributos numéricos, y con un intervalo temporal entre puntos de datos sucesivos idealmente constante. Existen una infinidad de datos que adoptan la forma de series temporales, incluyendo variables de naturaleza biométrica registradas por pulseras de actividad y dispositivos de similar naturaleza, de ahí la relevancia de esta técnica para el proyecto.

En una extensa y exhaustiva revisión del estado del arte publicada en 2015, los autores (Aghabozorgi et al., 2015) distinguen tres grandes categorías dentro de las cuales se pueden clasificar la mayor parte de trabajos que abordan la agrupación de series temporales: (1) *clustering* de series temporales completas; (2) *clustering* de sub-secuencias o segmentos, y (3) *clustering* de puntos temporales. El *clustering* de series temporales completas, como su nombre indica, busca agrupar un conjunto de series temporales completas en función del grado de similitud que exista entre las mismas. Los objetos a agrupar, por tanto, son cada una de las series temporales individuales. El *clustering* de sub-secuencias, en cambio, busca agrupar distintos segmentos de una misma serie temporal; y el *clustering* de puntos temporales trata de agregar puntos temporales de la serie en función de la proximidad temporal entre los mismos y la similitud de los valores correspondientes. Para el presente trabajo nos centraremos en los primeros dos enfoques.

Existen tres formas de aproximarse al problema (Aghabozorgi et al., 2015). Estas aproximaciones son válidas tanto para la agrupación de series completas como para la agrupación de sub-secuencias

1. Enfoques basados en la forma de la serie temporal (*shape-based*): este enfoque agrupa de acuerdo con la similitud de las propias formas de las series temporales, teniendo en cuenta sus patrones y tendencias. Las técnicas que siguen este enfoque trabajan directamente con los datos en bruto de las series temporales, en lugar de extraer características de la misma o ajustar un modelo paramétrico que sintetice su comportamiento. La complejidad de esta aproximación aparece a la hora de definir

y computar una medida de similitud que nos permita comparar la forma de distintas series temporales. Una de las métricas más utilizadas para este propósito es la DWT (*Dynamic Time Wrapping*), una medida de distancia que calcula el alineamiento óptimo entre dos series temporales estirando o contrayendo sus ejes temporales de manera no lineal. Definida la métrica de distancia, se suelen aplicar algoritmos convencionales como *k-means* para realizar el *clustering* propiamente dicho.

2. Enfoques basados en la extracción de vectores de características: estas técnicas buscan sintetizar la serie temporal en bruto representándola mediante un vector de características de menor dimensionalidad. Las características o atributos a extraer de la serie pueden recoger información de la misma en el dominio del tiempo o en el de la frecuencia, y sobre estos vectores se aplicarían los algoritmos de clustering. En este caso, la métrica de distancia resultaría más sencilla de definir (se podría emplear, por ejemplo, la distancia euclídea previa normalización de las variables) y la dificultad radicaría en definir qué atributos son los más apropiados para sintetizar la serie temporal.
3. Enfoques basados en modelos: este último enfoque se basaría en el ajuste de un modelo paramétrico a partir de la serie temporal que capture la estructura, patrones y dependencias subyacentes en los datos (por ejemplo, un modelo oculto de Markov (Smyth, 1996)). Después, tras definir una métrica de distancia apropiada para evaluar la similitud entre los parámetros de los modelos entrenados, se aplicaría un algoritmo de *clustering* (típicamente convencional) sobre los propios modelos.

Dado que estaremos empleando HMM como principal herramienta para caracterizar la serie temporal de ritmo cardiaco, puede ser interesante en un futuro explorar esta última aproximación al *clustering*. Sin embargo, dado que tiene mayor interés para el proyecto la clusterización de sub-secuencias que la clusterización de series completas, este enfoque añade más complejidad de la necesaria, por lo que se optó finalmente por una aproximación basada en vectores de características.

2.4.3 AUTOENCODERS

Los *autoencoders*, o autocodificadores, son una clase de redes neuronales de aprendizaje no supervisado empleadas para tareas muy variadas, entre las que se incluyen la reducción de dimensionalidad, la detección de anomalías, o el filtrado de ruido. Su arquitectura (en esencia, un perceptrón multicapa completamente conectado) consta de dos componentes: el encoder y el decoder. El encoder mapea los datos de entrada a una representación de dimensionalidad inferior (el espacio vectorial de dimensionalidad reducida se denomina ‘espacio latente’), y el decoder recibe como entrada el vector latente y, a partir del mismo, trata de reconstruir los datos originales (Figura 2).

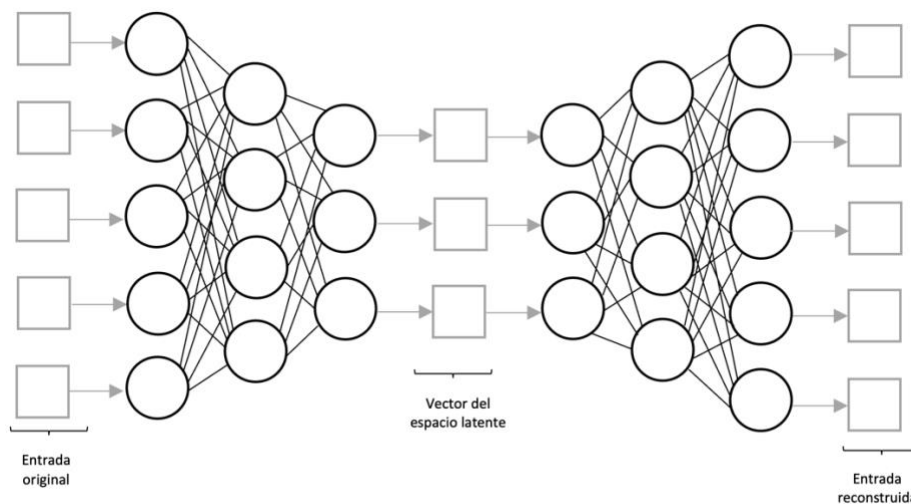


Figura 2: Ejemplo de un autoencoder básico de arquitectura simétrica.

Al ajustarse el modelo, la función de pérdida a minimizar es el error de reconstrucción, esto es, el error o diferencia entre los datos originales en la entrada del encoder y los datos reconstruidos en la salida del decoder. De este modo, se fuerza a la red a aprender como representar de manera comprimida y eficiente la información.

Su relevancia para la detección de anomalías deriva del hecho de que, al ajustarse, para poder comprimir correctamente los datos minimizando la pérdida de información (de modo que éstos sean después reconstruibles), la red debe aprender a caracterizar adecuadamente las

relaciones lineales y no lineales que existan entre las variables o dimensiones de los datos originales. Por ello, si aparecen observaciones cuyo comportamiento se separa mucho del observable en el grueso de los datos de entrenamiento, o que no parecen pertenecer a la misma distribución, la red tendrá mayores dificultades para retener la información de dicho dato tras la compresión y reconstruir la observación a partir de vector latente. Dicho de otra forma, podemos esperar que el error de reconstrucción suba. Por tanto, dado que para la clusterización de segmentos de series temporales se explorará la aproximación basada en extracción de vectores de características que sinteticen dichas sub-secuencias, esos mismos vectores podrán usarse para entrenar un autoencoder a partir del cual identificar aquellos segmentos cuyo comportamiento es anómalo.

Los *autoencoders* pueden además emplearse para reducción de dimensionalidad, lo cual de nuevo tiene aplicaciones interesantes de cara al proyecto: otra opción alternativa al empleo de vectores de características que sinteticen sub-secuencias de la serie temporal sería el uso de autoencoders que aprendan a comprimir y reconstruir las propias sub-secuencias en bruto, de modo que sean capaces de representarlas en un vector de dimensionalidad muy inferior. Tras el entrenamiento, se emplearía el encoder para obtener un vector de espacio latente por segmento, y sobre dichos vectores se podría aplicar un algoritmo de *clustering* convencional (por ejemplo, *k-means*). Esta aproximación evitaría la dificultad de tener que identificar qué atributos de cada segmento son verdaderamente relevantes de cara a construir los vectores de características, pero tendría como inconveniente la imposibilidad de interpretar los vectores latentes. Pese a ello, sí se podría caracterizar el comportamiento o rasgos de cada uno de los clusters en los que los vectores latentes son agrupados, a partir de las secuencias originales de ritmo cardíaco representadas por dichos vectores.

La arquitectura que se acaba de describir es la implementación básica del concepto de *autoencoders*, pero existen muchas otras variantes. Algunas de las más empleadas son las siguientes:

1. *Sparse autoencoders*: Se introduce un término de penalización que obliga a las neuronas a permanecer inactivas – entendiéndose como inactivas que su valor de

salida sea muy cercano a 0, asumiendo una función de activación sigmoïdal – la mayor parte del tiempo, conforme a un parámetro de ‘*sparsity*’ (escasez) que determina cual es el valor medio de activación que queremos para la neurona. Esto conduce a que la red aprenda a generar representaciones internas más robustas de las relaciones entre las variables de entrada (Ng and others, 2011).

2. *Autoencoders* de eliminación de ruido: Se entrenan usando versiones ruidosas de los datos de entrada, a partir de las cuales deben reconstruir los datos originales limpios de ruido. De este modo, la red aprende a extraer características significativas e ignorar todo lo que no contenga información (Vincent et al., 2010, 2008).
3. *Autoencoders* Variacionales (VAE): Modelan el espacio latente como una distribución de probabilidad, imponiendo la distribución como prior (por ejemplo, una distribución gaussiana) sobre las variables latentes (Kingma and Welling, 2019). Lo interesante de estas redes es que pueden emplearse como modelos generativos, pues son capaces de generar nuevas observaciones sintéticas de los datos al muestrear desde el espacio latente.
4. *Autoencoders* contractivos: Introducen un término de penalización adicional en la función de pérdida, lo cual incentiva que la representación del espacio latente sea insensible a pequeñas variaciones en los datos de entrada, buscando que el aprendizaje se centre en características estables e invariantes (Rifai et al., 2011).
5. *Autoencoders* convolucionales (Masci et al., 2011) y *autoencoders* secuenciales (Luong et al., 2015). Son *autoencoders* que emplean capas convolucionales o secuenciales, respectivamente, para poder procesar datos en malla (como imágenes) o secuenciales (por ejemplo, series temporales). Los *autoencoders* secuenciales suelen emplear arquitecturas recurrentes (por ejemplo, redes LSTM).

La variante específica basada en arquitecturas recurrentes son de particular relevancia para el proyecto, dada la naturaleza secuencial de los datos con los que se estará trabajando. Este

tipo de arquitectura ya se ha empleado para diversas aplicaciones basadas en datos de sensores *wearable*, incluyendo el reconocimiento de actividades (Wang, 2016) o la detección de recaídas en pacientes con trastornos psicóticos (Panagiotou et al., 2022).

Capítulo 3. METODOLOGÍA

El siguiente diagrama (Figura 3) resume la metodología completa desarrollada e implementada en el proyecto, desde el reclutamiento de voluntarios hasta el análisis de los modelos entrenados.



Figura 3: Metodología implementada

En las siguientes secciones se desglosa en detalle cada una de las etapas mostradas en el diagrama.

3.1 RECLUTAMIENTO DE VOLUNTARIOS Y TOMA DE MEDIDAS

3.1.1 SELECCIÓN DE VOLUNTARIOS

Para asegurar que la muestra sea suficientemente representativa de la población general y permita derivar conclusiones generalizables, se aplicaron los siguientes criterios de inclusión:

- a) Rango de edad: 18-60 años.
- b) Sin patologías conocidas ni antecedentes de enfermedades cardiovasculares graves
- c) Disposición para participar en el estudio y capacidad legal de dar un consentimiento informado.

Dada la naturaleza exploratoria de este estudio, y siempre dentro de los límites definidos por los criterios anteriores, reclutamos un grupo de nueve voluntarios bastante heterogéneo en cuanto a edad, sexo y nivel de actividad. La muestra consta de 3 mujeres y 6 hombres, con edades distribuidas de la siguiente manera: 1 menor de 20 años, 4 entre 20-30 años, 1 entre 40-50 años y 3 entre 50-60 años.

Si bien el tamaño de la muestra podría considerarse reducido, se ha de tener en cuenta que en este estudio se buscará fundamentalmente una caracterización **individualizada** de los parámetros biométricos de cada sujeto; este enfoque viene justificado por la constatable variabilidad en el comportamiento de las variables fisiológicas con las que se estará trabajando en función de la edad, hábitos de alimentación o nivel de actividad, como se verá en el siguiente capítulo. En cualquier caso, la variedad en la distribución de las variables demográficas hace posible evaluar la viabilidad del enfoque para cualquier sujeto, con independencia del impacto de dichos factores en el ritmo cardiaco del individuo.

Se hubo de descartar los datos de dos de los voluntarios por insuficiencia de muestras o errores de configuración de la pulsera. Por último, para dos de los restantes siete voluntarios (concretamente, ID6 e ID9) no se pudo contrastar las anomalías detectadas con la incidencia de eventos de salud reales mediante análisis retrospectivo, debido a que participaron en una fase muy temprana del estudio y no fue posible aplicar las técnicas de modelización de una manera sistematizada hasta mucho más tarde. Resultaba por ello inviable aplicar un análisis retrospectivo de eventos de salud con tanta distancia temporal.

3.1.2 PROCEDIMIENTO PARA LA EXTRACCIÓN DE DATOS

El procedimiento para la recogida y extracción de datos se efectuó de la siguiente manera: se dieron instrucciones a los voluntarios para llevar puestas las pulseras durante un periodo prolongado, que se extendió desde los dos meses en algunos casos, hasta los diez meses en otros. Se debe señalar que, durante dicho periodo, hubo intervalos de tiempo esporádicos en los que algunos de los voluntarios dejaron de llevar el dispositivo por diversos motivos, si bien en la mayoría de casos estos participantes retomaron las pruebas pasadas dichas interrupciones. Esto sucedió sobre todo con voluntarios que llevaron la pulsera durante

periodos de tiempo especialmente prolongados; en este sentido, se debe constatar que la integración de estos dispositivos en el día a día no siempre es sencilla para todos los individuos, dato a tener en cuenta a la hora de evaluar el alcance real de esta aproximación para la monitorización del estado de salud.

Varias veces a lo largo del periodo de prueba, y también al concluir el mismo, se procedió a la extracción de los datos recogidos por los dispositivos para su análisis. La periodicidad de las extracciones se situó en torno a los dos meses, lo que implica que en algunos voluntarios solo se llevó a cabo una única extracción. Dentro de este sub-grupo se tuvieron que descartar los datos de dos de los voluntarios, tal y como se comentó anteriormente: en un caso por insuficiencia de muestras, y en otro de los casos por un error en la configuración de la pulsera.

Antes de realizar cualquier procesamiento o análisis exploratorio, los datos se anonimizaron eliminando cualquier información personal identificable (IPI) y asignando identificadores únicos. Los voluntarios posteriormente informarían de manera retrospectiva sobre la ocurrencia de dichos sucesos, identificando eventos potencialmente asociados con los comportamientos atípicos de ritmo cardiaco identificados tras la modelización.

3.2 PREPROCESAMIENTO BÁSICO Y ANÁLISIS EXPLORATORIO

El propósito de esta fase fue el de homogeneizar la estructura de los datos para poder aplicar de manera idéntica y sistemática en todos los voluntarios el procesamiento necesario en las fases de análisis exploratorio, modelización e interpretación de resultados. Ello requirió acceder correctamente a los campos de los registros en los que se encuentra la información relevante (lo cual puede variar mucho en función del modelo de pulsera), calibrar las fechas entre registros de distinto tipo, y verificar – y en su caso forzar – la uniformidad en el periodo intermuestral a lo largo de la serie. Tras este procesamiento se procedió a un análisis exploratorio de la frecuencia cardiaca, que evidenció la necesidad de focalizar el análisis en el periodo nocturno (para volver los datos de distintas fechas más comparables). Por ello, a

modo de preparación final para la modelización, se hubieron de filtrar las observaciones que no se corresponden exclusivamente al periodo nocturno de sueño.

3.3 AJUSTE DE MODELOS Y ANÁLISIS DE RESULTADOS

Una vez procesados los datos, se aplicaron los modelos antes descritos de manera que permitiesen (1) caracterizar y entender el comportamiento normal de la frecuencia cardiaca nocturna, y (2) identificar desviaciones respecto al comportamiento caracterizado como normal. Estos objetivos no los han de cubrir cada tipo de modelo individual (por ejemplo, en el caso de los autoencoders la caracterización no es interpretable), sino el análisis conjunto resultante de la aplicación de los tres modelos sobre los datos de un mismo sujeto. A tal efecto, algunas de las técnicas aplicadas se orientarán específicamente al primero de los objetivos, otras al segundo, y la evaluación final buscará integrar las conclusiones derivadas de todas ellas contrastando entre sí los resultados. El siguiente diagrama (Figura 4) resume el procedimiento.

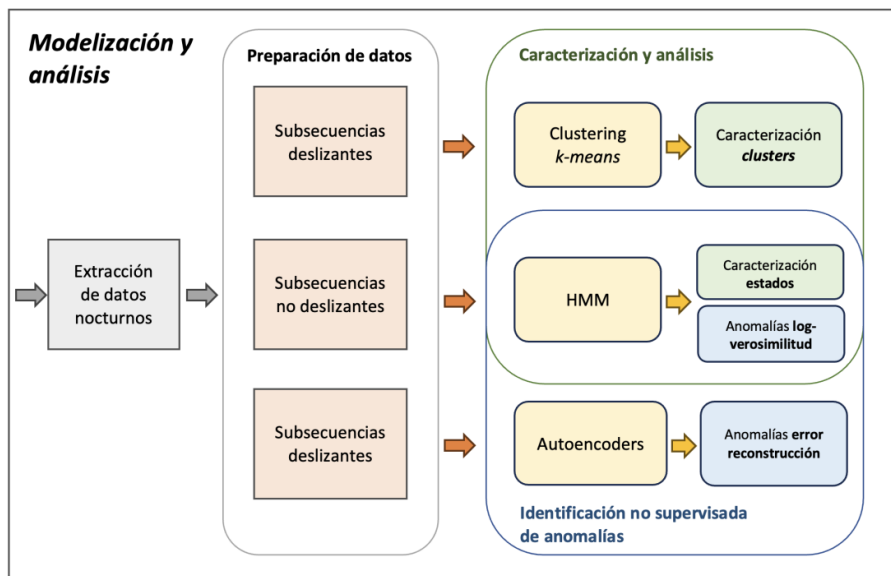


Figura 4: Diagrama de modelización

Las razones por las que el análisis se centra en las medidas de frecuencia cardiaca nocturna se exponen más adelante, en el capítulo dedicado al análisis exploratorio de los datos. Las fechas identificadas como anómalas se contrastaron con la evaluación retrospectiva de los voluntarios respecto a eventos de salud que recordaban o creyeron haber experimentado durante su participación en el estudio. Se evaluaron los resultados de los modelos tanto aisladamente como en conjunto.

Capítulo 4. PREPROCESAMIENTO Y EDA

4.1 PREPROCESAMIENTO BÁSICO DE LOS DATOS EN BRUTO

El formato de los datos en bruto presentaba importantes variaciones en función del tipo de dispositivo, por lo que se hubo de aplicar una etapa inicial de preprocesamiento para garantizar consistencia y facilitar su manejo.

Para el caso de los dispositivos **MiBand 6**, la totalidad de registros de ritmo cardíaco se encontraba almacenada en un único fichero CSV para cada participante, con un intervalo entre muestras constante de 1 minuto de duración. El registro de sueño también presentaba un formato de almacenamiento CSV, y de nuevo la información estaba agregada en un único fichero para cada individuo. Por ello, la lectura de los datos presentaba pocas dificultades. Únicamente se hubo de abordar un problema de inconsistencia entre el formato de fecha en los ficheros de ritmo cardíaco y en los ficheros de registro de sueño: las fechas para el registro de ritmo cardíaco estaban expresadas en horario GMT+1 (horario español de invierno) o GMT+2 (horario de verano), mientras que el registro de sueño estaba referenciado temporalmente en horario GMT+0.

Más allá de la inconsistencia entre husos horarios, el preprocesamiento de los datos de este modelo de pulsera resultó bastante sencillo, gracias a la sencillez en el formato de los ficheros de datos. Los registros de frecuencia cardíaca se encontraban unificados a nivel de individuo en un fichero CSV de tres campos: fecha (*date*), hora (*time*) y la lectura de frecuencia cardíaca correspondiente a dicho instante temporal (*heartRate*), en unidades de *ppm* (pulsaciones por minuto). Tras la lectura de los datos, se agregó la información temporal en un único campo – *dateTime* – que serviría como índice del conjunto de datos. En cuanto a los registros de sueño, estos también venían recogidos en un único fichero CSV por individuo, con siete campos fundamentales (*date*, *deepSleepTime*, *shallowSleepTime*, *wakeTime*, *start*, *stop*, y *REMTIME*) y varios adicionales para recoger los periodos de siesta.

El campo *date*, como cabe esperar, contiene la fecha correspondiente a cada observación; y los campos *start* y *stop* recogen el instante temporal en el que comienza y finaliza el periodo de sueño nocturno asociado a dicha fecha. Los otros cinco campos principales contienen la suma de minutos en los que el usuario ha estado en cada uno de los estados a los que hacen referencia (sueño ligero, sueño profundo, despierto y en fase REM) pero como no desglosan esta información ubicándola temporalmente dentro del periodo nocturno – esto es, indicando exactamente en qué instantes temporales tienen lugar dichas fases – no nos sirven para etiquetar con precisión los registros de ritmo cardiaco asociados a cada etapa. La única información verdaderamente relevante, por tanto, es la contenida en los campos *start* y *stop*, que se empleó para discriminar entre observaciones tomadas cuando el usuario se encontraba dormido ('1') y aquellas tomadas con el sujeto despierto ('0').

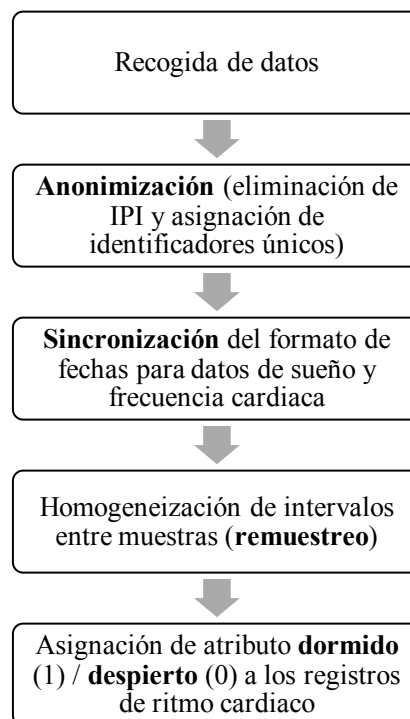


Figura 5: Flujo de preprocesamiento básico

Los datos MiBand presentaban por lo general intervalos entre muestras homogéneos de 1 minuto de duración, por lo que no se requirió remuestreo de la serie. Junto al ritmo cardiaco y el registro de sueño, las pulseras MiBand también proporcionaban información sobre el recuento de pasos diarios o niveles de actividad. Estas variables adicionales prácticamente no requirieron preprocesamiento; en cualquier caso, nuestro objetivo se centra en la caracterización del ritmo cardiaco, por lo que el contenido de estas variables solo será relevante en la medida en que pueda asociarse con las lecturas de frecuencia cardiaca.

Los datos procedentes de las pulseras **Fitbit** (tanto Charge como Inspire) presentaron mayores requisitos de procesamiento, si bien la información que proporcionan es más detallada. De entrada, los datos de sueño no estaban unificados para cada individuo sino guardados en múltiples ficheros de formato JSON, que desglosaban la información en múltiples campos y sub-campos anidados. Los ficheros estaban compuestos por una lista de objetos, cada uno de los cuales representa una única sesión de sueño. A su vez, cada objeto de sesión de sueño incluye varios campos que describen en detalle dicha sesión.

Los datos de frecuencia cardiaca también se encontraban guardados en múltiples ficheros JSON, si bien la estructura de los mismos es un poco menos compleja: cada uno de ellos está constituido por una lista de objetos, donde cada objeto recoge una observación de frecuencia cardiaca, con un campo de fecha y un campo denominado *value* que contiene la información propiamente dicha. Dicha información viene expresada en los sub-campos *bpm*, con la muestra de frecuencia cardiaca en pulsaciones por minuto (ppm); y *confidence*, cuyo valor es un entero de 0 a 3 que expresa la fiabilidad de la medición en cuestión.

Para la lectura de los datos se hubo de recorrer de manera secuencial cada uno de estos ficheros, extrayendo la información de los campos relevantes a fin de garantizar que la estructura final de los datos cargados en el entorno de ejecución fuese idéntica a la adoptada para datos MiBand; eso sí, no se ha de olvidar que la frecuencia de muestreo del ritmo cardiaco es en cualquier caso distinta según el tipo de pulsera. También se requirió de nuevo una sincronización de fechas entre los registros de sueño y los de ritmo cardiaco.

Por último, como ya se ha comentado, si bien el intervalo entre muestras mínimo en dispositivos Fitbit es de cinco segundos, este intervalo no siempre es constante para muestras consecutivas. Es fundamental que los intervalos entre muestras sean uniformes a lo largo de la serie temporal de cara a la aplicación de las técnicas de aprendizaje automático, por lo que se hizo necesario remuestrear la serie fijando el periodo intermuestral en cinco segundos. Se recurrió al método *spline* de interpolación, basado en el empleo de polinomios a trozos (en nuestro caso de orden 3) para aproximar la función continua subyacente. Concluido el remuestreo, de nuevo se aprovechó el registro de sueño para asignar un marcador indicando, para cada observación de ritmo cardiaco, si ésta se tomó estando el sujeto despierto o dormido.

4.2 ANÁLISIS EXPLORATORIO DE LOS DATOS (EDA)

Comenzaremos el análisis exploratorio con los datos del sujeto ID1. A continuación, se puede observar la representación gráfica de un pequeño segmento de la serie temporal de ritmo cardiaco, expresado en pulsaciones por minuto (ppm, o *bpm* en inglés).

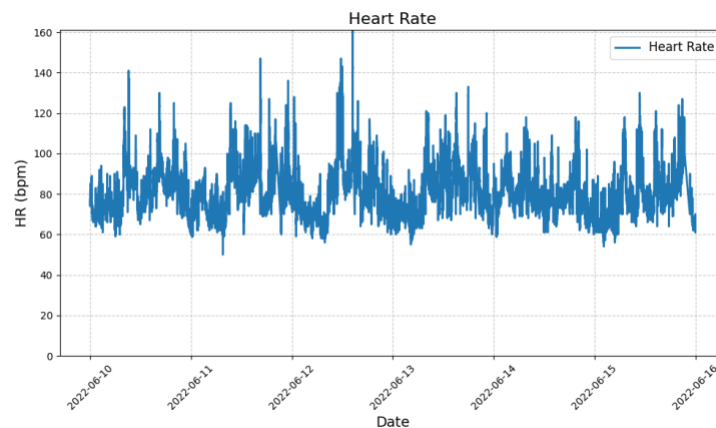


Figura 6: Fragmento de la serie temporal de ritmo cardiaco para el sujeto ID1

Además de poderse constatar la elevada variabilidad en las cifras de ritmo cardiaco para este sujeto, una primera inspección visual nos permite identificar ciertos patrones periódicos de aumento y disminución de la media y la desviación típica de la frecuencia cardiaca,

sucedándose uno detrás de otro con un periodo aproximado de 24 horas. Aún sin tener más información que esta, la periodicidad seguida por estas variaciones en el comportamiento de la serie temporal nos permite hipotetizar su probable correlación con los ciclos de sueño del sujeto. Y en efecto, si empleamos la información de sueño añadida a este registro para sombrear aquellos periodos en los que el sujeto está dormido, podemos verificar una clara correspondencia entre estas variaciones y el ciclo vigilia-sueño.

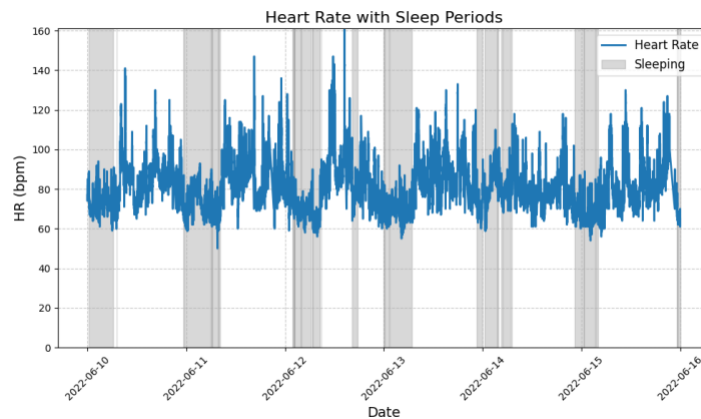


Figura 7: Fragmento con regiones correspondientes al sueño sombreadas

Este fenómeno no es exclusivo de este sujeto en particular. A continuación se exponen representaciones gráficas de las mismas variables para otros cuatro sujetos de la muestra. Nótese, sin embargo, que la diferencia entre el comportamiento de la frecuencia cardiaca dentro y fuera del periodo de sueño es más dramática en algunos sujetos, y menos pronunciada en otros.

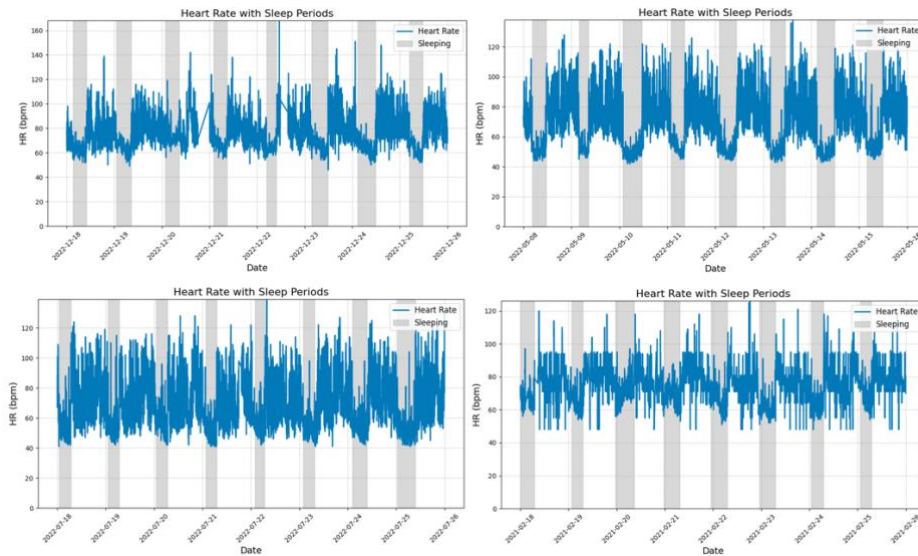


Figura 8: fragmentos de la serie de ritmo cardiaco para otros sujetos

La diferencia entre los valores de ritmo cardiaco cuando el voluntario está dormido frente a un estado de vigilia también es verificable si representamos mediante histogramas la distribución de estas mismas variables para cada uno de estos estados.

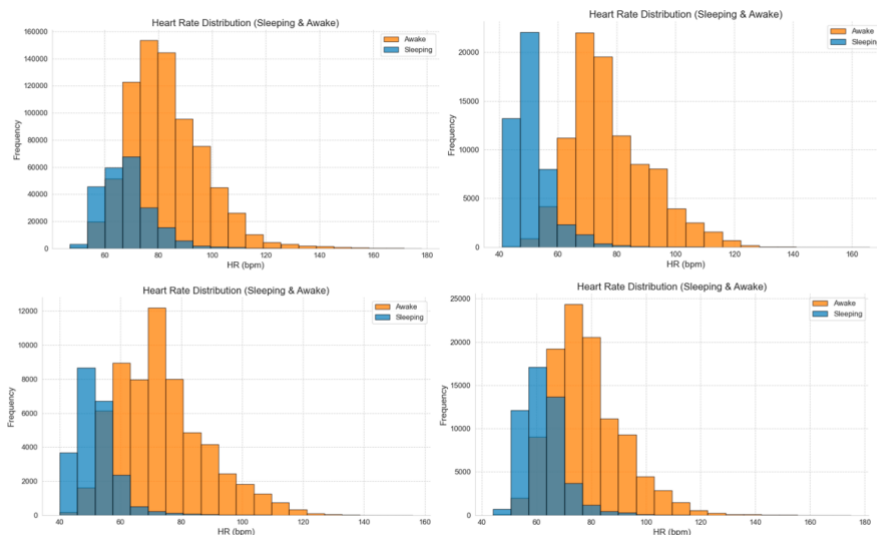


Figura 9: Comparativa de distribuciones de ritmo cardiaco dentro y fuera del periodo de sueño

Si bien hay solape entre ambas distribuciones, los registros correspondientes al periodo nocturno muestran una media y mediana notablemente inferiores a la que encontramos en las distribuciones de valores diurnos de ritmo cardiaco, y una frecuencia de valores elevados también muy inferior. En este sentido, otro punto reseñable es también la mayor longitud de las colas superiores de la distribución para valores correspondientes a un estado de vigilia, frente a las de las distribuciones de mediciones tomadas estando el sujeto dormido.

De este primer análisis podemos extraer dos conclusiones: por un lado, (1) el comportamiento de la frecuencia cardiaca diverge de manera significativa y predecible entre el periodo en el que el sujeto está dormido, y el periodo en el que está despierto; pero (2) pese a ser un patrón que se repite en todos los sujetos, estas mismas diferencias en la media y desviación típica son más o menos pronunciadas en función del sujeto en cuestión. Hay, en definitiva, una importante variabilidad *inter*-sujeto que si bien confirma la diversidad en la muestra, también apunta a la necesidad de seguir un procedimiento individualizado a la hora de modelar y caracterizar los parámetros normales o típicos de comportamiento de esta variable biométrica, ya que divergirán notablemente de una persona a otra.

Además, la conclusión (1) también pone de manifiesto una fuerte variabilidad *intra*-sujeto en el comportamiento de la serie temporal, pues vemos importantes cambios en su comportamiento en función del momento del día. Por ello, como ya apuntábamos en la introducción, a la hora de ajustar los tres modelos seleccionados para cubrir los objetivos del estudio (HMM, *autoencoders* y clustering), **se pondrá el enfoque exclusivamente en el periodo nocturno por razones de homogeneidad.**

Es preferible orientar la caracterización hacia el periodo nocturno en exclusiva, en lugar de hacia el periodo diurno en exclusiva, debido a que a lo largo del día hay una multiplicidad de factores (como la ingesta de cafeína, la realización de ejercicio físico, la aparición de imprevistos, etc.) que pueden afectar el comportamiento de variables fisiológicas como la frecuencia cardiaca, mientras que durante el periodo nocturno es más probable que cualquier anomalía sea un reflejo de una alteración interna en el sujeto, o al menos el efecto de una variable externa que influya de manera significativa en el estado de salud. La fiabilidad de

las medidas tomadas en reposo es además apreciablemente mayor debido a la ubicación del sensor en la muñeca.

Capítulo 5. DESARROLLO DE MODELOS

En este capítulo se detalla el procedimiento de modelización para la caracterización del ritmo cardiaco nocturno y el correspondiente análisis de resultados. Como se vio anteriormente, el siguiente diagrama expone las etapas de esta última fase de manera sintetizada.

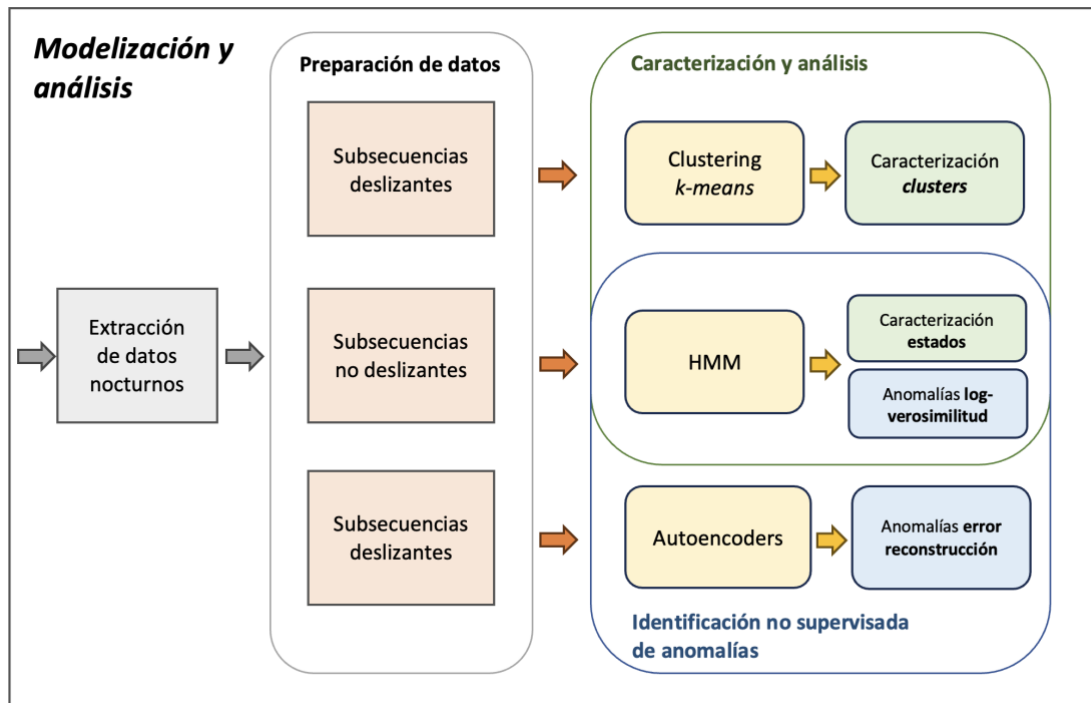


Figura 10: Flujo de modelización y análisis ampliado

Como se comentó en el capítulo 3, no solo se evaluarán los resultados para cada modelo individualmente, sino que también se contrastarán los modelos entre sí, dándose un mayor peso a aquellas anomalías identificadas tanto por HMM como por autoencoders.

5.1 *MODELOS OCULTOS DE MARKOV*

Tal y como se vio en la revisión del estado del arte, ajustar estos modelos a una secuencia observada implica estimar el vector de probabilidades iniciales π , la matriz de probabilidades de transición entre estados \mathbf{A} , y los parámetros θ que determinan la distribución de probabilidad de emisión del sistema, que varían en función del estado actual (nótese que el número de estados ocultos es un hiperparámetro pre-establecido antes de ajustar el modelo). Estos parámetros caracterizan completamente el comportamiento y evolución del sistema modelado. Para el ajuste del modelo a los datos (secuencia de observaciones de la variable de emisión, es decir, la serie temporal de frecuencia cardíaca nocturna) se empleó el algoritmo de Baum-Welch, descrito en el capítulo dedicado a la revisión del estado del arte.

Una vez ajustados, estos modelos permiten, por un lado, estimar la secuencia óptima de estados que maximiza la verosimilitud de los datos observables (y por ende la probabilidad a posteriori de que, para una secuencia dada de variables observables, el sistema se encuentre en uno u otro estado oculto en el momento de su emisión); y por otro lado, obtener la verosimilitud de que una secuencia dada de variables observables – vista o no en el entrenamiento – haya sido emitida por un sistema caracterizado por los parámetros π , \mathbf{A} , y θ del modelo ajustado. En nuestro caso, la variable observable sería el ritmo cardíaco, y el sistema modelado sería el propio sujeto. La estimación de la secuencia óptima de estados permite caracterizar el comportamiento de la serie temporal para cada estado y por ende entender mejor la evolución de la frecuencia cardíaca a lo largo del tiempo. Por otro lado, la estimación de la verosimilitud permite identificar de manera no supervisada desviaciones del comportamiento del ritmo cardíaco respecto a los patrones típicos caracterizados mediante los parámetros del HMM.

Estas desviaciones podrían estar potencialmente asociadas con eventos de salud, por lo que también evaluaremos la coincidencia de noches en las que se dan estas desviaciones con fechas en las que tienen lugar eventos de salud – física y mental – o interrupciones del ritmo circadiano identificadas por el sujeto. En adelante, nos referiremos a estos eventos con las siglas ESD, a modo de abreviatura. Dentro de esta etiqueta se englobará cualquier evento

que pueda ser relevante para el estado de salud general del sujeto, incluyendo episodios gripales, alteraciones en los ritmos circadianos del sujeto, consumo de alcohol o insomnio, por poner algunos ejemplos.

Existen una infinidad de distribuciones de probabilidad que se pueden asumir para modelar las variables observables; dado que no tenemos conocimiento previo suficiente para determinar con un criterio objetivo qué distribución es la más adecuada, un buen punto de partida sería el de asumir emisiones gaussianas. Con todo, dada la naturaleza de los datos, asumir una distribución gaussiana simple para cada estado supone simplificar en exceso el sistema, por lo que se modelará como una mezcla de gaussianas.

5.1.1 CARACTERIZACIÓN HMM DE LA SERIE TEMPORAL EN BRUTO

Comencemos nuestro análisis con los datos del primer voluntario (ID1). Como pudimos constatar en el anterior análisis exploratorio, dada la notable variabilidad inter-sujeto será preferible una caracterización individualizada del ritmo cardíaco. Por otra parte, dado que la serie de ritmo cardíaco se puede ver afectada en el periodo diurno por una infinidad de factores (nivel de estrés, ingesta de cafeína, ejercicio, día de la semana...) resulta de mayor interés para este proyecto poner el foco de atención en el periodo de sueño, por ser el más comparable de un día a otro. Cualquier alteración durante dicho periodo tendrá una mayor probabilidad de venir originado por el estado interno del organismo del sujeto. Por ello, aprovechando la variable añadida en la fase de preprocesado (para indicar si una observación de ritmo cardíaco se tomó estando el sujeto dormido o despierto), únicamente se emplearon datos tomados durante el periodo de sueño para el ajuste y evaluación del modelo.

El siguiente gráfico desglosa la metodología de modelización e identificación de anomalías con mayor detalle (Figura 11).

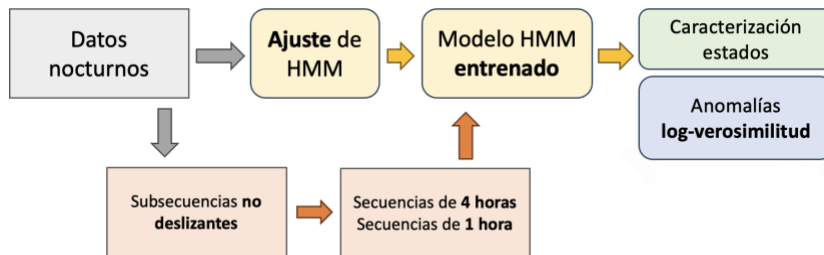


Figura 11: Metodología específica HMM

Lo primero a determinar es el número de estados ocultos que se asumirá para el modelo. Tal y como se ha dicho anteriormente, éste es un hiperparámetro previo fijado antes de poder realizar cualquier ajuste, de modo similar a la distribución de probabilidad de emisión. Dado que de nuevo no contamos con conocimiento previo suficiente para prefijar el número de estados, para tratar de seguir un procedimiento objetivo nos serviremos de la verosimilitud: ajustaremos sucesivamente varios modelos con diferente número de estados, computaremos la cifra de verosimilitud de los datos para cada modelo, y representaremos la evolución de la verosimilitud logarítmica en función del número de estados (Figura 12). El número de estados a emplear será aquel a partir del cual la cifra de verosimilitud parezca estabilizarse.

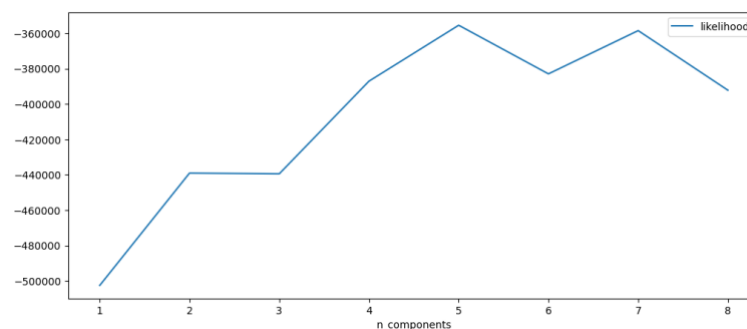


Figura 12: evolución de la verosimilitud logarítmica en función del número de estados.

Como puede observarse en la Figura 8, la verosimilitud parece estabilizarse en torno a los cinco-seis estados; a la vista de estos resultados, nos decantamos por $n = 5$ estados ocultos. Ajustado este hiperparámetro básico, se procedió a modelar la serie de frecuencia cardiaca.

Con el modelo ya entrenado, podemos obtener y representar las probabilidades a posteriori de que el sistema se encuentre en cada uno de los posibles estados para cada una de las observaciones de ritmo cardiaco, a continuación se muestra como ejemplo el resultado para una de las noches (Figura 13).

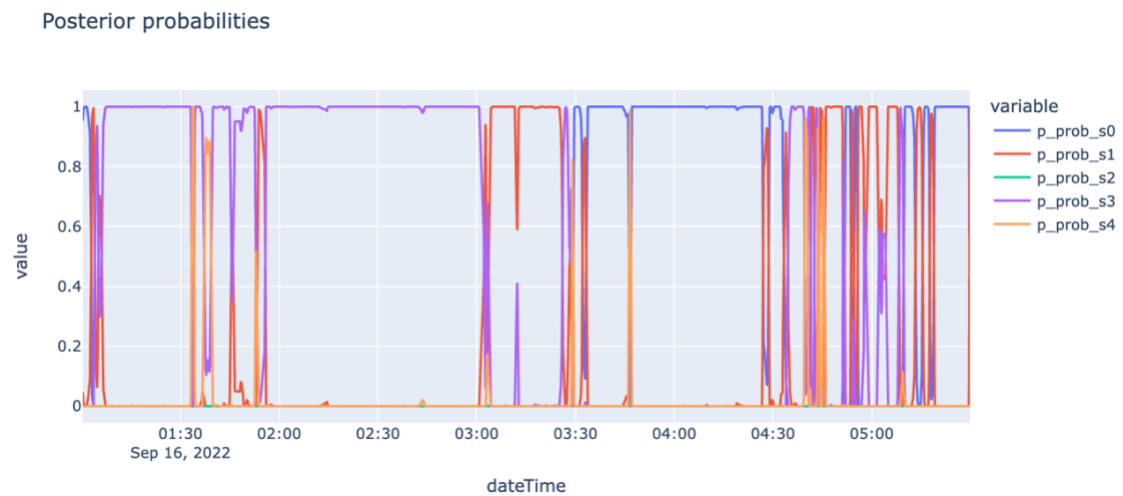


Figura 13: Evolución de probabilidades a posterior de cada uno de los cinco estados (noche del 16/9/22).

Siendo la evolución de la frecuencia cardiaca para el mismo intervalo temporal la siguiente (Figura 14).

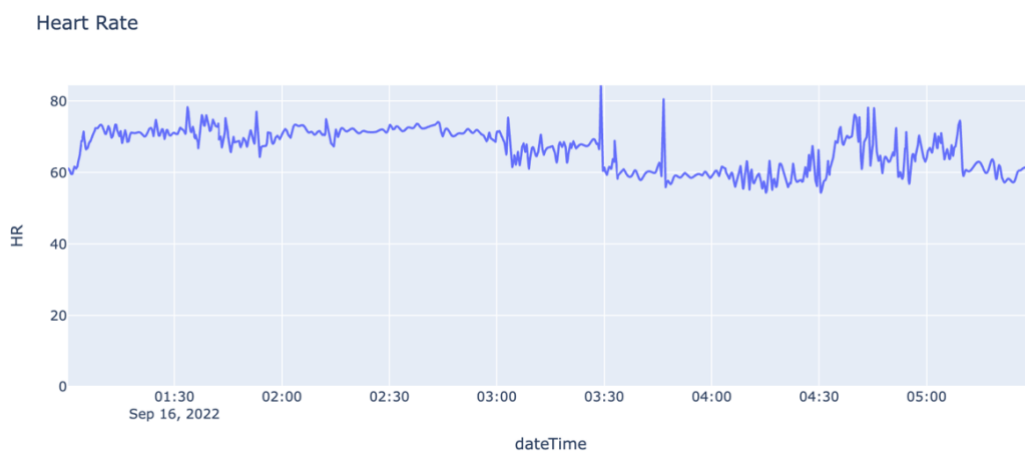


Figura 14: Frecuencia cardiaca para la noche del 16 de septiembre.

A priori, no resulta sencillo otorgar una interpretación clara al significado de estos estados, si bien sí se aprecia que algunos tienden a predominar sobre otros, siendo notablemente menos frecuente el estado 2 en la gran mayoría de las noches (sirva como ejemplo la noche representada en la figura 10, donde prácticamente no hay observaciones en las que la probabilidad asociada a este estado llegue siquiera a aproximarse a niveles de 0,5). El estado 4 también es poco frecuente, y si verificamos los valores que toma el ritmo cardiaco para aquellos periodos de tiempo en los que dicho estado se vuelve altamente probable, vemos que su ocurrencia tiende a estar correlacionada con comportamientos más inestables de la serie temporal, caracterizados por cambios bruscos y una mayor desviación típica.

A fin de poder dar una interpretación fisiológica a dichos estados, se representaron mediante histogramas cada una de las distribuciones de valores de frecuencia cardiaca asociadas a los mismos.

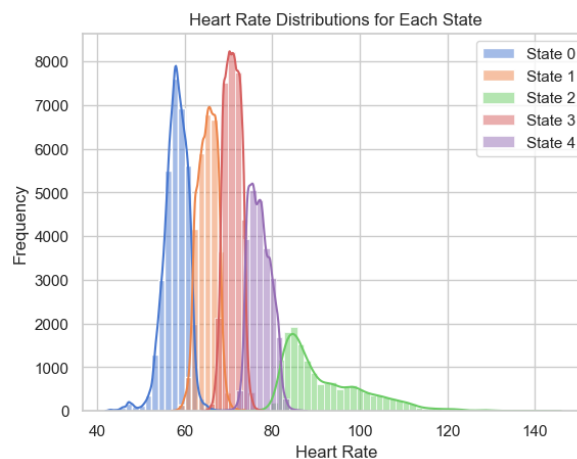


Figura 15: Distribución de los valores de ritmo cardiaco para cada estado (sujeto ID1).

Como puede constatar, y como veníamos previendo, el estado 2 (en verde) es el menos frecuente. Es interesante señalar que dicho estado también cuenta con la media de ritmo cardiaco más elevada, y en general con las observaciones de ritmo cardiaco de valor más extremo, además de una elevada desviación típica. No es por tanto sorprendente que a este estado venga asociada una porción apreciable de observaciones (más del doble que en otros

estados) en las que el sujeto se encuentra despierto o con insomnio. Esta información solo es posible derivarla a partir de los datos tomados por pulseras Fitbit, ya que las MiBand 6 no aportan información relativa a las estimaciones de fases de sueño desglosada por segmentos temporales, por lo que no será un análisis que replicaremos en los demás usuarios. En cualquier caso, si observamos que en los restantes sujetos alguno de los estados caracterizados por el modelo HMM resulta correlacionarse con un comportamiento de ritmo cardiaco muy similar al que vemos aquí en el estado 2, es presumible que también dicho estado englobe observaciones en las que el sujeto esté despierto o con insomnio.

El estado 0 (en azul), por otra parte, es el más predominante con un 25% de ocurrencia en el periodo de prueba, seguido de cerca por el 3 (en rojo, con un 23% de ocurrencia) y el 1 (naranja, con un 21% de observaciones asociadas a dicho estado). El estado 0 parece estar caracterizando periodos con niveles de frecuencia cardiaca más bajos y un comportamiento por lo general más estable que los segmentos asociados a s4 o s1. El comportamiento del ritmo cardiaco asociado al estado 3 también parece relativamente estable, distinguiéndose en este caso del estado 0 por su más elevado valor medio. Por último, es curioso constatar que en el estado 4, pese a no observarse la media más baja de ritmo cardiaco, hay una incidencia de observaciones tomadas en fase de sueño profundo significativamente mayor que en los restantes estados.

Tal y como comentábamos anteriormente, el modelo HMM ajustado también nos permite computar una cifra de verosimilitud logarítmica para una secuencia dada de valores de la variable observable (en este caso el ritmo cardiaco). Siguiendo un procedimiento adecuado, esto posibilita la identificación de noches en las que tienen lugar comportamientos anómalos de la serie temporal respecto a los patrones típicos a partir de los cuales se han ajustado los parámetros del modelo. Para ello, computaremos la verosimilitud logarítmica de subsecuencias no deslizantes (es decir, sin solapamiento) de longitud de 4 horas (considerando que estamos evaluando periodos nocturnos de 5 horas y 30 minutos de duración, cada subsecuencia englobará la mayor parte de una noche). La razón por la que se emplean secuencias no deslizantes – a diferencia de lo que sucederá con los otros dos modelos – se debe al coste computacional inasumible en cuanto a tiempo (y sin mejoras

apreciables en cuanto a calidad de los resultados) al emplearse ventanas deslizantes. Aplicando este procedimiento, a continuación se presenta una representación gráfica de la evolución de la verosimilitud logarítmica a lo largo del periodo de extracción de datos (de nuevo, seguimos con el sujeto ID1).

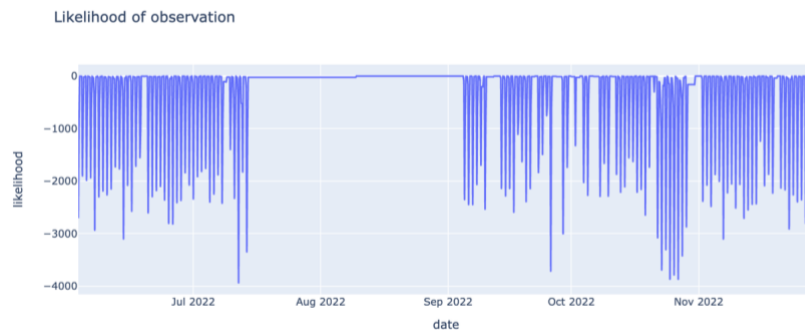


Figura 16: Verosimilitud logarítmica en ID1 para segmentos no deslizantes de 4 horas de duración.

Centrémonos de modo particular en el periodo posterior a septiembre de 2022. En este periodo, y a partir de una mera inspección visual, destaca de un modo particular un conjunto de noches ubicadas a finales de octubre (concretamente, del 22 al 28 de octubre), para las que vemos una caída muy notable en la verosimilitud respecto al rango de valores de verosimilitud observable en las noches restantes.

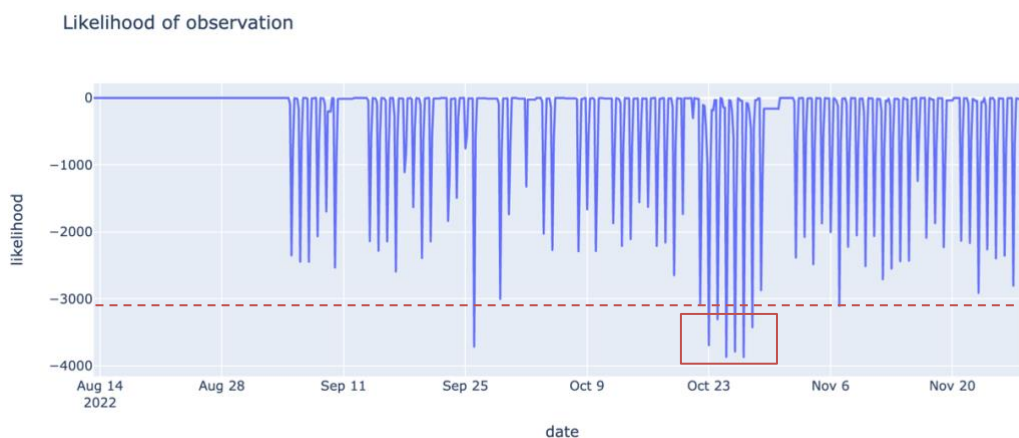


Figura 17: Detalle de verosimilitud logarítmica (ID1) para segmentos no deslizantes de 4 horas.

Tras identificar de manera no supervisada este conjunto de fechas como claramente atípicas, se pudo posteriormente verificar con el sujeto que en estas mismas fechas tuvo que trasladarse a una región con huso horario muy desfasado respecto al español. La caída en verosimilitud, por tanto, manifiesta las consiguientes alteraciones en la fisiología del sujeto durante el sueño fruto de la disrupción del su ritmo circadiano.

Resulta significativo a nuestro juicio que, únicamente siguiendo un criterio de detección de noches con picos de verosimilitud atípicamente baja, y sin ninguna otra variable que la del propio ritmo cardiaco durante el periodo de sueño, se ha podido identificar un conjunto de fechas en las que se da una coincidencia perfecta con un evento disruptivo como es la alteración del ciclo circadiano. Y todo ello, hemos de insistir, dentro de un paradigma puramente no supervisado. De por sí solo, este hallazgo ya apunta hacia la potencial validez de esta aproximación para identificar potenciales eventos de salud y efectuar un seguimiento del estado general del sujeto. Pero, como ahora se verá, no está exenta de problemas e inconvenientes.

En términos generales, la aproximación basada en inspección visual resulta por si sola suficiente cuando las anomalías son pocas y claramente distinguibles de los valores típicos de verosimilitud. Sin embargo, y como es el caso en este ejemplo específico (a excepción de los días que acabamos de describir, donde la caída en verosimilitud es especialmente significativa), cuando las potenciales anomalías empiezan a ser más numerosas y la frontera entre valores normales y valores atípicamente bajos es más ambigua, es preferible introducir un criterio objetivo que permita definir un umbral.

Por ello, partiendo de los valores de los picos de verosimilitud logarítmica de cada noche (es decir, el segmento de cada noche para el que la verosimilitud es más baja), si tomamos el percentil 90 como umbral de detección (representado en la figura 14 mediante una línea discontinua), podemos identificar hasta 11 noches anómalas, 6 de las cuales se correlacionan con ESD. Nótese que estos ESD son únicamente aquellos eventos que han podido ser identificados por el usuario al finalizar el periodo de prueba (el reconocimiento de ESD por parte del voluntario es **retrospectivo**, tal y como se clarificó anteriormente al exponer la

metodología). Dada la perfecta correlación entre una agrupación de noches atípicas y un evento disruptivo que tuvo lugar a lo largo de esas mismas fechas, resulta plausible que algunas de las restantes noches anómalas para las que no hay registrado ningún ESD puedan de hecho estar reflejando algún evento de salud que el sujeto no pudo recordar al finalizar el periodo, o no registró cuando tuvo lugar.

Hasta ahora, hemos estado computando la verosimilitud de subsecuencias de 4 horas de duración. Esta opción resulta a priori razonable porque nos permite obtener una evaluación de verosimilitud que tome en consideración la casi totalidad del periodo de sueño. En cualquier caso, se podrían definir subsecuencias de longitud temporal más reducida que permitan una mayor granularidad en nuestro análisis. A continuación se expone el resultado de este mismo análisis efectuado para subsecuencias de una hora de duración.

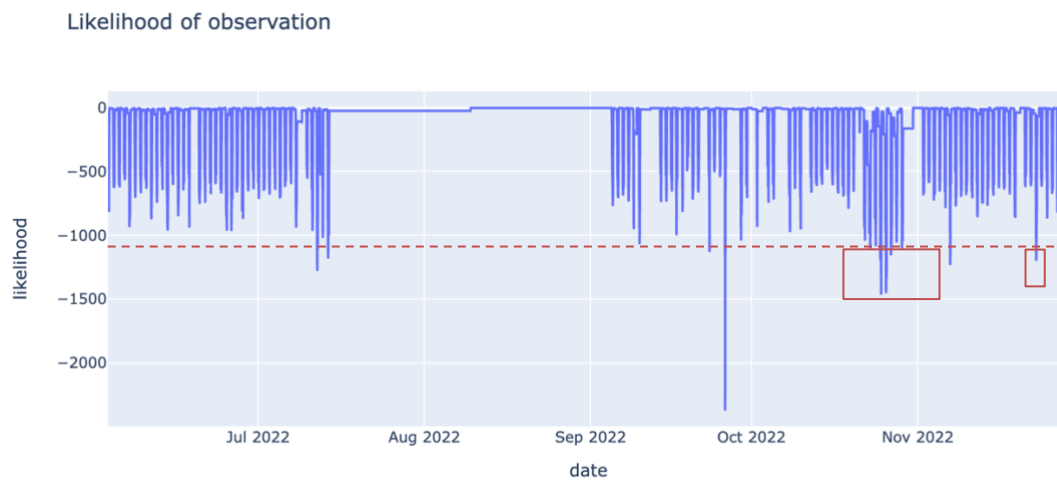


Figura 18: Verosimilitud logarítmica en ID1 para segmentos no deslizantes de 1 hora de duración.

Como vemos, el patrón general de valores de verosimilitud no cambia sustancialmente. Únicamente se pueden identificar una caída especialmente notable de la verosimilitud en la noche del 26 de septiembre, más pronunciada que la observada al evaluar la verosimilitud en subsecuencias de cuatro horas. Por otra parte, la escala de la propia verosimilitud

logarítmica es más reducida, pero esto no resulta sorprendente dado que se están evaluando sub-secuencias de menor longitud.

En cualquier caso, aplicando el criterio anterior, además de las noches correlacionadas con el cambio de huso horario (que vuelven a aparecer como atípicas), encontramos también una anomalía en la noche del 23 de noviembre. Aunque al realizar anteriormente la evaluación con subsecuencias de cuatro horas no contabilizamos esta fecha como atípica, sí se podía observar en cualquier caso que la cifra de verosimilitud se alejaba un poco de los valores más típicos y frecuentes. Pese a ello, la caída en verosimilitud no alcanzaba la magnitud necesaria para superar el umbral antes definido para contabilizar las fechas anómalas. Sin embargo, al emplear segmentos de una hora, el pico de caída de verosimilitud ahora sí entraría dentro del rango de valores que definimos como anómalos. Pues bien, esta fecha se corresponde con un episodio febril que se manifestó en el sujeto al contraer la enfermedad vírica Covid-19, cuyo periodo sintomático alcanzó su pico de intensidad ese mismo día. De nuevo, se ha podido identificar de manera no supervisada una desviación atípica respecto al comportamiento normal de la serie temporal, correlacionado con un ESD.

Este mismo procedimiento se replicó con los datos de los restantes sujetos. A continuación (Figura 24) se expone de manera resumida el análisis y caracterización de los estados HMM identificados tras el ajuste del modelo. Las gráficas ubicadas a la izquierda exponen para cada sujeto un breve segmento de la serie temporal de ritmo cardiaco, a fin de poder visualizar las variaciones en el comportamiento de la serie temporal entre el periodo de sueño y el periodo activo. El gráfico derecho muestra una comparativa de la distribución de valores de ritmo cardiaco para cada estado del HMM ajustado.

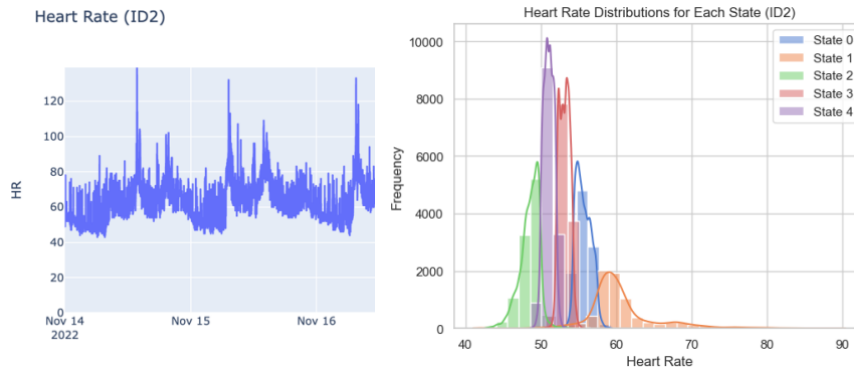


Figura 19: Caracterización de estados ocultos en ID2

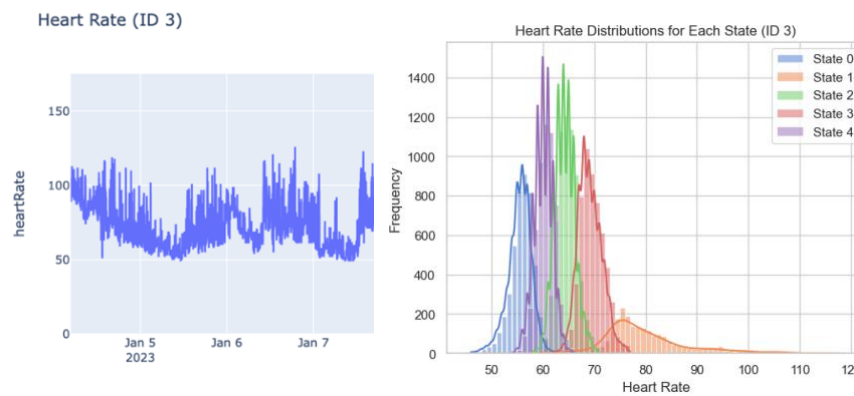


Figura 20: Caracterización de estados ocultos en ID3

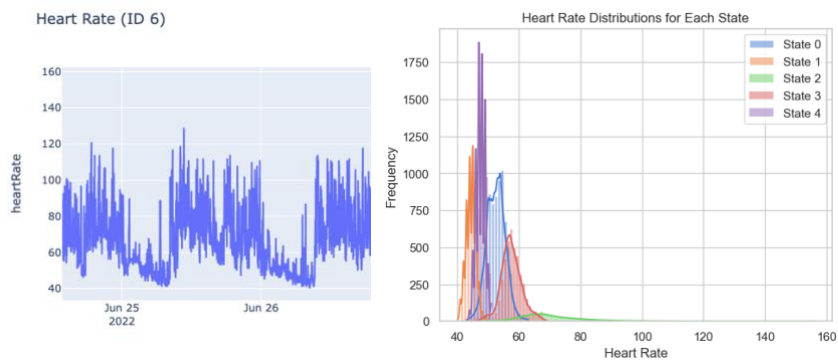


Figura 21: Caracterización de estados ocultos en ID6

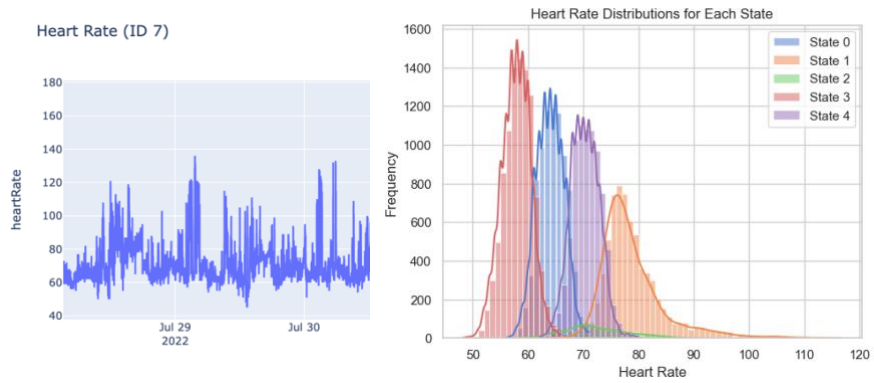


Figura 22: Caracterización de estados ocultos en ID7

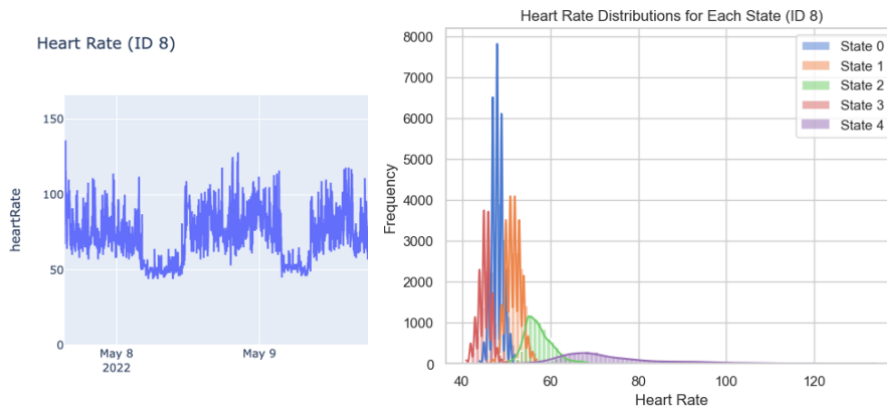


Figura 23: Caracterización de estados ocultos en ID8

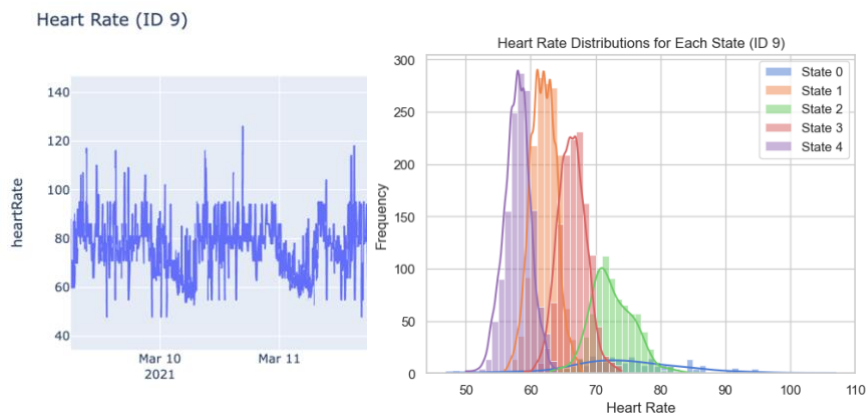


Figura 24: caracterización de estados ocultos en ID9

A modo de comentario general, se debe destacar que en casi todos los sujetos se ha obtenido consistentemente $n = 5$ como número óptimo de estados, aplicándose en todos ellos el mismo criterio seguido con ID1 para determinar este hiperparámetro. Por otra parte, también podemos constatar a partir de los histogramas superiores que, pese a la variabilidad inter-sujeto, la caracterización de los estados ocultos es muy similar en todos los casos: un estado menos frecuente para el cual el ritmo cardiaco tiende a exhibir valores elevados y una mayor desviación típica; y otros cuatro estados con mayor solapamiento entre sí en las distribuciones, y centrados en torno a medias progresivamente menos elevadas.

Únicamente en el caso de ID9 vemos una caracterización apreciablemente distinta al patrón general observable en los demás sujetos: posiblemente por la elevada cantidad de observaciones más extremas dentro del periodo de sueño, uno de los estados caracteriza en exclusiva estas observaciones, presentando una media mucho más elevada que los restantes estados y un rango intercuartílico muy amplio. La distribución de los restantes estados sigue la distribución típica en los demás sujetos, con la salvedad de que ahora el estado 0 y el 4 engloban observaciones que en otros casos vendrían asociadas a un estado distinto e intermedio entre estos dos.

Ajustados los modelos, se procedió a la identificación de anomalías obedeciendo al mismo criterio seguido con los resultados de verosimilitud en los datos de ID1: la identificación se realizará mediante inspección visual cuando las desviaciones respecto a los valores típicos de verosimilitud logarítmica sean muy pronunciadas. Para los casos en los que haya una mayor ambigüedad y la diferencia entre anomalía y no anomalía resulta más difícil de establecer visualmente, se fijará como umbral el percentil 90 de la distribución de picos de verosimilitud. La tabla que se presenta a continuación sintetiza los resultados de este análisis.

Sujeto	Segmentos de evaluación de 4h		Segmentos de evaluación de 1h	
	Noches anómalas	Anomalías correladas con ESD	Noches anómalas	Anomalías correladas con ESD
ID1	11	7	11	7
ID2	7	2	9	4
ID3	10	3	6	2
ID6	6	<i>Sin información</i>	3	<i>Sin información</i>
ID7	1	1	1	1
ID8	12	3	12	2

Tabla 2: Anomalías identificadas mediante verosimilitud logarítmica de segmentos.

Antes abordar y evaluar estos resultados, es necesario reincidir en el hecho de que aquellas fechas anómalas para las que no haya podido identificarse ningún ESD correlado no han de interpretarse como ‘falsos positivos’, entre otras razones, porque pueden estar reflejando un verdadero evento de salud en esa fecha no recordado por el sujeto. Nos movemos en todo momento dentro de un paradigma no supervisado a la hora de identificar fechas anómalas, apoyándonos en una caracterización previa de la serie temporal que capture sus parámetros típicos de comportamiento y permita identificar desviaciones respecto a los mismos. Al efectuar este posterior análisis, buscamos meramente constatar la existencia de correlaciones frecuentes entre ESD y caídas en la verosimilitud.

Entre los ESD cuya correlación con caídas en la verosimilitud se ha podido verificar encontramos picos de estrés intenso, episodios catarrales, bajadas en la calidad de sueño vinculadas con viajes, insomnio... En prácticamente todos los casos en los que se han verificado tres o más días sucesivos con valores de verosimilitud anómalos, se ha hallado una coincidencia de fechas con un ESD (tratándose típicamente de episodios gripales con fiebre alta, o alguna disrupción que impacte con fuerza los ritmos circadianos normales del sujeto). Por último, es interesante observar que para el sujeto ID8 una porción sustancial de las anomalías tienen lugar en la segunda mitad del periodo de medición (esto es visible tanto para evaluación de segmentos de una hora como de cuatro horas, aunque con una hora es aún más visible). Así mismo, el valor medio de verosimilitud logarítmica parece disminuir ligeramente respecto al que encontramos antes del mes de abril.

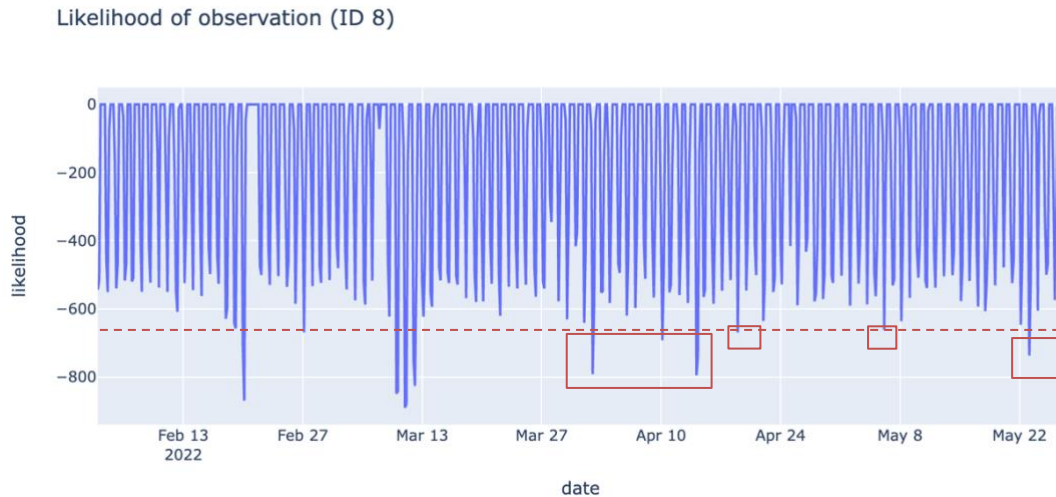


Figura 25: Posible cambio tendencial en ID8

Esto puede ser indicativo de un cambio estacional o tendencial en el comportamiento de la serie temporal, si bien solo un seguimiento más prolongado puede confirmar o desmentir esta conjetura. En términos generales, puesto que hay una pluralidad de casos en los que se ha podido identificar de manera no supervisada múltiples eventos de relevancia directa para este estudio (incluyendo episodios de fiebre moderada o alta, interrupciones circadianas o picos de estrés), la validez e interés de los modelos ocultos de Markov para nuestros objetivos queda acreditada. En los sucesivos apartados de esta memoria evaluaremos si otras aproximaciones pueden ampliar, reforzar o mejorar estos resultados preliminares, tanto individualmente como en conjunción con esta caracterización basada en HMM.

5.2 AUTOENCODERS

Tal y como se expuso en detalle en la sección segunda de esta memoria, los *autoencoders* son una clase de redes neuronales entrenadas para comprimir un vector de entrada en otro de inferior dimensionalidad (vector latente), y posteriormente reconstruir a partir de este vector latente la entrada original. Para ello, la red habrá de caracterizar internamente las correlaciones lineales y no lineales que se den en condiciones normales entre los atributos o campos del vector de entrada. Dado que la propia arquitectura de la red y la naturaleza de la

tarea para la que está entrenada, si aparecen observaciones cuyo comportamiento se desvía mucho del observable en el grueso de los datos de entrenamiento, la red tendrá mayores dificultades para retener la información tras la compresión y reconstruir la observación a partir de vector latente. Ello se manifestará en una subida en el error de reconstrucción.

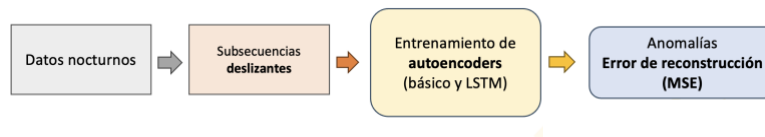


Figura 26: Metodología específica autoencoders.

Para abordar el problema específico de este proyecto – la identificación de anomalías en la serie temporal de ritmo cardiaco – entrenaremos dos variantes de autoencoder: una primera arquitectura básica basada en capas densas completamente conectadas, donde el *encoder* y el *decoder* son a todos los efectos dos perceptrones multicapa; y una segunda arquitectura basada en neuronas LSTM. La arquitectura basada en capas densas (*'feedforward'*) se concreta en la siguiente estructura de capas y neuronas:

- **Capa de entrada:** se emplearán ventanas de 180 muestras para pulseras Fitbit (15 minutos en términos de longitud temporal) o de 90 muestras para pulseras MiBand (lo cual equivale en términos de tiempo a ventanas de 1 hora y 30 minutos debido a la menor frecuencia de muestreo; en este sentido se ha buscado crear ventanas de longitud comparable a las empleadas con datos Fitbit en términos de número de muestras, pero procurando que el tamaño no llegue a suponer una duración temporal de ventana demasiado grande). Cada muestra ocupará una de las posiciones de entrada de la red.
- **Capas ocultas del encoder:** consta de una primera capa de 64 neuronas, una segunda capa de 32 neuronas, una tercera capa de 16 neuronas, y una última capa de 12 neuronas coincidente con la dimensión del vector latente (es decir, dicho vector será de longitud 12). En cada capa se aplicará un *dropout* del 10%, siendo esta una técnica

orientada a evitar el sobreajuste de la red consistente en desactivar aleatoriamente un 10% de las neuronas de cada capa durante el entrenamiento.

- **Capas ocultas del decoder:** sigue una arquitectura idéntica a la del encoder pero invertida: cuatro capas (con 12, 16, 32 y 64 neuronas, respectivamente), todas ellas con *dropout* del 10%.
- **Capa de salida:** finalmente, el decoder cuenta una última capa de salida en la que se reconstruye la ventana de 180 (o 90) muestras original.

En todas las neuronas se ha empleado la función de activación ReLU (siglas de *Rectified Linear Unit*), una de las funciones de activación más populares en el campo del aprendizaje profundo. Definida como $f(x) = \max(0, x)$, presenta ventajas importantes sobre otras alternativas, entre las que destacan principalmente su fácil diferenciación (la derivada de esta función es muy simple: es igual a 0 para valores de entrada negativos, y 1 para valores de entrada positivos, lo cual facilita enormemente el cálculo de gradientes durante la retropropagación); y su impacto positivo en la velocidad de convergencia durante el entrenamiento. Estas ventajas han conducido a su adopción generalizada para redes profundas, por lo que también la emplearemos con la arquitectura anterior.

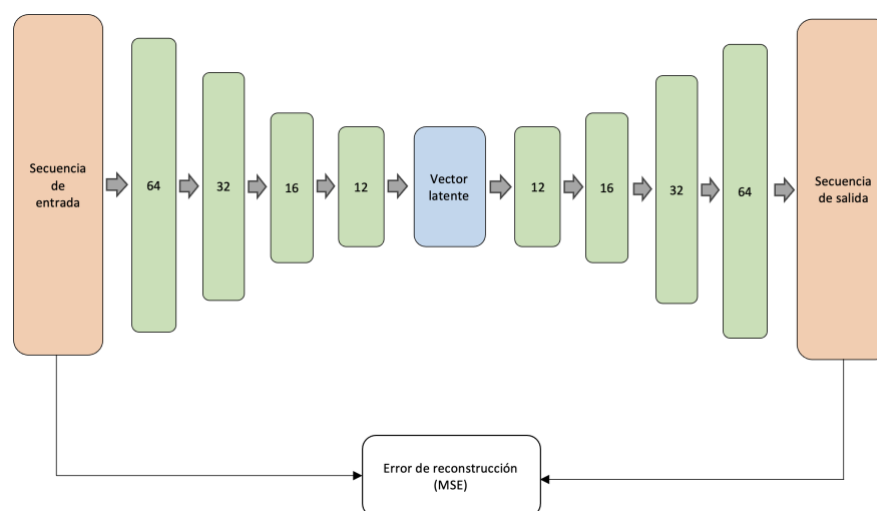


Figura 27: Arquitectura autoencoder 'feedforward'

Por otra parte, la arquitectura seguida para los autoencoders LSTM es la siguiente:

- **Capa de entrada:** en este caso, dado que se estarán empleando neuronas LSTM en el autoencoder, la capa de entrada estará constituida por 64 neuronas tanto para datos Fitbit como para datos MiBand. Pero nótese que ahora la interpretación de esta capa de entrada es muy distinta: en cada neurona se pasará la ventana deslizante completa, que de nuevo vuelve a estar constituida por 180 muestras para pulseras Fitbit y 90 muestras para pulseras MiBand. La naturaleza recurrente de las neuronas LSTM del encoder permitirá procesar secuencialmente esta ventana para finalmente condensar la información en un vector latente.
- **Capas ocultas del encoder:** consta de una primera capa LSTM de 64 neuronas, una segunda capa LSTM de 32 neuronas, y una tercera y última capa densa y *feedforward* de 8 neuronas cuya salida se corresponderá con el vector latente. De nuevo, en cada capa se aplicará un *dropout* del 10% para evitar el sobreajuste durante el entrenamiento.
- **Capas ocultas del decoder:** constituido por una capa LSTM de 32 neuronas, donde cada neurona recibe el vector latente transformado para que incorpore una dimensión secuencial temporal igual al tamaño de ventana de los datos originales (90 o 180, según los datos que estemos abordando). Básicamente se estarán introduciendo 90 o 180 repeticiones del vector latente para construir las ventanas. Esto es así porque las dimensiones del vector latente no tienen un sentido temporal o secuencial, mientras que las neuronas LSTM esperan datos secuenciales. Además, datos de salida que estamos tratando de reconstruir sí son secuenciales, por lo que esta transformación es necesaria. La capa de 32 neuronas viene seguida por una capa LSTM de 64 neuronas.

- **Capa de salida:** tras la capa de 64 neuronas LSTM, incorporamos una última capa densa de salida que se aplica de forma independiente sobre cada una de las ventanas deslizantes de la secuencia, reconstruyendo así la entrada.

Para esta arquitectura se usará como función de activación la tangente hiperbólica en lugar de la ReLU, debido a que el uso de la ReLU con neuronas LSTM presentó en esta arquitectura concreta un problema de gradientes explosivos que impedía la convergencia en entrenamiento. La naturaleza secuencial de los datos con los que estamos trabajando convierte a este segundo tipo de autoencoders en una opción particularmente interesante, pero también evaluaremos el desempeño de la arquitectura básica basada en capas densas *'feedforward'* por presentar esta una mayor simplicidad y menores costes de entrenamiento en términos computacionales y de tiempo.

A continuación se expone el procedimiento completo de preparación de datos y modelización seguido para ID1, los pasos seguidos fueron idénticos para los restantes sujetos. El procesamiento de los datos previo a la modelización constó de las siguientes fases:

1. **Normalización** de los datos de frecuencia cardiaca.
2. **Segmentación** de la serie temporal en subsecuencias deslizantes (es decir, a diferencia de las subsecuencias empleadas para computar la verosimilitud con los modelos HMM, en este caso sí hay solapamiento entre los segmentos). Estas ventanas deslizantes tendrán una longitud de 90 muestras para datos MiBand, y de 180 para datos FitBit, por las razones antes expuestas. El tamaño de ventana queda limitado de esta manera debido a que se estará trabajando con modelos recurrentes. Aún pese a tratarse de neuronas LSTM, ventanas excesivamente grandes no pueden evitar completamente el fenómeno de desvanecimiento de gradiente, por lo que en la práctica la red no podrá recordar muestras muy atrás en el tiempo.
3. Partición en *train* y *test*. Se reservaron las primeras 70% de observaciones para el ajuste del modelo, y las restantes 30% para su evaluación con nuevos datos. A

consecuencia de ello, únicamente identificaremos y evaluaremos fechas anómalas ubicadas dentro de este periodo de validación: el mayor interés del modelo radica en su capacidad para identificar desviaciones anómalas en datos nunca vistos. La razón por la que no seguimos un procedimiento similar de partición train-test con los modelos HMM responde a la naturaleza distinta de estos modelos; en la sección 4.1 se ofrece una explicación más detallada de las razones que justifican esta decisión. Los datos de test se dividieron a su vez en un conjunto de entrenamiento propiamente dicho, y un conjunto de validación que nos permita identificar cualquier signo de sobreajuste durante el entrenamiento.

De cara al entrenamiento, se definió como función de pérdida el error cuadrático medio (ECM, o MSE por sus siglas en inglés) de reconstrucción. A continuación se muestra una gráfica representando la evolución del MSE durante el entrenamiento, tanto para el conjunto de entrenamiento como para el de validación. El entrenamiento quedó limitado a 160 épocas debido a que, a partir de este punto, el valor de pérdida en validación comenzaba a ascender y a alejarse del MSE de *train*, fruto de un probable sobreajuste del modelo. En cualquier caso, la evolución de esta métrica para estas primeras 160 épocas y su pronunciado descenso respecto a los valores iniciales evidencia que el modelo ha logrado converger en torno a errores de reconstrucción aceptables.

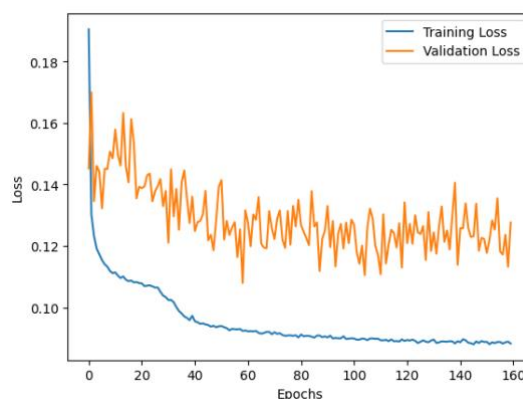


Figura 28: Evolución del error de reconstrucción durante el entrenamiento.

Tras el entrenamiento, el modelo ya ajustado se empleó para reconstruir los datos del conjunto de test y computar el MSE de reconstrucción asociado a cada subsecuencia deslizando. Tras ello, se marcaron como anómalos todos aquellos segmentos en los cuales este error de reconstrucción se alejase un mínimo de tres desviaciones típicas respecto a la media. Este criterio (criterio de tres sigmas, o 3σ) resultaba excesivamente restrictivo a la hora de clasificar como anómalas las cifras de verosimilitud logarítmica asignadas por los modelos HMM, pero a nuestro juicio es el más adecuado para identificar desviaciones atípicas del MSE, dada la naturaleza y significado diferente de esta métrica respecto a la verosimilitud logarítmica.

Aunque los autoencoders posibilitan una evaluación temporal más fina de los instantes temporales precisos en los que tienen lugar comportamientos atípicos del ritmo cardiaco (dado el enfoque de detección basado en ventanas deslizando), para mantener los resultados comparables con los obtenidos mediante caracterización HMM, nos interesará una detección a nivel de noche. Por ello, contabilizaremos como anómalas todas aquellas noches en las que se detecte alguna subsecuencia cuyo error de reconstrucción supere el umbral 3σ definido anteriormente. En la siguiente figura (Figura 29), se muestra el ritmo cardiaco durante el periodo de sueño para varias noches consecutivas. Aparecen sombreadas en rojo aquellas fechas en las que aparecen comportamientos atípicos conducentes a un error de reconstrucción superior a 3σ .

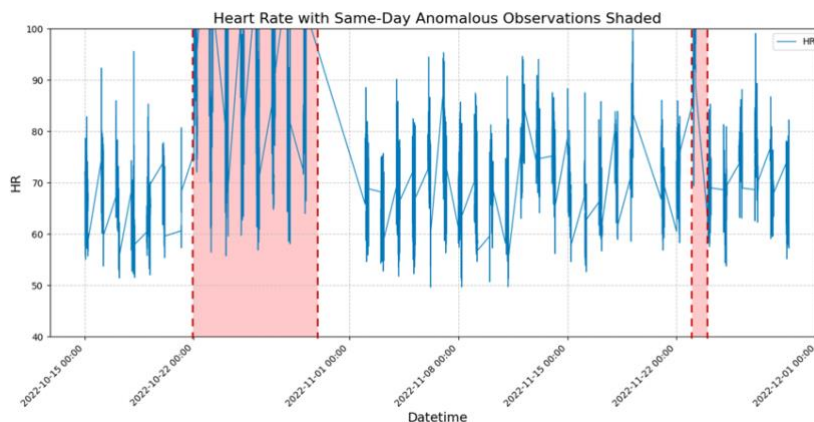


Figura 29: Registro de ritmo cardiaco de ID1 limitado al periodo nocturno (fechas atípicas sombreadas).

Tal y como vimos en la anterior sección, los días comprendidos entre el 22 y el 31 de octubre de 2022 (que aquí de nuevo aparecen marcados como atípicos) se correspondían con un viaje del sujeto a una zona horaria muy desfasada respecto al huso horario español. De nuevo, las fechas vinculadas a este evento de disrupción circadiana han podido ser identificadas de manera no supervisada con una aproximación basada en autoencoders. El 23 de noviembre, como también verificamos anteriormente, viene asociado con un episodio de fiebre provocada por la enfermedad Covid-19 en su fase sintomática.

Comprobemos ahora los resultados con el autoencoder LSTM. El gráfico bajo estas líneas (Figura 30) muestra la evolución del desempeño del modelo con los datos de train y validación a lo largo del entrenamiento. Esta vez se han fijado muchas menos épocas, por dos razones fundamentales: (1) la red ha mostrado signos de convergencia en una fase mucho más temprana del entrenamiento; y (2) el carácter recurrente de esta red obliga a que el entrenamiento se realice de forma secuencial, ralentizando mucho el proceso de ajuste.

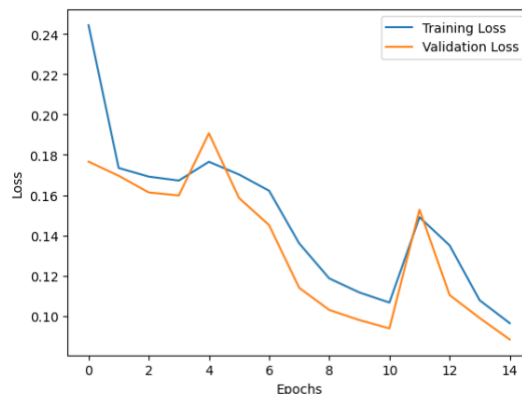


Figura 30: Evolución del error de reconstrucción durante el entrenamiento.

Pese al inconveniente de un mayor tiempo de entrenamiento, los errores de reconstrucción al final del mismo apuntan a una caracterización interna de la serie temporal claramente superior. Aplicando el mismo criterio de detección de anomalías descrito para la arquitectura

básica de autoencoder, las noches clasificadas como anómalas se muestran en la siguiente figura (Figura 31).

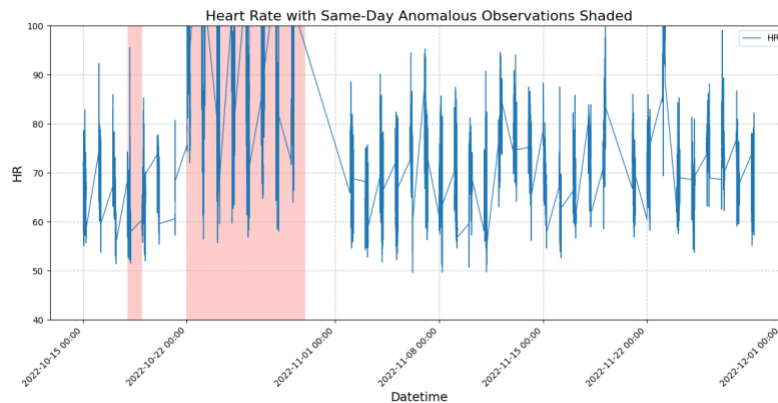


Figura 31: Registro de ID1 limitado al periodo nocturno (fechas atípicas sombreadas).

De nuevo, vuelven a identificarse como atípicos los días asociados al viaje, pero resulta interesante que con esta red, el error de reconstrucción para el 23 de noviembre ya no supere el umbral de anomalía. Por otro lado, aparece una nueva fecha – 18 de octubre – clasificada como anómala, si bien al limitar el sombreado anterior solo a las observaciones concretas en las cuales el error de reconstrucción ha superado el umbral de anomalía (en lugar de marcar la noche entera), observamos que el número de horas anómalas es muy reducido (Figura 32), hasta el punto de no ser visible en el propio gráfico. Al ser mejor la caracterización, posiblemente el ritmo cardiaco tenga que desviarse más sustancialmente de sus parámetros normales de comportamiento de forma que quede clasificado como atípico.

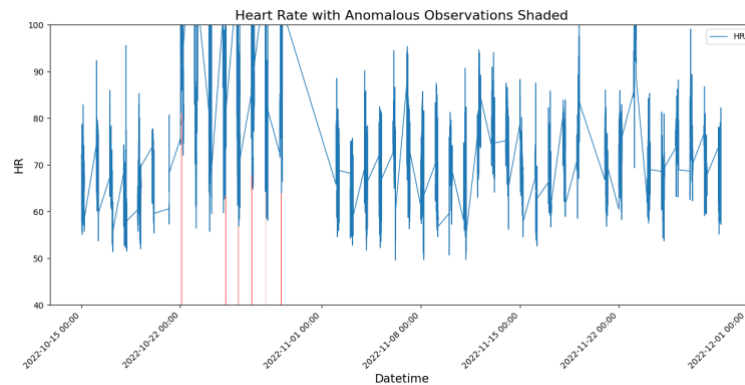


Figura 32: Registro para ID1 limitado al periodo nocturno, con segmentos atípicos sombreados.

Un modelo de este tipo podría ser de interés si buscamos capturar únicamente desviaciones claramente atípicas. Sin embargo, el mayor coste computacional de entrenamiento para redes LSTM, unido al hecho de que de cara a este proyecto resulta más relevante un modelo con mayor sensibilidad a cualquier indicio de desviación anómala, parece aconsejar el empleo de las versiones básicas de autoencoder con datos Fitbit.

Esta conclusión, sin embargo, no es necesariamente extensible a los datos MiBand, debido a la frecuencia de muestreo sustancialmente inferior para estas pulseras (1 minuto entre muestras) frente a la que encontramos en Fitbit (con un periodo intermuestral de solo 5 segundos). A consecuencia de ello, tenemos que (1) las ventanas deslizantes a partir de las cuales estaremos computando el error de reconstrucción son de longitud más reducida en términos de número de muestras, lo cual acelera notablemente el tiempo de entrenamiento de las redes LSTM, y (2) aún pese a tratarse de ventanas de menos muestras, estas subsecuencias comprenden un mayor intervalo temporal, por lo que al computar el error de reconstrucción para cada ventana se estará evaluando una mayor cantidad de minutos. Esto puede implicar que el modelo en su versión básica tenga más dificultades para reconstruir los datos.

Para evaluar el autoencoder más conveniente con este tipo de datos, tomaremos como ejemplo los resultados con los datos del sujeto ID8. Recuérdese que con datos procedentes de pulseras MiBand nos hemos visto obligados a usar ventanas más grandes en términos

temporales (1 hora y 30 minutos), precisamente por las razones que se acaban de exponer, e incluso así la secuencia es de menos muestras (90 muestras) que para el caso de ventanas de 15min con datos Fitbit. Por todo ello debemos hacer esta evaluación y no valen las conclusiones para la Charge 5.

Pese a esta diferencia, la fase de preparación y modelización únicamente cambia en lo siguiente: tomaremos para el conjunto de *test* el 30% inicial de muestras (en lugar del 30% final), debido a que como pudimos ver con los modelos HMM, uno de los eventos anómalos más distinguibles se ubica dentro de dicha porción inicial. Dado que tenemos este evento bien localizado, puede ser interesante comprobar si los autoencoders son capaces de identificar también este ESD. Las fechas anómalas detectadas con el autoencoder básico son las siguientes (Figura 33):

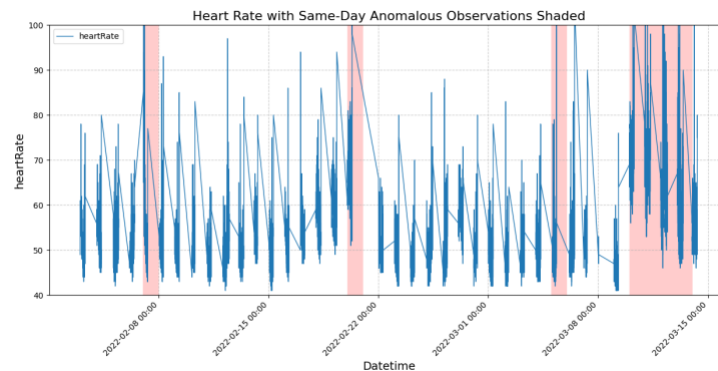


Figura 33: Registro de ID8 limitado al periodo nocturno (fechas atípicas sombreadas).

Y los resultados de detección para el autoencoder LSTM se muestran a continuación (Figura 34).

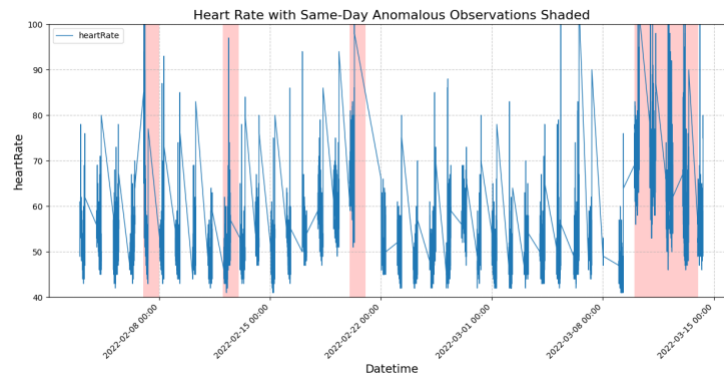


Figura 34: Registro de ID8 limitado al periodo nocturno (fechas atípicas sombreadas).

No hay diferencia sustancial; en cualquiera de los casos todas las fechas coincidentes con el ESD al que nos referíamos antes (episodio de fiebre alta entre el 9 y el 12 de marzo) son identificadas como anómalas. Ambos modelos también detectan comportamientos atípicos en las noches del 8 y el 20 de febrero, y el propio 20 de febrero también aparece como fecha notoriamente anómala de acuerdo con el modelo HMM ajustado para estos mismos datos, pese a que el sujeto no ha podido identificar un ESD coincidente con dicha fecha. El hecho de que múltiples modelos la estén señalizando como fuertemente atípica es indicativo de que, con toda probabilidad, exista alguna causa de tipo fisiológica o vinculada con la calidad de sueño que subyazca a este comportamiento anómalo de los datos, pese a que no haya quedado registrada por el sujeto. Si a todo ello le sumamos el hecho de que, al ser las ventanas de menos muestras, el entrenamiento del autoencoder LSTM para los datos MiBand 6 es más rápido, emplearemos estos mismos autoencoders para caracterizar las series de ritmo cardíaco extraídas de estas pulseras.

Pese a que las diferencias entre ambos modelos no son muy marcadas en cuanto a resultados, las razones antes expuestas para justificar esta evaluación y comparación entre las versiones básicas y LSTM (ventana de menos muestras pero con mayor longitud temporal) nos llevan a decantarnos por la versión LSTM para datos procedentes de las pulseras de marca MiBand. El procedimiento descrito anteriormente se ha aplicado uniformemente con los datos de los restantes sujetos, con la única salvedad de que en todos los casos distintos de ID8 se han tomado en torno al 30% final de muestras para la evaluación tras el entrenamiento. En la

siguiente tabla se sintetizan los resultados, comparándolos además con las fechas anómalas detectadas por los HMM en el mismo periodo de evaluación. Las anomalías detectadas por ambos métodos pueden ser especialmente relevantes, en tales casos es muy probable que haya una causa subyacente aun cuando no haya podido ser identificada.

<i>Sujeto</i>	<i>Dispositivo</i>	<i>Modelo</i>	<i>Noches anómalas en test (recuento total)</i>	<i>Noches anómalas en test coincidentes con ESD identificados</i>	<i>Noches anómalas en test detectadas también con HMM</i>
<i>ID1</i>	Charge 5	Básico	9	9 (100%)	7
<i>ID2</i>	Inspire 2	Básico	4	3 (75%)	1
<i>ID3</i>	MiBand 6	LSTM	4	1 (25%)	1
<i>ID6</i>	MiBand 6	LSTM	7	<i>Sin información</i>	3
<i>ID7</i>	MiBand 6	LSTM	3	3 (100%)	1
<i>ID8</i>	MiBand 6	LSTM	7	4 (57%)	4
<i>ID9</i>	MiBand 6	LSTM	2	<i>Sin información</i>	0

Tabla 3: Identificación de anomalías mediante autoencoders, comparativa con HMM.

A modo de comentario general, nótese que aunque el número de anomalías tiende a ser inferior al registrado con modelos HMM (la columna del extremo derecho de la tabla solo está contabilizando el número de anomalías que han sido identificadas tanto por los autoencoders como por modelos HMM, no el número total de anomalías HMM), la proporción de las mismas para las cuales se han podido identificar en estas mismas fechas eventos ESD es superior en el caso de los autoencoders. Como contraparte, y a diferencia de la caracterización HMM, es muy poco el conocimiento que podemos inferir respecto a cómo se ha caracterizado internamente el comportamiento típico de la serie temporal dentro de las redes. Son modelos poco interpretables y muy opacos, mientras que los modelos ocultos de Markov quedan definidos por un número de parámetros sustancialmente inferior (número de estados, probabilidades de transición entre estados y distribución de probabilidad de emisión para cada estado) a los cuales se les puede asignar una interpretación más directa y

comprensible; e incluso con un poco de análisis se les puede dar a los propios estados un sentido fisiológico, caracterizando el comportamiento de la serie temporal asociado a cada uno de ellos. Por ello, la detección de anomalías basada en autoencoders supone una interesante herramienta de apoyo para la identificación no supervisada de desviaciones atípicas, pero es preferible entender su papel complementario al análisis efectuado con HMM.

5.3 CLUSTERING

El término *clustering*, como vimos anteriormente, hace referencia a un conjunto de técnicas englobadas dentro de la familia de algoritmos de aprendizaje no supervisado, que busca particionar un conjunto de datos en grupos o *clusters* en base a su similitud. El *clustering* de series temporales aplica esta idea general para datos secuenciales, realizando la agrupación en base a similitudes estadísticas, patrones, tendencias y otros atributos temporales de las series a agrupar. En la sección 2.3.2 identificábamos tres grandes enfoques a la hora de abordar la agrupación de series temporales: *clustering* de series temporales completas, *clustering* de subsecuencias o segmentos, y *clustering* de puntos temporales. De estos tres enfoques, dado que estamos realizando un análisis y caracterización de los datos a nivel de individuo (por razón de la elevada variabilidad inter-sujeto que constatábamos al realizar el análisis exploratorio de los datos), el más interesante para nosotros es la clusterización de subsecuencias. Ello implica tener que realizar una fragmentación de la serie temporal de frecuencia cardíaca nocturna en ventanas deslizantes, para lo cual se aplicó el mismo procedimiento seguido con los autoencoders, fijando tamaños de ventana de 90 muestras para datos MiBand y de 180 muestras para datos Fitbit.

Generadas las subsecuencias, existen tres maneras de computar la similitud entre las mismas: en base a la similitud de las propias formas de las series temporales en bruto, en base a la distancia entre vectores de características de dimensionalidad muy inferior a la secuencia original, que sinteticen sus atributos más relevantes; o en base a la similitud o disimilitud entre modelos paramétricos ajustados a partir de los datos de cada secuencia. Aunque podría ser interesante aplicar este último enfoque usando HMM para parametrizar las secuencias,

ello podría complicar en exceso el proceso en su conjunto, y obligaría a usar ventanas más grandes para que el ajuste de parámetros se apoye en una cantidad suficiente de datos. El enfoque basado en vectores de características proporciona una importante ventaja en términos de simplicidad y coste computacional y, a diferencia de la aproximación basada en similitud de forma, definir una métrica de distancia adecuada resulta mucho más sencillo (en nuestro caso se usará la distancia euclidiana).

Estos vectores se podrían construir aplicando una técnica de reducción de dimensionalidad (por ejemplo, empleando los *encoders* entrenados anteriormente para mapear las subsecuencias a sus respectivos vectores latentes) pero, para facilitar la interpretabilidad de los resultados, construiremos los vectores a partir de una serie de atributos que sintetizen el comportamiento de la secuencia. Existen una infinidad de variables, tanto en el dominio del tiempo como en el de la frecuencia, que podrían emplearse para construir estos vectores. Tras ponderar las distintas opciones existentes, los atributos elegidos son los siguientes (Maximilian Christ et al., n.d.):

- **Media:** expresa la tendencia central de cada segmento. Su relevancia de cara a la caracterización se justifica en el hecho de que, tal y como se pudo constatar anteriormente, algunos de los episodios anómalos asociados con estados febriles tienden a concretarse en una elevación del valor medio de la frecuencia cardiaca.
- **Desviación típica:** medida de la dispersión de los datos en el segmento. Una desviación típica más alta reflejará una mayor volatilidad en el segmento en cuestión; por el contrario, una cifra más reducida señala un comportamiento más estable. Como pudimos observar, existen a lo largo del periodo nocturno segmentos de tiempo en los que el comportamiento de la serie es más estable, y segmentos en los que se verifica una mayor dispersión en los valores de frecuencia cardiaca. La predominancia de uno u otro comportamiento dentro de un segmento quedaría parametrizada mediante esta variable.
- **Mediana:** al igual que la media, expresa la tendencia central de las observaciones en cada segmento, pero es menos sensible a observaciones extremas.

- **Pendiente de la línea de regresión:** refleja la tendencia general de evolución de las observaciones del segmento. Una pendiente positiva indica que el segmento sigue una tendencia creciente, un valor negativo expresa lo contrario. Es por tanto un dato relevante para una caracterización completa del segmento
- **Número de cruces:** número de veces que la serie temporal (cada segmento es una sub-secuencia del registro completo de ritmo cardiaco, y por ende una pequeña serie temporal) cruza su valor mediano. Actúa como indicador del comportamiento oscilatorio del segmento, permitiendo identificar aquellos con fuertes oscilaciones.
- **Autocorrelación:** expresa el grado de similitud entre el segmento y una versión retardada del mismo, lo cual puede ayudar a parametrizar y distinguir aquellos segmentos que exhiben patrones repetitivos frente a aquellos que no los exhiben.

Se aplica la fórmula

$$\frac{1}{(n-l) \cdot \sigma^2} \cdot \sum_{t=1}^{n-l} (X_t - \mu)(X_{t+l} - \mu)$$

Donde n representa la longitud de la subsecuencia X , σ^2 denota la varianza, μ representa la media, y l denota el retardo ('lag') expresado en términos de número de muestras, que en este caso se ha fijado a 1.

- **Entropía de la muestra:** mide la complejidad o irregularidad del segmento, permitiendo distinguir entre aquellos más estables y aquellos con un comportamiento aparentemente más caótico.
- **Sesgo:** mide la asimetría en la distribución de los valores de ritmo cardiaco para cada segmento, permitiendo de este modo una parametrización más completa de dicha distribución en cada vector de características (junto con las cifras de media, mediana y desviación típica).

Para cada subsecuencia, se extrajeron estos atributos y con ellos se construyeron los vectores de características sobre los que se aplicaría el algoritmo de *clustering*. Tras construir estos vectores, se efectuó una partición train-test usando el mismo criterio de partición empleado con los autoencoders: el 70% inicial de observaciones se empleó para la estimación del número óptimo de *clusters* y el entrenamiento; el 30% restante se reservó para validación. Por último, dadas las muy distintas escalas de cada una de las anteriores variables, para poder emplear la distancia euclidiana como métrica de disimilitud entre vectores se hubo de normalizar todas las variables.

Para efectuar la agrupación propiamente dicha, se aplicó el algoritmo de *k-means* (Hartigan et al., 1979). Se trata de un método de *clustering* muy popular, que busca particionar el conjunto de datos en un número *k* de grupos, donde *k* sería un hiperparámetro cuyo valor se ha de prefijar antes del entrenamiento. El algoritmo comienza con una inicialización aleatoria o heurística de los centroides de los clústeres y, a continuación, alterna recursivamente entre dos pasos: asignación, donde se asignan los puntos de datos a los *clusters* representados por sus centroides más cercanos; y actualización, donde se recalculan los centroides como la media de los puntos de datos dentro de cada *cluster* resultante tras aplicar el paso de asignación. La convergencia se alcanza cuando las posiciones de los centroides se estabilizan o se cumple un criterio de parada especificado antes del entrenamiento. El algoritmo cuenta con una complejidad de tiempo lineal que lo vuelve muy eficiente en términos computacionales, es fácilmente escalable y produce resultados interpretables.

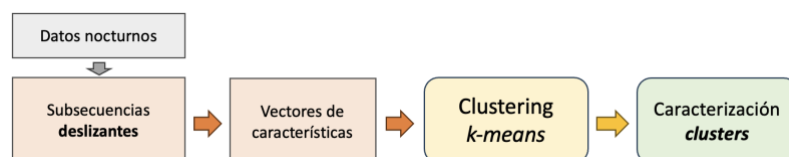


Figura 35: Metodología completa clustering.

A continuación se describe el procedimiento completo seguido con los datos del sujeto ID1. Para determinar el número óptimo de *clusters*, se recurrió al método del codo (Figura 36).

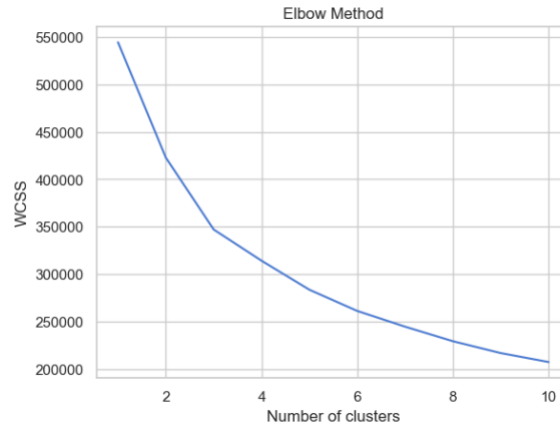


Figura 36: Evolución de la suma de cuadrados intra-cluster según el número de clusters.

En la gráfica no se visualiza ningún punto concreto para el que se observe una clara suavización y aplanamiento en la tasa de decremento de la suma de cuadrados intra-*cluster* (WCSS, o inercia), si bien sí se puede apreciar una cierta estabilización en torno a los 3-5 *clusters*. Por ello, se optó por $k = 4$ *clusters*. Los resultados de la clusterización se sintetizan en la siguiente tabla (Tabla 4).

<i>Cluster</i>	<i>Coef. Silhouette en train</i>	<i>Coef. Silhouette en test</i>	<i>% obs. train</i>	<i>% obs. test</i>
0	0.138326	0.031574	21.06 %	39.14 %
1	0.215918	0.219430	24.87 %	26.04 %
2	0.215679	0.250908	39.04 %	25.15 %
3	0.164607	0.165703	15.01 %	9.65 %

Tabla 4: Resumen de resultados tras la aplicación de clustering.

Los coeficientes de *Silhouette* nos proporcionan una medida del nivel de cohesión interna de cada *cluster* y separación respecto a los demás *clusters* (Ashutosh Bhardwaj, 2020). El coeficiente se mueve en un rango comprendido entre 1 y -1, donde un valor alto implica que la distancia promedio entre observaciones dentro del *cluster* es pequeña, y la distancia promedio entre observaciones de dicho *cluster* y las de otros *clusters* es muy grande. Por el contrario, un valor bajo implica que las observaciones de dicho *cluster* están más dispersas, o el *cluster* está muy cerca de grupos vecinos. Un valor negativo puede darse cuando la clusterización es muy pobre, y muchas de las observaciones pertenecientes a dicho *cluster* guardan mayor similitud con observaciones de otros grupos que con las del suyo propio.

A partir de la tabla podemos ver que los clústeres 1 y 2 presentan un mayor nivel de cohesión interna y una separación más definida respecto a los restantes datos; así mismo, parecen más estables y generalizables, pues sus tamaños relativos y coeficientes de *Silhouette* mantienen una relativa consistencia. Por otro lado, el coeficiente de *Silhouette* del *cluster* 0 experimenta una importante caída en test, y su tamaño relativo prácticamente se dobla. En el *cluster* 3 también observamos una diferencia notable en el tamaño relativo de test respecto al de train, si bien en este caso el coeficiente de *Silhouette* mantiene un valor muy próximo para ambos conjuntos. Considerando el valor de *Silhouette* tan bajo que presenta súbitamente el *cluster* 0 en los datos de test, parece probable que sea en éste *cluster* donde hayan quedado agrupados los segmentos anómalos o que más se desvían de los patrones típicos de comportamiento de la serie, si bien también se plantean preguntas sobre su capacidad de generalización. Se hace necesario efectuar un análisis en mayor profundidad sobre el comportamiento de las subsecuencias dentro de cada grupo.

A continuación se expone una sucesión de histogramas mostrando las distribuciones para cada *cluster* de algunos de los atributos más sencillos de interpretar (media, desviación típica, número de cruces por el valor mediano, y pendiente de regresión para el segmento). Téngase en cuenta que los histogramas representan las distribuciones para todos los datos, y no únicamente para los conjuntos de entrenamiento o evaluación.

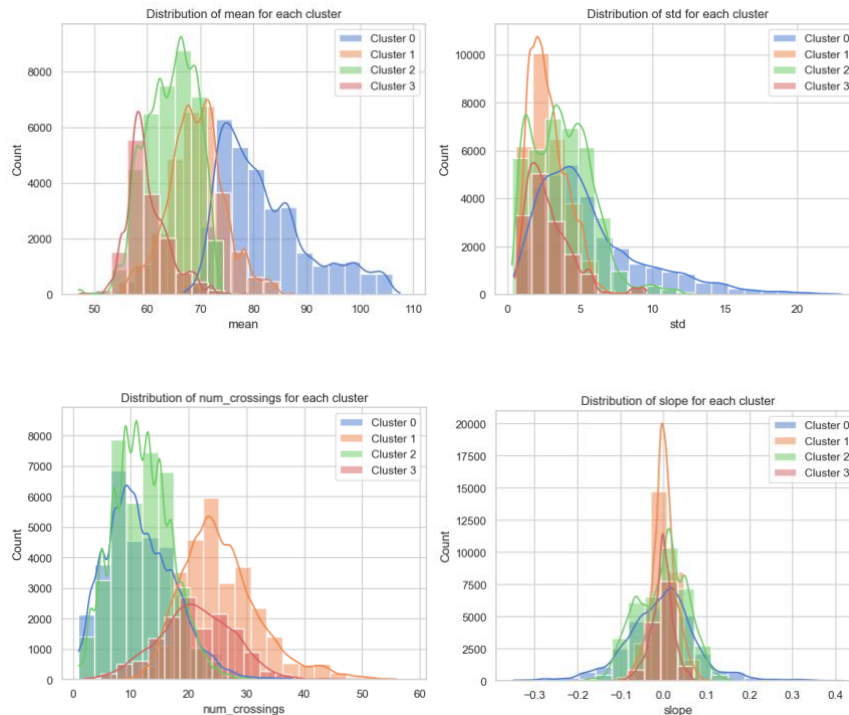


Figura 37: Resumen de clustering para ID1

Como puede verse, el *cluster* 0 (recordemos, el que presentaba un coeficiente de Silhouette particularmente bajo en test) es también el que cuenta con los segmentos de media más alta. Parte de la explicación del bajo valor de su coeficiente de Silhouette debe encontrarse en el hecho de que su distribución de ritmo cardiaco medio presenta una fuerte dispersión, englobando todos aquellos segmentos con valores más extremos, contando así con un rango muy amplio que se extiende desde las 70 hasta las 100 pulsaciones por minuto. Dentro de este mismo *cluster* se incluyen también los segmentos con mayor desviación típica, y con las pendientes de regresión más pronunciadas (tanto negativas como positivas).

Por todo ello, resulta llamativo que sea en cambio el *cluster* 1 el que agrupe a los segmentos con un mayor carácter oscilatorio (esto es, con mayor número de cruces por el valor mediano). La media de cruces para el *cluster* 3 también es curiosamente alta, pese a ser el *cluster* cuyos segmentos cuentan con las medias más bajas y pendientes de regresión casi nulas, esto último refleja un comportamiento tendencial de las subsecuencias bastante estable, pese a su carácter simultáneamente más oscilatorio. En la tabla 7 pueden encontrarse

los centroides de cada *cluster* para todas las dimensiones del vector de características. A la luz de estas últimas observaciones, no sorprende que los segmentos de los *clusters* 1 y 3 sean los que presentan una mayor entropía y menor autocorrelación.

<i>Clust.</i>	<i>mean</i>	<i>std</i>	<i>median</i>	<i>slope</i>	<i>Num crossings</i>	<i>autocorrelation</i>	<i>entropy</i>	<i>skewness</i>	<i>abs energy</i>
0	79.84	5.09	79.96	-0.0058	11.25	0.92	0.33	-0.14	1.16e+06
1	68.37	2.87	68.59	0.0029	25.92	0.75	0.70	-0.27	8.47e+05
2	63.83	3.77	63.54	-0.0007	11.87	0.91	0.33	0.24	7.40e+05
3	59.46	2.66	58.97	-0.0034	21.34	0.72	0.49	1.90	6.40e+05

Tabla 5: Centroides de los clusters

A modo de comparativa entre la caracterización obtenida mediante la aplicación de *clustering* de secuencias temporales y la caracterización derivada de los HMM, en la siguiente tabla se presenta la distribución de los estados ocultos dentro de cada *cluster*.

<i>Estado oculto HMM</i>	<i>Estado 0</i>	<i>Estado 1</i>	<i>Estado 2</i>	<i>Estado 3</i>	<i>Estado 4</i>
<i>Cluster</i>					
0	2.44 %	8.65 %	38.96 %	14.86 %	35.05 %
1	15.85 %	26.13 %	1.70 %	38.21 %	18.08 %
2	38.70 %	29.48 %	0.56 %	24.12 %	7.11 %
3	65.84 %	24.03 %	0.43 %	8.04 %	1.63 %

Tabla 6: Proporción de observaciones de cada cluster pertenecientes a cada estado oculto HMM.

En el *cluster* 0 vemos una predominancia de los estados ocultos 2 y 4 a partes iguales, con un 74% de las observaciones de dicho *cluster* pertenecientes a uno de estos dos estados. Recordemos que estos estados se caracterizaban principalmente por contar con las distribuciones de ritmo cardiaco de media más alta, el estado 4 presentaba además una incidencia comparativamente mayor de observaciones tomadas estando el sujeto despierto (aún dentro del periodo de sueño, es decir, despierto pese a estar tratando de dormir). Con todo, lo más llamativo a nuestro juicio es el hecho de que cerca de la totalidad de observaciones pertenecientes al estado 2 han quedado englobadas en el *cluster* 0, con una incidencia mínima en los demás *clusters*. Esto puede apuntar al hecho de que, posiblemente, la caída en el nivel de agregación de este *cluster* en test derive de una mayor incidencia del estado oculto 4 en los datos de evaluación, que han quedado agrupados en el *cluster* 1 por ser éste el más similar pese a que las observaciones que realmente estaban siendo caracterizadas en el *cluster* 0 tendían a pertenecer al estado 1. También llama la atención la elevada proporción de observaciones del *cluster* 3 pertenecientes al estado 2, casi dos tercios del total. A este estado pertenecían las observaciones con valores más bajos de frecuencia cardiaca, por lo que es interesante constatar que pese a ello, una proporción apreciable de las observaciones de dicho estado oculto se caractericen por manifestar un comportamiento más oscilatorio.

Por último, queda por verificar la hipótesis planteada anteriormente respecto a la alta probabilidad de que en el *cluster* 0 estuviesen agrupados los segmentos más atípicos de la serie. La siguiente tabla muestra el porcentaje de segmentos anómalos detectados por el autoencoder que caen dentro de cada *cluster*. Nos apoyamos concretamente en la detección basada en autoencoders porque estos modelos se han entrenado con las mismas subsecuencias que las empleadas para la clusterización.

Cluster	% del total de anomalías detectadas por el autoencoder que caen dentro del cluster
0	100%
1	0%
2	0%
3	0%

Tabla 7: Proporción de anomalías en cada cluster.

La hipótesis se confirma con contundencia: todas las subsecuencias anómalas caen dentro del *cluster 0*. Con todo, el *cluster 0* es bastante grande en test, por lo que una aproximación basada en *clustering* no es suficiente para aislar anomalías del resto de datos. De todas formas, sí nos ha permitido caracterizar algunos de los rasgos generales que estas anomalías tienden a presentar, infiriéndolas a partir de los del propio *cluster* en el que están agrupadas. A continuación se expone una versión resumida de los resultados para los sujetos restantes tras aplicar el mismo procedimiento.

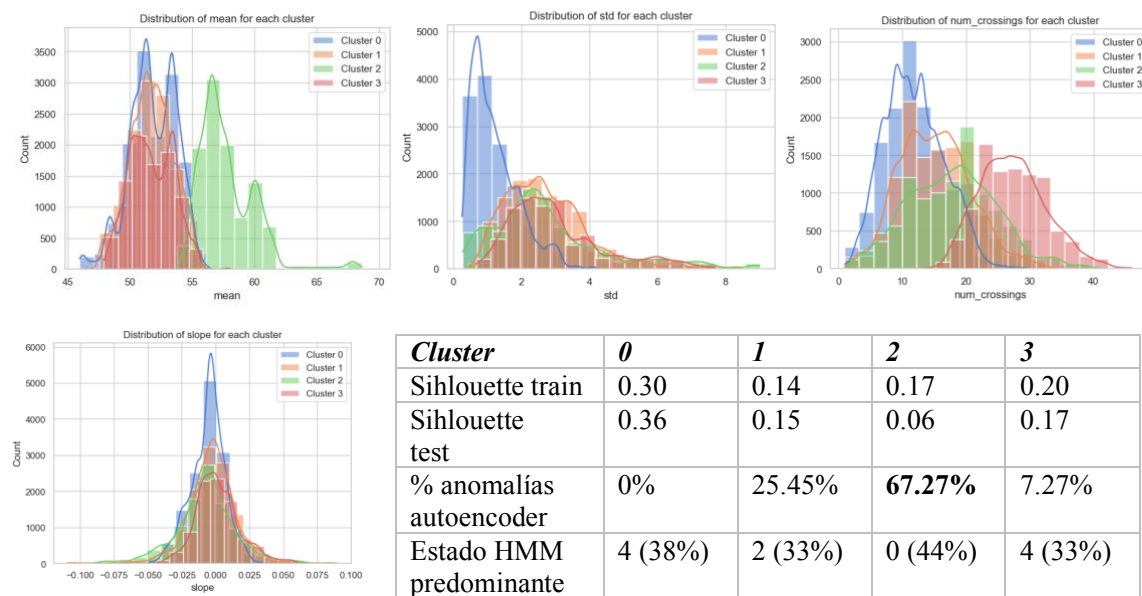


Figura 38: Resumen de clustering para ID2

Como vemos, la clusterización para ID2 nos ofrece unos resultados muy similares a los obtenidos con los datos del sujeto ID1: un *cluster* (en este caso el 2) mucho más distanciado de los demás en cuanto a la distribución de promedios de ritmo cardiaco a nivel de segmento, que agrupa la mayor parte de las anomalías identificadas de manera independiente con los autoencoders (en este caso un 67.27% del total); un *cluster* (en este caso el 3) con valores medios de ritmo cardiaco bajos pero un alto número de cruces por la mediana; y un *cluster* grande con alto coeficiente de agregación, medias de frecuencia cardiaca bajas, más estable

y con segmentos de tendencia casi plana. Es curioso que en el *cluster 2*, el que agrupa la mayor parte de anomalías, el estado oculto HMM predominante sea de nuevo el que presenta la segunda media más alta (y no el que contaba con las observaciones más extremas, véase la caracterización de estados HMM expuesta en la sección 4.1 para este sujeto).

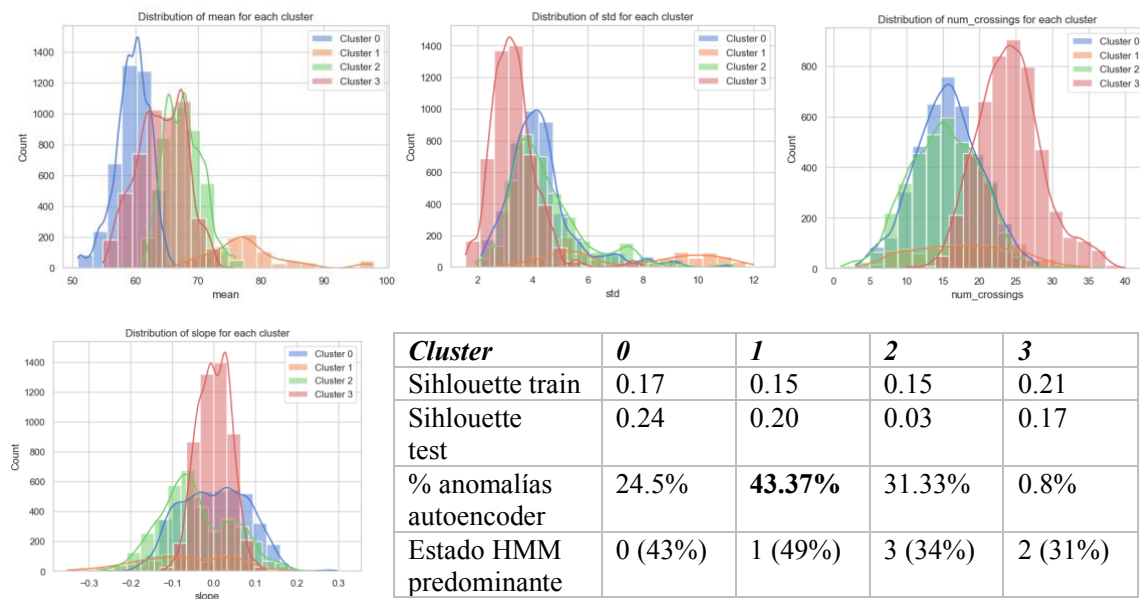


Figura 39: Resumen de clustering para ID3

La clusterización arroja para este sujeto resultados un poco distintos, si bien observamos patrones que nuevamente se repiten: el *cluster* que engloba a la mayor parte de anomalías (el 1) es el de valores medios más elevados y pendientes más fuertes; y uno de los *clusters* de frecuencia cardiaca media más baja (*cluster 3*) presenta por otra parte un fuerte comportamiento oscilatorio y simultáneamente pendientes de regresión próximas a 0, además de valores muy bajos de desviación típica. Ha de tenerse en cuenta, por otra parte, que los datos de este sujeto se tomaron con pulseras MiBand, por lo que las subsecuencias que se están clusterizando tienen menores muestras pero una mayor longitud temporal. Lo más reseñable lo encontramos en el hecho de que los segmentos del *cluster 1* (recordemos, el más atípico) en su mayoría presentan pendientes de regresión negativas.

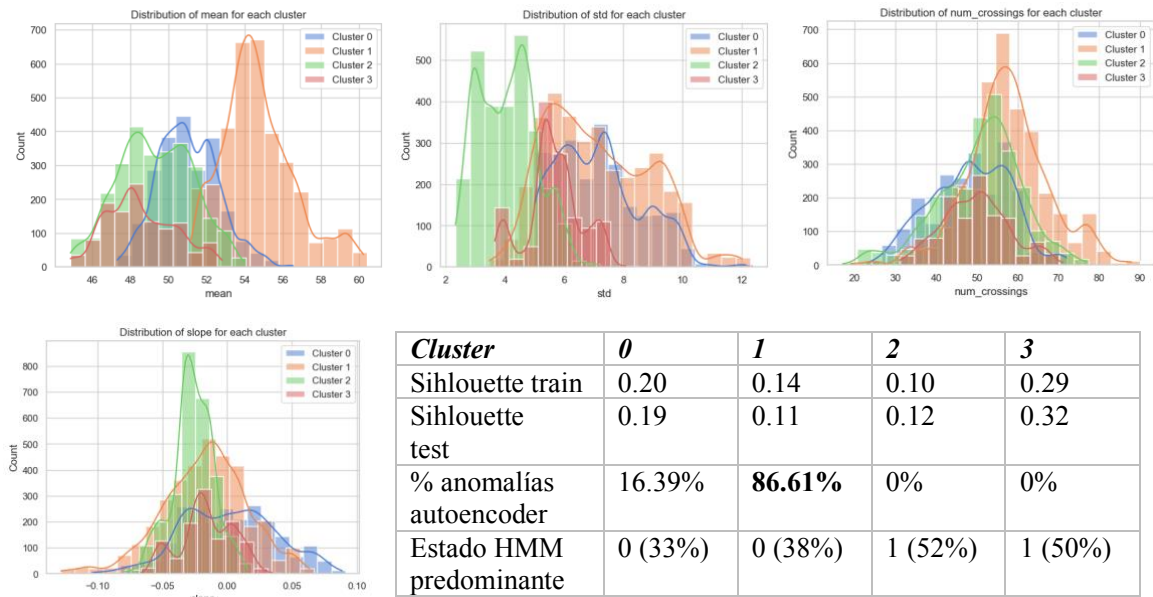


Figura 40: Resumen de clustering para ID6

Con ID6 (Figura 40) se repite el patrón general: todas las anomalías quedan agrupadas en el *cluster* de frecuencia cardiaca media más alta (*cluster* 1). Resulta curioso, sin embargo, que en este caso no se observe una bajada tan abrupta en el coeficiente de Silhouette en los datos de test. Por lo demás, no hay diferencias muy sustanciales respecto a los anteriores ejemplos.

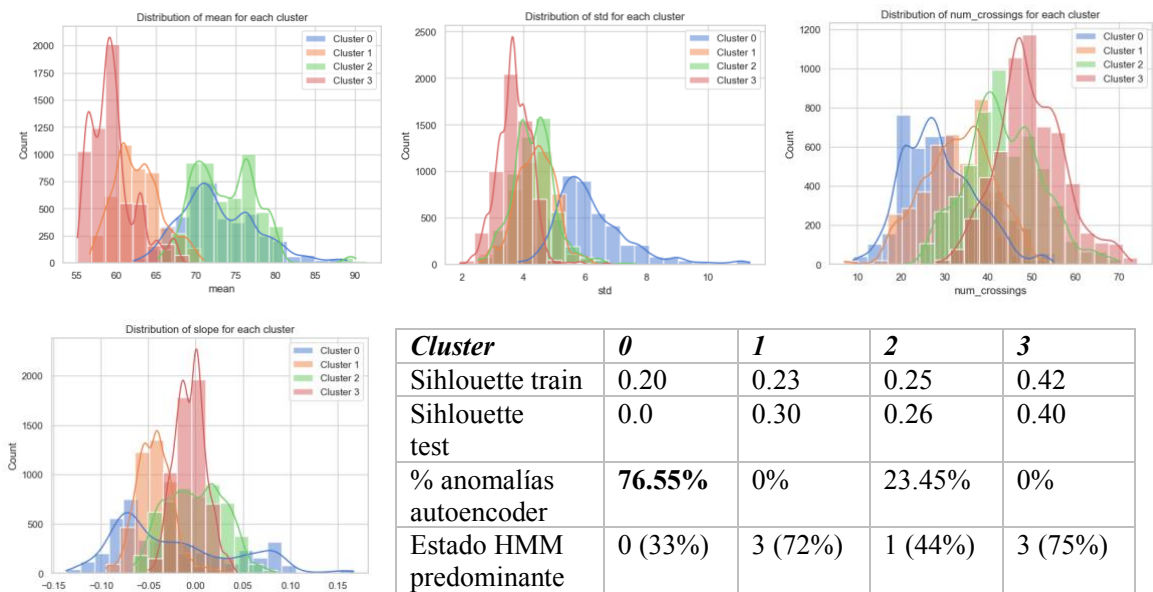


Figura 41: Resumen de clustering para ID7

Una vez más, las subsecuencias anómalas de ID7 se concentran en su mayor parte en los dos *clusters* – los *clusters* 0 y 2, concretamente – de frecuencia cardiaca media más alta (gráfica superior izquierda de la Figura 41). Los segmentos agrupados en el *cluster* 2 también presentan valores altos de ritmo cardiaco, pero se distinguen del *cluster* 0 en su dispersión notablemente menor. Las secuencias del *cluster* 0 son también las de mayor pendiente (tanto positiva como negativa), repitiéndose también en esto el patrón general que vemos con otros sujetos. Por último, como es de esperar (por el hecho de contener las subsecuencias atípicas), el nivel de agregación del *cluster* 0 en test cae hasta niveles cercanos a 0.

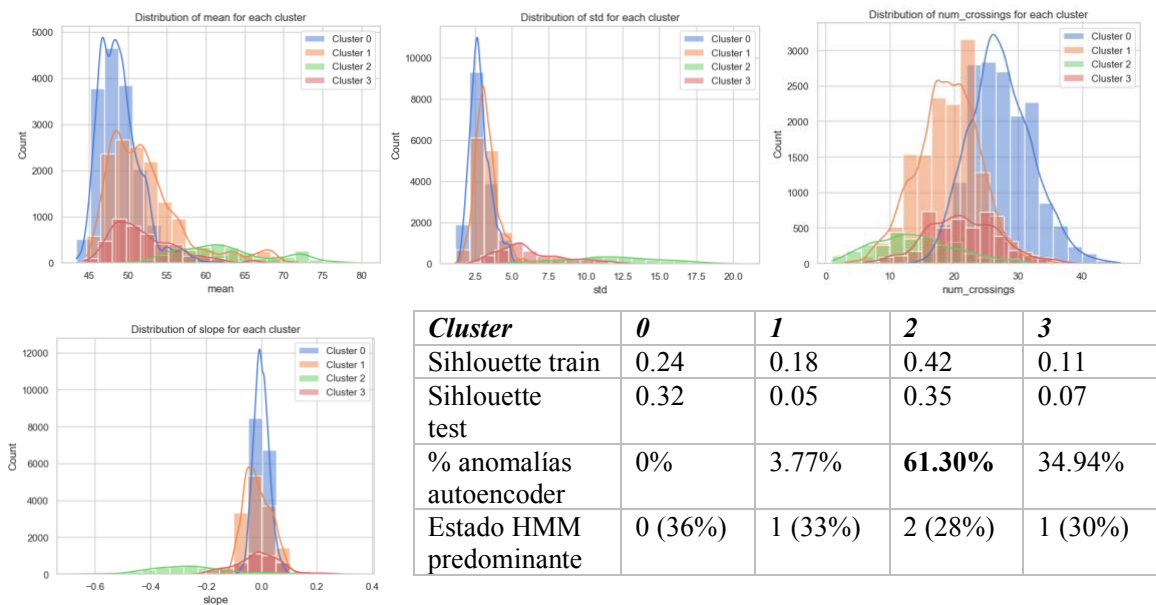


Figura 42: Resumen de clustering para ID8

Los resultados para ID8 resultan muy similares a los obtenidos con ID7 en lo que se refiere a los *clusters* propiamente dichos y el comportamiento de las distintas variables dentro de cada *cluster*. Las anomalías parecen concentrarse en los *clusters* 2 y 3, que considerados en conjunto son muy similares al *cluster* 0 de ID7.

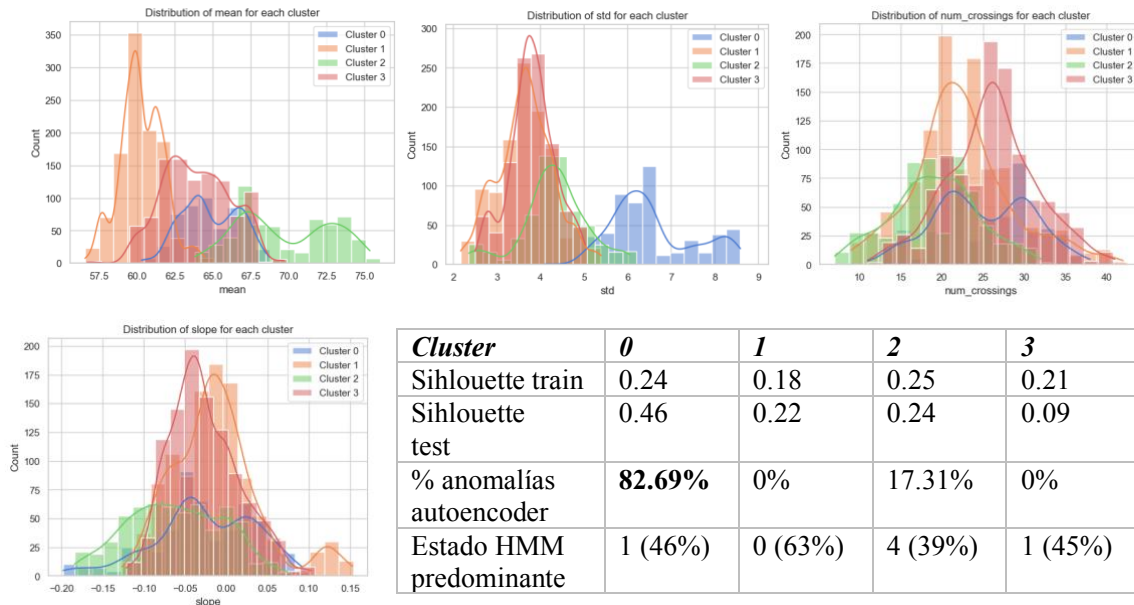


Figura 43: Resumen de clustering para ID9.

Por último, los resultados de clusterización para ID9 los podemos examinar sobre estas líneas. Sus resultados son los que más se alejan de los restantes voluntarios, y en general manifiestan un comportamiento más disperso, oscilatorio e inestable de la serie temporal durante el sueño. Con todo, sí hay algunas similitudes que se repiten: el *cluster* 0 (en el que se agrupan la mayor parte de anomalías) contienen algunos de los segmentos de pendiente más extrema, y su distribución de valores promedio de ritmo cardiaco está desplazada hacia valores altos (si bien no es el *cluster* con promedios más elevados). Su mayor divergencia respecto a los restantes *clusters* la encontramos en la distribución de los valores de desviación típica de sus segmentos: presenta un grado muy superior de dispersión respecto a los restantes *clusters*. En todo caso, se trata de resultados bastante idiosincráticos que reflejan el comportamiento globalmente atípico del ritmo cardiaco de este sujeto respecto al de los demás voluntarios. Son, por tanto, poco generalizables al resto de la muestra, lo cual por otra parte subraya la importancia de haber seguido una aproximación metodológica basada en el análisis y modelización a nivel de individuo.

Capítulo 6. CONCLUSIONES

6.1 *LIMITACIONES EN LA METODOLOGÍA Y ALCANCE DEL ESTUDIO*

Acabada la exposición y análisis de resultados, y antes de realizar una valoración global de los mismos, se expondrán brevemente las principales limitaciones identificadas en el estudio. La discusión de estas limitaciones es de suma importancia para poder valorar adecuadamente el alcance de estos resultados y su interpretación dentro del contexto general en el que se inserta este trabajo investigativo.

6.1.1 DISCUSIÓN DE LIMITACIONES EN LA MUESTRA DE VOLUNTARIOS

A pesar de las consideraciones defendidas en la sección 3.1.2, el tamaño de la muestra no deja de ser un factor limitante, dado que puede no captar todo el espectro de variabilidad en el comportamiento de la frecuencia cardiaca. Otra limitación la encontramos en la ligera infrarrepresentación de mujeres, y en la falta de uniformidad en la distribución de voluntarios según la edad, si bien ambas radican en última instancia en el carácter reducido de la muestra. Por ello, futuras líneas de investigación deberán trabajar con muestras más amplias buscando una distribución más equilibrada de las variables demográficas de control, de forma que puedan validar nuestros hallazgos. A pesar de estas limitaciones, nuestra muestra está constituida por un grupo suficientemente heterogéneo de individuos y ha permitido explorar a fondo la caracterización individualizada de la frecuencia cardiaca para patrones de comportamiento muy diverso. Los conocimientos adquiridos en este estudio pueden servir de base para futuras líneas de investigación con muestras más amplias y equilibradas que permitan validar y ampliar nuestros hallazgos. Entre tanto, las razones antes expuestas y la naturaleza exploratoria de la investigación nos llevan a considerar este tamaño como suficiente para cubrir los objetivos del proyecto y alcanzar conclusiones con suficiente respaldo experimental.

6.1.2 DISCUSIÓN DE LIMITACIONES EN EL PROCEDIMIENTO DE ADQUISICIÓN DE DATOS

En lo relativo a la metodología seguida para la extracción de los datos, el principal punto de mejora lo encontramos en el procedimiento de registro de ESD. El planteamiento seguido durante el periodo de extracción de muestras dejaba recaer exclusivamente sobre los voluntarios la tarea de registro de ESD, recabándose posteriormente la información de manera retrospectiva mediante una entrevista con los voluntarios. Se sugirió a los sujetos mantener un registro periódico de estos eventos de salud, especialmente los más importantes, para facilitar posteriormente dicha tarea de evaluación retrospectiva; o al menos preservar suficiente información para que al final del periodo pudiesen identificarse y ubicarse temporalmente. Sin embargo, durante la recopilación de datos quedo evidenciado que esta forma de proceder resultó poco adecuada.

Se pueden identificar tres grandes deficiencias en el proceso: por un lado, (1) las instrucciones no fueron suficientemente precisas en lo relativo a exactamente qué eventos se consideraban relevantes para el proyecto y por ende se debían registrar. Así mismo, (2) el propio procedimiento de registro periódico era poco sistemático y en la práctica no fue muy seguido por muchos de los voluntarios. Ello condujo a que la mayoría de los ESD tuviesen que ser identificados y ubicados temporalmente en base a lo que los sujetos recordaban, resultando en que únicamente se identificasen aquellos de mayor impacto (por ejemplo, episodios de fiebre de varios días de duración). Esto nos lleva a la tercera limitación: (3) el procedimiento dependía en exceso de la propia iniciativa de los voluntarios.

A modo de propuesta de cara a futuras líneas de investigación, se podrían solventar simultáneamente estos tres problemas mediante el uso de cuestionarios diarios en los que se solicite al sujeto una valoración subjetiva de su estado de salud para ese día. El cuestionario habría de ser breve, y podrían emplearse escalas visuales analógicas (EVA) comprendidas entre el 0 y el 10 en las que el voluntario evalúe alguna dimensión concreta de su salud (nivel de estrés, calidad de sueño auto percibida, sensación de descanso...) además de variables binarias para registrar viajes a zonas de huso horario distinto, episodios de fiebre, o cualquier

otro evento similar que admita respuestas de tipo sí/no. Estos cuestionarios habrían de ser accesibles a través del móvil para facilitar su relleno. Existen en el mercado de aplicaciones de salud algunas opciones que incorporan cuestionarios de naturaleza similar a la que se acaba de describir, por lo que se recomienda explorar dichas posibilidades antes de confeccionar una herramienta propia.

6.2 VALORACIÓN CRÍTICA DE LOS RESULTADOS

El objetivo central de este trabajo es, en síntesis, el de caracterizar y modelar el comportamiento normal del ritmo cardíaco para una cohorte de voluntarios suficientemente diversa en términos demográficos, analizar los distintos patrones de comportamiento capturados mediante esta caracterización, y detectar desviaciones respecto al comportamiento normal que pudieran asociarse con algún tipo de anomalía de salud. A la luz de los resultados expuestos y analizados en la sección cuarta de esta memoria, hemos visto que (1) se ha podido identificar, segregar y analizar separadamente distintos comportamientos y patrones de la serie temporal, e identificar algunos de los cambios en estos patrones cuando aparecen comportamientos anómalos. Aun cuando los modelos de Markov se han constituido en la herramienta principal de nuestro análisis, ha sido el análisis combinado del comportamiento de la frecuencia cardíaca para cada estado oculto y la examinación exhaustiva del comportamiento de subsecuencias en la clusterización lo que nos ha permitido cubrir plenamente este primer punto básico.

Por otra parte, la caracterización interna (no interpretable) efectuada mediante el ajuste de autoencoders, y el cómputo de verosimilitud logarítmica de subsecuencias a partir de los modelos de Markov ajustados, han permitido conjuntamente identificar desviaciones atípicas en el comportamiento de la serie. En este sentido, (2) se han podido verificar múltiples instancias de correlación entre estas desviaciones anómalas identificadas de manera no supervisada, y eventos de salud o de disrupción del ritmo circadiano, validando la hipótesis de que un paradigma no supervisado basado en la caracterización de los comportamientos normales de ritmo cardíaco y la identificación de desviaciones atípicas respecto a estos comportamientos permite realizar un seguimiento de la evolución del estado

de salud basada exclusivamente en esta variable biométrica. Además, en al menos un caso hemos podido identificar también un cambio más tendencial en el comportamiento de la frecuencia cardiaca, si bien la interpretación médica precisa excede el marco de este proyecto. Con todo, la detección basada en autoencoders parece más adecuada como método sistemático de identificación de anomalías, aunque ha de concebirse como herramienta de apoyo a un modelo base más transparente (por ejemplo, los HMM).

A modo de comentario final, se ha de recalcar que la totalidad del análisis se apoya en la caracterización del ritmo cardiaco durante el sueño. Este enfoque podría ampliarse integrando otras variables biométricas; sin embargo, el interés de esta variable concreta radica en la facilidad de su adquisición mediante sensores de fotopleletismografía con una precisión aceptable, siendo de hecho una prestación básica en prácticamente todas las pulseras de actividad disponibles en el mercado, incluyendo los modelos más accesibles en términos económicos. Es esto último lo que confiere especial relevancia a la aproximación se-guida y los resultados obtenidos. En cualquier caso, se deja planteada como posible línea de avance sobre estos resultados, la posibilidad de un análisis integrado de múltiples marcadores biométricos.

6.3 CONCLUSIÓN

Pese a las limitaciones discutidas en las secciones anteriores, todos los objetivos planteados inicialmente se han logrado cubrir, implementándose en el proceso un procedimiento sistematizado para el tratamiento de los datos y la construcción de los modelos (el código se proporciona en el enlace incluido en el anexo). Con todo, estos resultados no son aún accionables, y todavía se requiere una mayor profundización en algunos de los hallazgos expuestos para poder integrar esta forma de aproximación al seguimiento del estado de salud en una aplicación o herramienta de nivel comercial en torno a la cual se pueda desarrollar una idea de negocio. En la sección 5.1 se han planteado algunas mejoras en el procedimiento para poder avanzar en este sentido. Entre tanto, este trabajo se posiciona como una contribución de alcance académico que acredita la viabilidad de un enfoque de seguimiento

de salud basado exclusivamente en la caracterización del ritmo cardiaco durante el periodo de sueño, sentando las bases para futuras investigaciones.

Capítulo 7. BIBLIOGRAFÍA

- Aghabozorgi, S., Seyed Shirخورshidi, A., Ying Wah, T., 2015. Time-series clustering – A decade review. *Inf Syst* 53, 16–38. <https://doi.org/10.1016/j.is.2015.04.007>
- Altini, M., Amft, O., 2016. HRV4Training: Large-scale longitudinal training load analysis in unconstrained free-living settings using a smartphone application, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 2610–2613. <https://doi.org/10.1109/EMBC.2016.7591265>
- Altini, M., Dunne, L.E., 2021. What’s Next For Wearable Sensing? *IEEE Pervasive Comput* 20, 87–92. <https://doi.org/10.1109/MPRV.2021.3108377>
- Altini, M., Plews, D., 2021. What Is behind Changes in Resting Heart Rate and Heart Rate Variability? A Large-Scale Analysis of Longitudinal Measurements Acquired in Free-Living. *Sensors* 21, 7932. <https://doi.org/10.3390/s21237932>
- Arbo, J.E., Lessing, J.K., Ford, W.J.H., Clark, S., Finkelsztein, E., Schenck, E.J., Sharma, R., Heerdt, P.M., 2020. Heart rate variability measures for prediction of severity of illness and poor outcome in ED patients with sepsis. *Am J Emerg Med* 38, 2607–2613. <https://doi.org/10.1016/j.ajem.2020.01.012>
- Arocha Rodulfo, J.I., 2019. Sedentarism, a disease from xxi century. *Clínica e Investigación en Arteriosclerosis (English Edition)* 31, 233–240. <https://doi.org/10.1016/j.artere.2019.04.001>
- Ashutosh Bhardwaj, 2020. Silhouette Coefficient [WWW Document]. Towards Data Science. URL <https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c> (accessed 6.19.23).
- Bashkirtsev, O., Sagan, V., Gaevska, V., Zimba, O., 2021. BIOLOGICAL AGE ESTIMATION BASED ON HEART RATE VARIABILITY: A PILOT STUDY.

- Proceedings of the Shevchenko Scientific Society. Medical Sciences 65.
<https://doi.org/10.25040/ntsh2021.02.21>
- Chalmers, T., Hickey, B.A., Newton, P., Lin, C.-T., Sibbritt, D., McLachlan, C.S., Clifton-Bligh, R., Morley, J., Lal, S., 2021. Stress Watch: The Use of Heart Rate and Heart Rate Variability to Detect Stress: A Pilot Study Using Smart Watch Wearables. *Sensors* 22, 151. <https://doi.org/10.3390/s22010151>
- Cho, S.-B., Park, H.-J., 2003. Efficient anomaly detection by modeling privilege flows using hidden Markov model. *Comput Secur* 22, 45–55. [https://doi.org/10.1016/S0167-4048\(03\)00112-3](https://doi.org/10.1016/S0167-4048(03)00112-3)
- Choi, J., Cha, W., Park, M.-G., 2020. Declining Trends of Heart Rate Variability According to Aging in Healthy Asian Adults. *Front Aging Neurosci* 12. <https://doi.org/10.3389/fnagi.2020.610626>
- Cosoli, G., Spinsante, S., Scalise, L., 2020. Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. *Measurement* 159, 107789. <https://doi.org/10.1016/j.measurement.2020.107789>
- Dunn, J., Runge, R., Snyder, M., 2018. Wearables and the medical revolution. *Per Med* 15, 429–448. <https://doi.org/10.2217/pme-2018-0044>
- Eatock, D., Guidi, C.F., Kunz, V., MARGARAS, V., SCHOLZ, N., UMBACH, G., ZAMFIR, I., 2017. Demographic outlook in the European Union 2017. European University Institute.
- Garavaglia, L., Gulich, D., Defeo, M.M., Thomas Mailland, J., Irurzun, I.M., 2021. The effect of age on the heart rate variability of healthy subjects. *PLoS One* 16, e0255894. <https://doi.org/10.1371/journal.pone.0255894>

- Ghassempour, S., Girosi, F., Maeder, A., 2014. Clustering Multivariate Time Series Using Hidden Markov Models. *Int J Environ Res Public Health* 11, 2741–2763. <https://doi.org/10.3390/ijerph110302741>
- Guerville, F., De Souto Barreto, P., Ader, I., Andrieu, S., Casteilla, L., Dray, C., Fazilleau, N., Guyonnet, S., Langin, D., Liblau, R., Parini, A., Valet, P., Vergnolle, N., Rolland, Y., Vellas, B., 2019. Revisiting the hallmarks of aging to identify markers of biological age. *J Prev Alzheimers Dis* 1–9. <https://doi.org/10.14283/jpad.2019.50>
- Hartigan, J.A., Wong, M.A., others, 1979. A k-means clustering algorithm. *Appl Stat* 28, 100–108.
- James, P.T., Leach, R., Kalamara, E., Shayeghi, M., 2001. The Worldwide Obesity Epidemic. *Obes Res* 9, 228S–233S. <https://doi.org/10.1038/oby.2001.123>
- Joshi, S.S., Phoha, V. V., 2005. Investigating hidden Markov models capabilities in anomaly detection, in: *Proceedings of the 43rd Annual Southeast Regional Conference - Volume 1*. ACM, New York, NY, USA, pp. 98–103. <https://doi.org/10.1145/1167350.1167387>
- Kang, S.-J., Jung, S.I., 2020. Age-Related Morbidity and Mortality among Patients with COVID-19. *Infect Chemother* 52, 154. <https://doi.org/10.3947/ic.2020.52.2.154>
- Karaman, S., Benois-Pineau, J., Dovgalecs, V., Mégret, R., Pinquier, J., André-Obrecht, R., Gaëstel, Y., Dartigues, J.-F., 2014. Hierarchical Hidden Markov Model in detecting activities of daily living in wearable videos for studies of dementia. *Multimed Tools Appl* 69, 743–771. <https://doi.org/10.1007/s11042-012-1117-x>
- Kawamoto, R., Nazir, A., Kameyama, A., Ichinomiya, T., Yamamoto, K., Tamura, S., Yamamoto, M., Hayamizu, S., Kinosada, Y., 2013. Hidden Markov model for analyzing time-series health checkup data, in: *MEDINFO 2013*. IOS Press, pp. 491–495.

- Kingma, D.P., Welling, M., 2019. An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning* 12, 307–392. <https://doi.org/10.1561/22000000056>
- KRAUS, W.E., JANZ, K.F., POWELL, K.E., CAMPBELL, W.W., JAKICIC, J.M., TROIANO, R.P., SPROW, K., TORRES, A., PIERCY, K.L., 2019. Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health. *Med Sci Sports Exerc* 51, 1206–1212. <https://doi.org/10.1249/MSS.0000000000001932>
- Lee, K.H., Byun, S., 2023. Age Prediction in Healthy Subjects Using RR Intervals and Heart Rate Variability: A Pilot Study Based on Deep Learning. *Applied Sciences* 13, 2932. <https://doi.org/10.3390/app13052932>
- Li, J., Pedrycz, W., Jamal, I., 2017. Multivariate time series anomaly detection: A framework of Hidden Markov Models. *Appl Soft Comput* 60, 229–240. <https://doi.org/10.1016/j.asoc.2017.06.035>
- Li, Z., Wei, Z., Yue, Y., Wang, H., Jia, W., Burke, L.E., Baranowski, T., Sun, M., 2015. An Adaptive Hidden Markov Model for Activity Recognition Based on a Wearable Multi-Sensor Device. *J Med Syst* 39, 57. <https://doi.org/10.1007/s10916-015-0239-x>
- Lim, W.K., Davila, S., Teo, J.X., Yang, C., Pua, C.J., Blöcker, C., Lim, J.Q., Ching, J., Yap, J.J.L., Tan, S.Y., Sahlén, A., Chin, C.W.-L., Teh, B.T., Rozen, S.G., Cook, S.A., Yeo, K.K., Tan, P., 2018. Beyond fitness tracking: The use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. *PLoS Biol* 16, e2004285. <https://doi.org/10.1371/journal.pbio.2004285>
- Lippi, G., Plebani, M., 2013. Biomarker research and leading causes of death worldwide: a rather feeble relationship. *Clin Chem Lab Med* 51. <https://doi.org/10.1515/cclm-2013-0210>
- Luong, M.-T., Le, Q. V., Sutskever, I., Vinyals, O., Kaiser, L., 2015. Multi-task Sequence to Sequence Learning.

- Mariano, L.J., Poore, J.C., Krum, D.M., Schwartz, J.L., Coskren, W.D., Jones, E.M., 2015. Modeling strategic use of human computer interfaces with novel hidden Markov models. *Front Psychol* 6. <https://doi.org/10.3389/fpsyg.2015.00919>
- Masci, J., Meier, U., Cireşan, D., Schmidhuber, J., 2011. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. pp. 52–59. https://doi.org/10.1007/978-3-642-21735-7_7
- Maximilian Christ et al., n.d. tsfresh package documentation [WWW Document]. Blue Yonder GmbH. URL https://tsfresh.readthedocs.io/en/latest/api/tsfresh.feature_extraction.html (accessed 6.19.23).
- Mendez, M.O., Matteucci, M., Castronovo, V., Strambi, L.F., Cerutti, S., Bianchi, A.M., 2010. Sleep staging from Heart Rate Variability: time-varying spectral features and Hidden Markov Models. *Int J Biomed Eng Technol* 3, 246. <https://doi.org/10.1504/IJBET.2010.032695>
- Nemati, S., Ghassemi, M.M., Ambai, V., Isakadze, N., Levantsevych, O., Shah, A., Clifford, G.D., 2016. Monitoring and detecting atrial fibrillation using wearable technology, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 3394–3397. <https://doi.org/10.1109/EMBC.2016.7591456>
- Ng, A., others, 2011. Sparse autoencoder. *CS294A Lecture notes* 72, 1–19.
- Panagiotou, M., Zlatintsi, A., Filntisis, P.P., Roumeliotis, A.J., Efthymiou, N., Maragos, P., 2022. A comparative study of autoencoder architectures for mental health analysis using wearable sensors data, in: 2022 30th European Signal Processing Conference (EUSIPCO). IEEE, pp. 1258–1262. <https://doi.org/10.23919/EUSIPCO55093.2022.9909697>

- Ponikowski, P., Anker, S.D., AlHabib, K.F., Cowie, M.R., Force, T.L., Hu, S., Jaarsma, T., Krum, H., Rastogi, V., Rohde, L.E., Samal, U.C., Shimokawa, H., Budi Siswanto, B., Sliwa, K., Filippatos, G., 2014. Heart failure: preventing disease and death worldwide. *ESC Heart Fail* 1, 4–25. <https://doi.org/10.1002/ehf2.12005>
- Quer, G., Radin, J.M., Gadaleta, M., Baca-Motes, K., Ariniello, L., Ramos, E., Kheterpal, V., Topol, E.J., Steinhubl, S.R., 2021. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med* 27, 73–77. <https://doi.org/10.1038/s41591-020-1123-x>
- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286. <https://doi.org/10.1109/5.18626>
- Radin, J.M., Quer, G., Ramos, E., Baca-Motes, K., Gadaleta, M., Topol, E.J., Steinhubl, S.R., 2021. Assessment of Prolonged Physiological and Behavioral Changes Associated With COVID-19 Infection. *JAMA Netw Open* 4, e2115959. <https://doi.org/10.1001/jamanetworkopen.2021.15959>
- Radin, J.M., Wineinger, N.E., Topol, E.J., Steinhubl, S.R., 2020. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. *Lancet Digit Health* 2, e85–e93. [https://doi.org/10.1016/S2589-7500\(19\)30222-5](https://doi.org/10.1016/S2589-7500(19)30222-5)
- Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y., 2011. Contractive auto-encoders: Explicit invariance during feature extraction, in: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. pp. 833–840.
- Sano, A., Picard, R.W., 2013. Stress Recognition Using Wearable Sensors and Mobile Phones, in: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, pp. 671–676. <https://doi.org/10.1109/ACII.2013.117>
- Shin, G., Jarrahi, M.H., Fei, Y., Karami, A., Gafinowitz, N., Byun, A., Lu, X., 2019. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A

- systematic literature review. *J Biomed Inform* 93, 103153.
<https://doi.org/10.1016/j.jbi.2019.103153>
- Smyth, P., 1996. Clustering sequences with hidden Markov models. *Adv Neural Inf Process Syst* 9.
- Tison, G.H., Singh, A.C., Ohashi, D.A., Hsieh, J.T., Ballinger, B.M., Olgin, J.E., Marcus, G.M., Pletcher, M.J., 2017. Abstract 21042: Cardiovascular Risk Stratification Using Off-the-Shelf Wearables and a Multi-Task Deep Learning Algorithm. *Circulation* 136.
https://doi.org/10.1161/circ.136.suppl_1.21042
- Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A., 2008. Extracting and composing robust features with denoising autoencoders, in: *Proceedings of the 25th International Conference on Machine Learning - ICML '08*. ACM Press, New York, New York, USA, pp. 1096–1103. <https://doi.org/10.1145/1390156.1390294>
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., Bottou, L., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research* 11.
- Wang, L., 2016. Recognition of Human Activities Using Continuous Autoencoders with Wearable Sensors. *Sensors* 16, 189. <https://doi.org/10.3390/s16020189>
- Witt, D.R., Kellogg, R.A., Snyder, M.P., Dunn, J., 2019. Windows into human health through wearables data analytics. *Curr Opin Biomed Eng* 9, 28–46.
<https://doi.org/10.1016/j.cobme.2019.01.001>
- Wu, J.W., Yaqub, A., Ma, Y., Koudstaal, W., Hofman, A., Ikram, M.A., Ghanbari, M., Goudsmit, J., 2021. Biological age in healthy elderly predicts aging-related diseases including dementia. *Sci Rep* 11, 15929. <https://doi.org/10.1038/s41598-021-95425-5>
- Yanai, H., 2020. Metabolic Syndrome and COVID-19. *Cardiol Res* 11, 360–365.
<https://doi.org/10.14740/cr1181>

ANEXO I: CARACTERÍSTICAS TÉCNICAS DE LAS PULSERAS EMPLEADAS

La elección de los modelos de pulsera empleados en el proyecto vino determinada por los resultados de un análisis comparativo llevado a cabo con anterioridad al inicio del trabajo de investigación propiamente dicho. Tras comparar (para un mismo sujeto) las lecturas de diversos modelos de pulsera, y contrastar también estos resultados con los análisis desarrollados a tal efecto en la literatura primaria, nos decantamos finalmente por los tres modelos siguientes:

1. Fitbit modelo Charge 5
2. Fitbit modelo Inspire 2
3. Xiaomi modelo Mi Smart Band 6 (MiBand 6)

Se decidió emplear varios modelos de pulsera en lugar de uno solo para contemplar el efecto de las divergencias en calidad y prestaciones a la hora de aplicar las técnicas empleadas en el proyecto, buscando explorar si la metodología propuesta es suficientemente robusta y sus resultados no varían en exceso en función de la fuente original de los registros (con tal de que se cumplan unos requisitos mínimos de precisión). La siguiente tabla recoge las principales diferencias entre los tres modelos:

Característica	Fitbit Charge 5	Fitbit Inspire 2	Xiaomi Mi Smart Band 6
Ritmo cardiaco	Registro continuo en intervalos mínimos de <u>5 segundos</u>	Registro continuo en intervalos mínimos de <u>5 segundos</u>	Registro continuo en intervalos mínimos de <u>1 minuto</u>
Variabilidad de ritmo cardiaco (HRV) ¹	Solo durante el periodo de sueño	Solo durante el periodo de sueño	No

¹ Si bien se debe tener en cuenta que la precisión de las estimaciones de variabilidad del ritmo cardiaco basada en mediciones PPG de la muñeca cuenta con importantes limitaciones.

Oxígeno en sangre	Sí	No	Sí
Recuento de pasos diarios	Sí	Sí	Sí
Registro de sueño	Sí	Sí	Sí
Temperatura cutánea	Sí	No	No
Autonomía de la batería	Hasta 7 días	Hasta 10 días	Hasta 14 días
Precio	150 € (moderado)	70 € (relativamente económico)	35 € (económico)

Tabla 8: Comparativa de prestaciones técnicas de las pulseras

Los tres dispositivos emplean la tecnología de fotopletismografía (PPG) para realizar las mediciones de frecuencia cardíaca. Se trata de una técnica de medición indirecta que busca detectar las variaciones en el volumen de sangre en los capilares bajo la piel, en lugar de registrar mediante electrodos las diferencias de potencial eléctrico asociadas a las pulsaciones del corazón. Para tal fin se proyecta un haz de luz sobre la piel y un sensor mide la intensidad de la luz reflejada. Dado que la sangre tiende a absorber muy bien la luz visible en la zona verde del espectro, es posible inferir el ritmo cardíaco a partir de los cambios periódicos en la cantidad de luz reflejada. La señal en bruto requiere de tratamiento de artefactos, especialmente cuando el sensor se encuentra ubicado en la muñeca, por lo que los datos verdaderamente relevantes son los registros de valor promedio de la frecuencia cardíaca para un intervalo de tiempo que será más o menos reducido en función de las prestaciones de la pulsera (entre unos pocos segundos y varios minutos, según el modelo). Como se puede ver en la tabla, para los modelos Fitbit este intervalo es de cinco segundos, mientras que para la pulsera MiBand 6 es de un minuto o cinco minutos, según como se configure.

ANEXO II: CÓDIGO FUENTE

En el siguiente enlace se puede consultar un ejemplo completo (implementado en Python) del procedimiento de lectura de datos, preprocesamiento, modelización y análisis para el sujeto ID1. El código es casi idéntico en los demás voluntarios: <https://github.com/AlfonsoSerranoS/Heart-rate-characterization>

No se incluyen los datos debido a que, aun estando anonimizados, se trata de registros biométricos sensibles cuyo uso para otros fines ajenos a los de la investigación desarrollada en este informe ha de contar con el consentimiento expreso de los voluntarios.