Inferencia afectiva en tiempo real mediante wearables: equilibrio entre precisión y latencia

Gadea Lucas-Pérez[1 [0009−0009−2258−8507]], David Martínez-Acha[1 [0009−0005−4214−2845]], Rodrigo Pascual-García[1 [0009−0005−8791−4159]], Ana Serrano-Mamolar[1 [0009−0009−2258−8507]], and Álvar Arnaiz-González[1 [0000−0001−6965−0237]]

Departamento de Ingeniería Informática, Universidad de Burgos, España
glucas@ubu.es

Resumen Este trabajo estudia la viabilidad de clasificar estados afectivos en tiempo real utilizando señales fisiológicas adquiridas mediante dispositivos *wearables* de bajo coste y *open-source*. En particular, se emplea EmotiBit para registrar señales de fotopletismografía y actividad electrodérmica.

La experimentación incluye cuatro estados: relajación, alta activación, asco y tristeza. Estos eventos se modelan utilizando el espacio bidimensional de valencia-activación. A partir de las señales adquiridas, se desarrolla un *pipeline* completo de aprendizaje automático que incluye preprocesado de señal, segmentación temporal mediante ventanas deslizantes, normalización por usuario, extracción automática de características, selección de variables y entrenamiento de clasificadores. Este pipeline no solo permite transformar las señales fisiológicas en un espacio de representación adecuado para el aprendizaje automático, sino que también facilita el estudio de cómo distintas decisiones de diseño afectan a la capacidad de generalización del modelo y a su futura aplicación en escenarios de inferencia afectiva en tiempo real.

Los resultados muestran rendimientos cercanos a 0.9 de *F1-weighted* con ventanas de 8–10 segundos, lo que constituye un compromiso razonable entre precisión y latencia para aplicaciones orientadas a tiempo real. La selección de características basada en importancia Gini reduce el espacio original a 174 variables, de las cuales el 81 % proceden de actividad electrodérmica, sugiriendo un papel dominante de esta señal en la discriminación afectiva.

Keywords: Computación afectiva · Señales fisiológicas · *Wearables* · Ventanas temporales · Aprendizaje automático

1. Introducción

La clasificación de estados afectivos a partir de señales fisiológicas se ha convertido en una tarea cada vez más viable gracias al desarrollo de dispositivos

portátiles y sensores de bajo coste. Esto abre la puerta al desarrollo de sistemas capaces de responder a cambios del estado afectivo en tiempo real.

En este contexto, el uso de *wearables* en combinación con técnicas de aprendizaje automático ha demostrado resultados prometedores. Sin embargo, existen numerosos desafíos abiertos. En primer lugar, muchos sistemas se desarrollan utilizando dispositivos propietarios o entornos altamente controlados, lo que reduce su posible uso en aplicaciones reales [5,6,12].

Otro problema habitual en este tipo de aplicaciones es lo que se denomina *wide datasets* [9] (*i.e.*, alta dimensionalidad y pocas instancias), donde el número de características puede superar ampliamente el número de usuarios disponibles. Herramientas como `tsfresh` permiten generar cientos de descriptores temporales y estadísticos a partir de series temporales de forma automática [3], pero esto incrementa el riesgo de sobreajuste y hace necesaria una adecuada selección de características.

Asimismo, la evaluación de modelos fisiológicos exige prestar especial atención para evitar *data leakage*. En particular, cuando ventanas procedentes de un mismo usuario aparecen simultáneamente en entrenamiento y *test*. Por este motivo, resulta fundamental emplear estrategias de validación que garanticen la separación completa entre usuarios durante la evaluación.

Otro desafío relevante es el desequilibrio entre clases. En protocolos experimentales de inducción afectiva, la duración de cada fase no siempre es similar y, además, la respuesta fisiológica de los participantes puede generar un número desigual de ventanas válidas por condición. Como consecuencia, algunas clases pueden quedar subrepresentadas durante el entrenamiento, favoreciendo que el modelo aprenda patrones asociados a los estados mayoritarios y reduciendo su capacidad para reconocer condiciones menos frecuentes.

Por otro lado, en aplicaciones orientadas a tiempo real, el tamaño de ventana utilizado para segmentar las señales fisiológicas constituye un parámetro crítico. Ventanas demasiado cortas pueden no contener suficiente información fisiológica, mientras que ventanas excesivamente largas incrementan la latencia del sistema y reducen su capacidad de respuesta en tiempo real. A pesar de su importancia, este aspecto no siempre se ha analizado de forma sistemática en la literatura.

Aunque la literatura reciente muestra resultados prometedores en computación afectiva basada en señales fisiológicas, persisten desafíos relacionados con el diseño de pipelines de aprendizaje automático robustos y reproducibles para aplicaciones de inferencia afectiva en tiempo real. Aspectos como el tamaño de ventana, el tratamiento del desbalanceo de clases o la selección de características pueden influir de forma significativa en el rendimiento del modelo. Sin embargo, su análisis de manera conjunta es limitado, especialmente en estudios basados en dispositivos wearables *open-source*, no intrusivos y de bajo coste.

En este trabajo se estudia la viabilidad de clasificar cuatro estados afectivos diferentes a partir de señales fisiológicas adquiridas mediante un *wearable* de bajo coste y *open-source*: EmotiBit. Para ello, se recogen señales de fotopletismografía (PPG) y actividad electrodérmica (EDA o GSR). Con estas señales de entrada se desarrolla un *pipeline* completo de aprendizaje automático que inclu-

ye: preprocesado de señal, segmentación temporal, normalización por usuario, extracción automática de características mediante `tsfresh`, selección de características y entrenamiento de modelos. Con el objetivo de mejorar el equilibrio entre clases, se emplean técnicas de remuestreo, mientras que la evaluación se realiza utilizando estrategias de validación basadas en separación por usuario para evitar *data leakage*. Finalmente, se analiza la influencia del tamaño de ventana sobre el rendimiento del sistema, con el objetivo de evaluar la viabilidad futura de estos modelos en aplicaciones de inferencia afectiva en tiempo real.

2. Trabajos relacionados

El reconocimiento automático de estados afectivos a partir de señales fisiológicas adquiridas mediante *wearables* ha adquirido una creciente atención en los últimos años. Saganowski et al. [12] presentaron una revisión sistemática centrada en el reconocimiento de estados efectivos mediante señales fisiológicas procedentes de *wearables*, demostrando así el potencial de este tipo de dispositivos, así como las dificultades asociadas a su uso en escenarios reales: variabilidad entre individuos, artefactos de movimiento y la dependencia del contexto.

De forma similar, Li y Zhang [5] hicieron una revisión de los trabajos recientes de computación afectiva en los que se recogen señales fisiológicas con dispositivos *wearables*. Los autores identificaron que, aunque los modelos de aprendizaje automático han alcanzado resultados prometedores, siguen existiendo retos importantes relacionados con la generalización entre usuarios, la calidad de la señal, la interpretabilidad de los modelos y la disponibilidad de datos recogidos en condiciones reales.

Otros estudios han abordado la clasificación de estados afectivos con sensores comerciales cerrados. Mattern et al. [6], por ejemplo, emplearon Empatica E4 para comparar escenarios de laboratorio con contextos más próximos a situaciones reales. Sus resultados muestran la viabilidad de los sensores portátiles, aunque también evidencian la dificultad de generalizar modelos entre participantes y contextos. Por otra parte, Siirtola et al. [14] compararon distintos enfoques de aprendizaje automático para la predicción de estados afectivos a partir de señales fisiológicas, reforzando la importancia de evaluar modelos en condiciones compatibles con un uso continuado.

En los últimos años han comenzado a aparecer sensores *open-source* como alternativa a los dispositivos comerciales cerrados. Montgomery et al. [7] presentaron EmotiBit, un sensor multimodal *open-source* capaz de registrar señales fisiológicas de calidad (PPG, EDA, temperatura y movimiento), con almacenamiento local así como transmisión inalámbrica, una característica que resulta especialmente relevante para el desarrollo de futuras aplicaciones en tiempo real. Por ejemplo, Morris et al. [8] emplearon EmotiBit para recoger señales de HR, PPG y EDA en un contexto de interacción realista, reforzando su interés como plataforma accesible para estudios de computación afectiva.

3. Configuración experimental

Con el objetivo de evaluar la viabilidad de la clasificación de estados afectivos, se diseñó un protocolo experimental orientado a inducir diversos estados afectivos y niveles de activación fisiológica[1].

Esta sección describe tanto el dispositivo utilizado, los participantes, el protocolo experimental seguido y el proceso de adquisición de señales fisiológicas empleado para la construcción del conjunto de datos.

3.1. Dispositivo y participantes

El *wearable* empleado fue EmotiBit[2], con el objetivo de registrar señales fisiológicas de forma continua y no invasiva. En concreto, se adquirieron señales de PPG y EDA, utilizadas posteriormente para la extracción de características fisiológicas y la inferencia de estados afectivos.

El total de participantes fueron 10 (5 hombres y 5 mujeres), con una edad media de 27.9 $\pm$ 7.1 años. Todos los participantes firmaron un consentimiento informado previo a su inclusión en el estudio, de acuerdo con los procedimientos aprobados por el Comité de Ética de la Universidad de Burgos (código de protocolo IR-14/2022, aprobado en diciembre de 2022). Asimismo, todos los procedimientos experimentales se llevaron a cabo conforme a los principios establecidos en la Declaración de Helsinki [1].

3.2. Protocolo experimental

En cuanto a la ubicación, el *wearable* se situó en la muñeca, como recomienda el fabricante. En este caso, se orientó a la zona volar (cara interna) de la muñeca[3].

La sesión experimental se estructuró en las siguientes fases (ver Figura 1):

- ***Baseline* (en estado de relajación):** dos minutos en reposo sentado para establecer una referencia fisiológica (condición basal).
- ***Squat test*:** realización de 30 sentadillas en 45 segundos para inducir activación cardiovascular, similar al *test* de Ruffier-Dickson [13].
- **Relajación 1:** periodo de recuperación de un minuto tras la actividad física.
- **Vídeo 1:** fragmento audiovisual de 29 segundos utilizado para inducir asco (extracto de «Pink Flamingos») [10].
- **Relajación 2:** periodo de recuperación de un minuto tras el estímulo.
- **Vídeo 2:** fragmento audiovisual de 130 segundos utilizado para inducir tristeza (extracto de «El Rey León») [10].

[1] El código fuente del análisis de los datos para la predicción de estados afectivos está disponible en https://github.com/gadea-lucas/ventanas-emociones-affective-prediction

[2] https://www.emotibit.com/

[3] La elección de la zona volar para la medición de EDA se justifica por la mayor densidad de glándulas sudoríparas en esta región, lo que permite una mayor sensibilidad en la captación de la actividad electrodérmica [4].



Figura 1: Fases de la sesión experimental

3.3. Procesado de señal

Las señales recogidas fueron procesadas de forma *offline*. El procesamiento incluyó varias etapas: 1) remuestreo para homogeneizar las frecuencias de muestreo, 2) detección y eliminación de valores atípicos, y 3) aplicación de filtros para reducir el ruido y mejorar la calidad de las señales.

En particular, se empleó el método del Rango Intercuartílico para la detección de valores atípicos (*Interquartile Range*, IQR), interpolando posteriormente los huecos generados tras su eliminación. En cuanto al filtrado, se aplicó uno gaussiano a la señal de EDA y un filtro Butterworth de quinto orden (1-15 Hz) a la señal de PPG.

Además, a partir de las señales de PPG, se derivó la frecuencia cardíaca (HR) empleando el algoritmo de EmotiBit[4].

3.4. Definición de clases

Tras la adquisición y procesamiento de las señales, se inicia la fase de modelado orientada a la predicción de estados afectivos. Los distintos eventos (condición basal, activación física y visualización de vídeos) se etiquetan y posteriormente se proyectan en el modelo bidimensional de estados afectivos de Modelo Circunflejo de las Emociones propuesto por Russell [11]. Este modelo organiza los estados afectivos en un espacio definido por dos dimensiones principales: valencia (positiva-negativa) y activación (alta-baja).

A partir de esta conceptualización, los distintos eventos se proyectan sobre dicho espacio (como se muestra en la Figura 2) con el fin de definir las clases utilizadas en el posterior proceso de aprendizaje automático. En particular, se consideran cuatro condiciones principales: estado basal (relajación), activación física (sentadillas), y dos estados afectivos inducidos mediante estímulos audiovisuales (asco y tristeza).

El estado basal (relajación) se caracteriza por una baja activación fisiológica y una valencia neutra o ligeramente positiva, asociada a condiciones de reposo y

[4] Código disponible en https://github.com/EmotiBit/EmotiBit_FeatherWing/blob/master/EmotiBit.cpp#L3289.

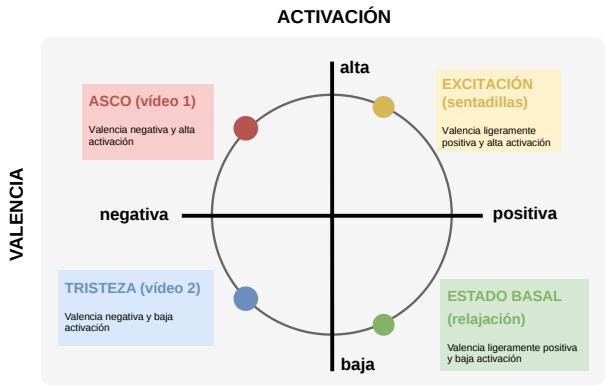


Figura 2: Representación de los eventos en el modelo de valencia-activación.

ausencia de estímulos relevantes. Por este motivo, se sitúa en la región de baja activación y valencia neutra o positiva del modelo.

Por otro lado, la condición de sentadillas, aunque no constituye una emoción en sentido estricto, provoca un incremento significativo de la activación fisiológica, reflejado en variables como el HR o el EDA. Dado que no implica necesariamente una evaluación afectiva negativa, se posiciona en la zona de alta activación y valencia neutra o ligeramente positiva. La inclusión de esta condición resulta especialmente relevante, ya que permite disponer de ejemplos de alta activación fisiológica no asociada a estados afectivos negativos, facilitando la diferenciación entre activación física y activación afectiva en el modelo.

En cuanto a los estímulos emocionales, la condición de asco se asocia típicamente a una valencia negativa y a un nivel elevado de activación, debido a su carácter aversivo y a la respuesta fisiológica intensa que genera. Por ello, se ubica en la región de alta activación y valencia negativa.

Por último, la tristeza se caracteriza por una valencia negativa pero con un nivel de activación relativamente bajo, asociado a estados de apatía o baja energía. En consecuencia, se sitúa en la zona de baja activación y valencia negativa.

3.5. *Pipeline* de aprendizaje automático

Tras establecer las clases, se desarrolla un *pipeline* completo de aprendizaje automático para predecir el estado afectivo a partir de señales fisiológicas.

Un aspecto clave en este proceso es la segmentación temporal de las señales, ya que, de cara a una futura implementación en tiempo real, resulta fundamental determinar el tamaño de ventana óptimo que permita un equilibrio entre rapidez de respuesta y precisión del modelo. Para ello, se generan múltiples versiones del *dataset* utilizando ventanas de duración variable, comprendidas entre 1 y 10 segundos. Cada ventana constituye una de estas versiones, permitiendo analizar cómo afecta la resolución temporal al rendimiento del sistema.

Se evaluaron dos estrategias: ventanas sin solapamiento y ventanas con solapamiento. La configuración con solapamiento se considera más próxima a un escenario en tiempo real, ya que permite actualizar la predicción con mayor frecuencia, aunque la latencia mínima sigue determinada por la duración de la ventana utilizada.

Previo a la extracción de características, se realiza una normalización por usuario, utilizando el periodo basal como referencia. Para cada individuo se calculan la media y desviación estándar de las señales HR y EDA en estado de reposo, y posteriormente se normalizan las señales en el resto de fases. Este paso es esencial para reducir la variabilidad inter-individuo y mejorar la capacidad de generalización del modelo.

A continuación, se lleva a cabo la extracción automática de características sobre todas las ventanas temporales de cada usuario. Para ello se emplea la librería `tsfresh`, que calcula un amplio conjunto de descriptores estadísticos y temporales a partir de las señales. Las características se obtienen de manera independiente para HR y EDA, combinándolo posteriormente en un único vector. El resultado es un *dataset* tabular estructurado donde cada usuario está representado por múltiples filas, una por cada ventana temporal. Cada fila incluye las características asociadas, la etiqueta de clase y el identificador del usuario.

Antes del modelado, se aplica una fase de limpieza de datos, eliminando variables con valores nulos o varianza cero, garantizando así que todas las características aporten información relevante al modelo.

El proceso de entrenamiento y evaluación se realiza de forma sistemática para cada tamaño de ventana mediante validación cruzada. En concreto, se utiliza un esquema de `StratifiedGroupKFold` con 10 particiones, que permite mantener la proporción de clases y, al mismo tiempo, asegurar que los datos de un mismo usuario no aparecen simultáneamente en entrenamiento y *test*. La idea es simular una validación *leave-one-out*, para que ninguna ventana del usuario de *test* se incluya en el entrenamiento.

Para abordar el desequilibrio de clases se compararon diversas estrategias de remuestreo en entrenamiento, generando muestras sintéticas de las clases minoritarias sin alterar el conjunto de *test*. En particular, *random oversampling* (ROS), SMOTE [2], SMOTEENN y SMOTETomek, todas disponibles en la biblioteca `imbalanced-learn`[5].

Por otro lado, para valorar la capacidad predictiva de los modelos se consideraron tres clasificadores: *Random Forest*, *kNN* y *Logistic Regression*. Dado el desbalanceo presente en el conjunto de datos, se utilizó la métrica *F1 weighted* para su comparación.

4. Resultados y discusión

El análisis de resultados se organiza en torno a tres decisiones principales del *pipeline*. En primer lugar, se evalúan diversos clasificadores y algoritmos de remuestreo. En segundo lugar, se analiza la influencia del tamaño de ventana, dado

[5] https://imbalanced-learn.org

que cada duración temporal genera una versión distinta del conjunto de datos y condiciona directamente el equilibrio entre rendimiento y latencia. Finalmente, sobre la configuración seleccionada se aborda la reducción de dimensionalidad mediante la importancia de Gini y un barrido sistemático de umbrales, con el objetivo de identificar el mejor subconjunto de características.

4.1. Efecto del tamaño de ventana y del solapamiento temporal

El aspecto más relevante para construir un modelo de predicción en tiempo real es el tamaño de ventana. Esto influye tanto en la calidad de predicciones como la capacidad de obtener predicciones en intervalos razonables. Para ello, se evaluaron ventanas comprendidas entre 1 y 10 segundos, generando un modelo independiente para cada configuración.

La Figura 3 (a y b) muestra los resultados con diversos tamaños de ventana (sin y con solapamiento). Se observa que el rendimiento mejora al pasar de ventanas muy cortas a ventanas intermedias (con picos alrededor de 5 y 8 segundos). Como se ha mencionado, la configuración con solapamiento temporal resulta especialmente relevante para un escenario cercano al tiempo real, ya que permite actualizar la predicción con mayor frecuencia una vez acumulada la ventana inicial. En ambos escenarios, *Random Forest* presentaa el comportamiento más competitivo frente a kNN y regresión logística.

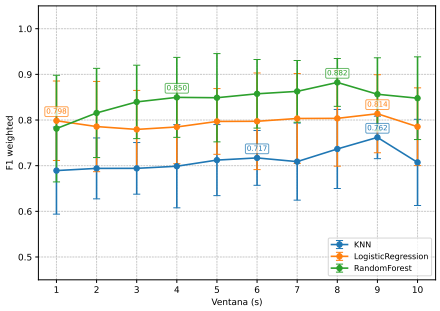
La Figura 3 (c y d) muestra la evolución del *F1-weighted* utilizando *Random Forest* con diversas técnicas de remuestreo para ventanas de 1 a 10 segundos (se comparan las configuraciones con y sin solapamiento).

Los resultados muestran, nuevamente, una tendencia creciente desde ventanas muy cortas hasta ventanas intermedias. En torno a 5 segundos se observan valores razonables de rendimiento, cercanos al 0.85. Por otro lado, en torno a 8 segundos se alcanza ya un rendimiento de 0.90 de *F1-weighted*, lo que representa un buen compromiso entre precisión y latencia para escenarios de inferencia afectiva en tiempo real: ventanas más largas pueden mejorar el rendimiento, pero también incrementan el tiempo necesario para emitir la predicción.

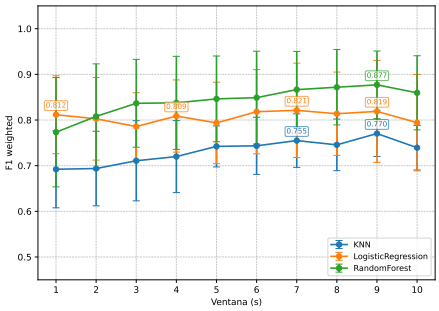
En la configuración con solapamiento, ROS obtiene el mejor resultado en ventanas de 8 segundos (0.8996 ± 0.0874), seguido de SMOTEENN (0.8966 ± 0.1084), SMOTE (0.8955 ± 0.1126) y SMOTETomek (0.8921 ± 0.1098). El clasificador sin remuestreo alcanza 0.8718 ± 0.0826. En ventanas de 10 segundos, SMOTE obtiene el mejor resultado (0.9022 ± 0.1037), seguido de SMOTETomek (0.9009 ± 0.0983) y SMOTEENN (0.9001 ± 0.1087).

La Tabla 1 resume los resultados en ventanas de 8 y 10 segundos para las configuraciones con y sin solapamiento.

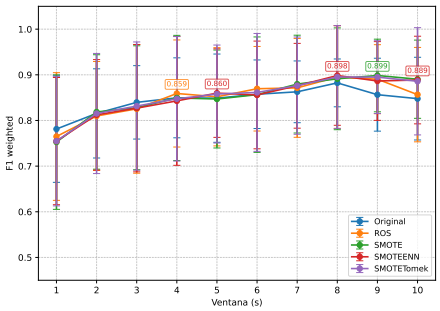
En resumen, se observa que las estrategias de remuestreo mejoran el rendimiento frente a no balancear, aunque las diferencias entre ROS y las variantes de SMOTE son pequeñas y dependen del tamaño de ventana. Esto sugiere que el problema no se resuelve únicamente mediante una técnica concreta de sobremuestreo, sino mediante la combinación de una ventana temporal adecuada y una estrategia que reduzca el sesgo del modelo hacia las clases mayoritarias.



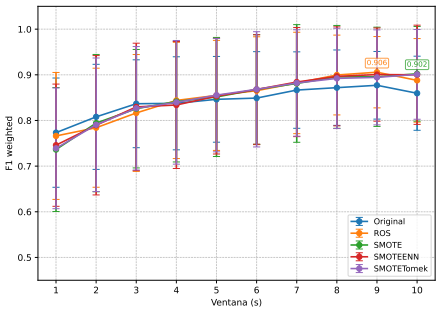
(a) Sin solapamiento.



(b) Con solapamiento.



(c) Sin solapamiento.



(d) Con solapamiento.

Figura 3: Evolución del *F1-weighted* en función del tamaño de ventana con y sin solapamiento. En (a) y (b) los clasificadores se entrenaron sin remuestreo; en (c) y (d) los resultados de Random Forest con remuestreo.

Tabla 1: Resultados de *Random Forest* para ventanas de 8 y 10 segundos. Se reporta *F1-weighted* medio ± desviación estándar.

Solapamiento	**Balanceo**	**Ventanas de 8 s**	**Ventanas de 10 s**
Con solapamiento	Sin balanceo	0.8718 ± 0.0826	0.8596 ± 0.0812
	ROS	$\mathbf{0.8996} \pm 0.0874$	0.8878 ± 0.0915
	SMOTE	0.8955 ± 0.1126	$\mathbf{0.9022} \pm 0.1037$
	SMOTEENN	0.8966 ± 0.1084	0.9001 ± 0.1087
	SMOTETomek	0.8921 ± 0.1098	0.9009 ± 0.0983
Sin solapamiento	Sin balanceo	0.8823 ± 0.0524	0.8478 ± 0.0906
	ROS	0.8952 ± 0.1127	0.8566 ± 0.1034
	SMOTE	0.8913 ± 0.1115	$\mathbf{0.8904} \pm 0.0857$
	SMOTEENN	$\mathbf{0.8985} \pm 0.1090$	0.8886 ± 0.0960
	SMOTETomek	0.8953 ± 0.1126	0.8858 ± 0.1174

Desde el punto de vista fisiológico, las ventanas muy cortas pueden no capturar adecuadamente la dinámica de respuestas autonómicas como la actividad electrodérmica, cuya evolución no es instantánea. Por el contrario, ventanas más largas agregan más información, pero aumentan la latencia del sistema. Por ello, ventanas de 8 segundos parecen constituir una solución de compromiso: permiten alcanzar valores cercanos a 0.9 de rendimiento sin requerir ventanas excesivamente largas.

4.2. Selección de características

Posteriormente, se realiza un análisis de importancia de variables basado en la medida Gini del modelo *Random Forest* con el fin de reducir la dimensionalidad y mejorar la eficiencia. A partir del *dataset* original de 745 características (563 de EDA y 182 de HR), se evalúa el rendimiento del modelo mediante un barrido de umbrales sobre la importancia Gini.

En la Figura 4 se observa que la mayor parte de la importancia está agrupada en la parte inicial. Esto indica que es posible disminuir el número de características sin afectar de forma significativa el rendimiento.

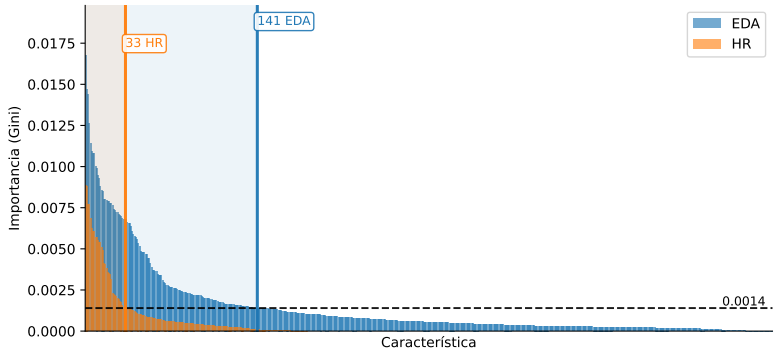


Figura 4: Importancia de cada característica según su importancia de Gini, ordenadas de mayor a menor. Combinación experimental: ventana de 8 segundos con solapamiento y ROS para balanceo.

En el caso estudiado, el mejor resultado se obtiene con un umbral de 0.0014, seleccionando 174 variables y alcanzando un *F1-weighted* de 0.9127 ± 0.0796, lo que confirma la eficacia de la selección de características para simplificar el modelo manteniendo (y mejorando) su capacidad predictiva.

5. Conclusiones y líneas futuras

Los resultados obtenidos confirman la viabilidad de la predicción de estados afectivos a partir de señales fisiológicas PPG y EDA recogidas mediante un

wearable (EmotiBit) en la zona volar de la muñeca. En particular, el algoritmo *Random Forest* demuestra una elevada capacidad para capturar relaciones no lineales entre las variables fisiológicas y los estados afectivos.

Uno de los resultados más relevantes del trabajo es que el tamaño de ventana constituye un parámetro crítico en el diseño de sistemas de inferencia afectiva en tiempo real. Los resultados muestran que ventanas de alrededor de 8 segundos proporcionan un compromiso adecuado entre capacidad predictiva y latencia, alcanzando valores cercanos a 0.9 de *F1-weighted* sin requerir intervalos temporales excesivamente largos para emitir una predicción.

Por otra parte, las estrategias de remuestreo mejoran notablemente el rendimiento frente al entrenamiento sin balanceo. En ventanas de 8 segundos con solapamiento temporal, ROS alcanza el mejor resultado, mientras que en ventanas de 10 segundos destaca SMOTE. No obstante, las diferencias entre métodos son pequeñas, por lo que el resultado principal debe interpretarse como evidencia de la utilidad del balanceo, más que como superioridad absoluta de una técnica concreta.

La selección de características basada en importancia Gini reduce el conjunto original de 745 variables a 174, manteniendo un buen rendimiento. Además, el análisis muestra una clara predominancia de características procedentes de EDA, que representan el 81 % del subconjunto seleccionado. Una vez validada la configuración, este subconjunto de características permite entrenar un modelo definitivo sobre el conjunto completo de ventanas disponibles, orientado a su futura serialización e integración en aplicaciones de inferencia afectiva en tiempo real.

Aunque el número de participantes es limitado, la evaluación se ha realizado mediante una estrategia de validación con separación estricta por usuario, evitando situaciones de data leakage y proporcionando una estimación más realista de la capacidad de generalización del modelo que una validación basada en ventanas aleatorias. No obstante, la ampliación del conjunto de participantes constituye una línea futura necesaria para confirmar la robustez de los resultados obtenidos y aumentar la capacidad de generalización del modelo. También se contemplan como trabajos futuros estudiar estrategias de adaptación por usuario y desplegar el modelo en un prototipo que permita validar la inferencia afectiva en tiempo real.

Agradecimientos

Este trabajo es parte del proyecto de I+D+i PID2023-150694OA-I00, financiado por MICIU/AEI/10.13039/501100011033 y "FEDER/UE".

Referencias

1. Association, W.M.: World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. JAMA **333**(1), 71–74 (01 2025). https://doi.org/10.1001/jama.2024.21972

2. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002). https://doi.org/10.1613/jair.953
3. Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W.: Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). Neurocomputing **307**, 72–77 (2018). https://doi.org/10.1016/j.neucom.2018.03.067
4. Hossain, M.B., Kong, Y., Posada-Quintero, H.F., Chon, K.H.: Comparison of electrodermal activity from multiple body locations based on standard eda indices' quality and robustness against motion artifact. Sensors **22**(9) (2022). https://doi.org/10.3390/s22093177
5. Li, F., Zhang, D.: Multimodal physiological signals from wearable sensors for affective computing: A systematic review. Intelligent Sports and Health **1**(4), 210–222 (2025). https://doi.org/https://doi.org/10.1016/j.ish.2025.10.001
6. Mattern, E., Jackson, R.R., Doshmanziari, R., Dewitte, M., Varagnolo, D., Knorn, S.: Emotion recognition from physiological signals collected with a wrist device and emotional recall. Bioengineering (Basel, Switzerland) **10**(11), 1308 (Nov 2023). https://doi.org/10.3390/bioengineering10111308
7. Montgomery, S.M., Nair, N., Chen, P., Dikker, S.: Introducing emotibit, an open-source multi-modal sensor for measuring research-grade physiological signals. Science Talks **6**, 100181 (2023). https://doi.org/https://doi.org/10.1016/j.sctalk.2023.100181
8. Morris, C., Danry, V., Maes, P.: Wearable systems without experiential disruptions: exploring the impact of device feedback changes on explicit awareness, physiological synchrony, sense of agency, and device-body ownership. Frontiers in Computer Science **Volume 5 - 2023** (2023). https://doi.org/10.3389/fcomp.2023.1289869
9. Ramos-Pérez, I., Barbero-Aparicio, J.A., Canepa-Oneto, A., Arnaiz-González, A., Maudes-Raedo, J.: An extensive performance comparison between feature reduction and feature selection preprocessing algorithms on imbalanced wide data. Information **15**(4) (2024). https://doi.org/10.3390/info15040223
10. Rottenberg, J., Ray, R.D., Gross, J.J.: Emotion elicitation using films: Appendix – film stimulus instructions. In: Coan, J.A., Allen, J.J.B. (eds.) The handbook of emotion elicitation and assessment. Oxford University Press (2007). https://doi.org/10.1080/02699939508408966
11. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology **39**(6), 1161–1178 (1980). https://doi.org/10.1037/h0077714
12. Saganowski, S., Perz, B., Polak, A.G., Kazienko, P.: Emotion Recognition for Everyday Life Using Physiological Signals From Wearables: A Systematic Literature Review . IEEE Transactions on Affective Computing **14**(03), 1876–1897 (jul 2023). https://doi.org/10.1109/TAFFC.2022.3176135
13. Sartor, F., Bonato, M., Papini, G., Bosio, A., Mohammed, R.A., Bonomi, A.G., Moore, J.P., Merati, G., La Torre, A., Kubis, H.P.: A 45-second self-test for cardiorespiratory fitness: Heart rate-based estimation in healthy individuals. PLoS One **11**(12) (2016). https://doi.org/10.1371/journal.pone.0168154
14. Siirtola, P., Tamminen, S., Chandra, G., Ihalapathirana, A., Röning, J.: Predicting emotion with biosignals: A comparison of classification and regression models for estimating valence and arousal level using wearable sensors. Sensors (Basel, Switzerland) **23**(3), 1598 (Feb 2023). https://doi.org/10.3390/s23031598