

Jornadas de Automática

¿Puede una monitorización fisiológica detectar carga mental en operadores UAV?

Alberto Jiménez Hormeño^a, Raul Fernández Matellán^{a,*}, Carlos Castellanos Ormeño^a,
David Martín Gómez^a, Arturo de la Escalera Hueso^a

^aIntelligent Systems Lab, Universidad Carlos III de Madrid, Leganés, Madrid, España.

To cite this article: Jiménez Hormeño, A., Fernández Matellán, R., Castellanos Ormeño, C., Martín Gómez, D., de la Escalera Hueso, A. 2026. Can physiological monitoring detect mental workload in UAV operators? Jornadas de Automática, 47. <https://doi.org/10.17979/ja-cea.2026.47.13714>

Resumen

La operación de vehículos aéreos no tripulados (UAV) combina supervisión visual, control manual y toma de decisiones; por lo tanto, la carga mental del operador es crítica para la seguridad. Este trabajo presenta una validación preliminar de un protocolo intrasujeto de medidas repetidas para estudiar dicha carga durante vuelos reales en interior. Se registraron 10 sesiones con un operador certificado UAS A1/A3, incluyendo bloques de 2 min bajo tres condiciones: *Estacionario*, *Fácil* y *Difícil*. La adquisición combinó telemetría del UAV, señales fisiológicas de muñeca y vídeo facial, sincronizados fuera de línea. Los resultados muestran una progresión coherente de la carga percibida y de la exigencia cinemática. Además, una línea base personalizada basada en EDA y PPG discriminó entre *Estacionario* y *Difícil* en sesiones no vistas, sugiriendo el potencial de este enfoque en el ámbito intrasujeto y respaldando la continuidad de la investigación.

Palabras clave: Modelado del rendimiento humano, UAVs, Teleoperación, Fusión de datos de sensores, Validación del modelo

Can physiological monitoring detect mental workload in UAV operators?

Abstract

Unmanned Aerial Vehicle (UAV) operation combines visual supervision, manual control, and decision-making, making the operator's mental workload relevant to safety. This paper presents a preliminary validation of an intra-subject repeated-measures protocol to study workload during real indoor flights. Ten sessions were recorded with a certified UAS A1/A3 operator, including 2-min blocks under three conditions: *Hover*, *Easy*, and *Difficult*. The acquisition combined UAV telemetry, wrist-worn physiological signals, and facial video, synchronized offline. The results show a coherent progression of perceived workload and kinematic demand across conditions. In addition, a personalized baseline based on EDA and PPG discriminated between *Hover* and *Difficult* in unseen sessions, suggesting the potential of this intra-subject approach and supporting further research.

Keywords: Modeling of human performance, UAVs, Teleoperation, Sensor data fusion, Model validation

1. Introducción

La operación remota de UAV (*Unmanned Aerial Vehicle*) es una tarea de supervisión humano-máquina crítica para la seguridad. Durante una misión se pueden alternar periodos de baja demanda vinculados a la vigilancia del entorno, con maniobras que exigen control preciso e intervención rápida. Estos cambios afectan al rendimiento del piloto al requerir su-

pervisión, control y toma de decisiones veloz. En sistemas de control supervisado, la transición de un estado pasivo a una intervención crítica se relaciona con la “ironía de la automatización” (Bainbridge, 1983), y con los riesgos asociados al uso y abuso de estas plataformas (Parasuraman and Riley, 1997). Por tanto, estimar el estado del operador es clave para caracterizar la tarea, con el fin de diseñar sistemas de apoyo adaptativo, estrategias de entrenamiento y mecanismos de prevención.

*Autor para correspondencia: raulfern@ing.uc3m.es

Formato de citación: APA (American Psychological Association)

Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Las señales fisiológicas registradas mediante sensores portátiles facilitan la monitorización continua y no invasiva del operador durante la tarea. La actividad electrodérmica (EDA), la fotopletismografía (PPG) y la temperatura cutánea pueden proporcionar indicadores indirectos de activación autonómica y respuesta fisiológica ante cambios en la demanda (Luzzani et al., 2024). A diferencia de cuestionarios como NASA-TLX (Hart and Staveland, 1988) o RSME (Zijlstra, 1993), que recogen la percepción del operador una vez finalizada la tarea, estas señales permiten analizar variaciones temporales y complementar la evaluación subjetiva.

La carga de trabajo en operadores de UAV se ha estudiado mediante escalas subjetivas, indicadores oculares, señales fisiológicas y técnicas de aprendizaje automático (Russo et al., 2024; Singh et al., 2021; Dell'Agnola et al., 2022; Kostenko et al., 2022). Sin embargo, muchos trabajos se apoyan en simuladores, registros de corta duración o modelos generalistas. Estas aproximaciones son útiles para identificar tendencias generales, pero son menos efectivas para un operador concreto durante varias sesiones reales. En el análisis fisiológico, esta limitación es especialmente importante, ya que las señales pueden modificarse por diferencias individuales, fluctuaciones del estado basal, movimiento, contacto sensor-piel y condiciones propias de cada sesión (Bolpagni et al., 2024; Giorgi et al., 2021). Además, los patrones grupales no siempre describen adecuadamente la dinámica de un individuo (Fisher et al., 2018). En consecuencia, antes de plantear modelos generalizables, se debe comprobar si un protocolo personalizado permite obtener datos estables y distintivos sobre la carga.

Este trabajo plantea la estimación del estado de un operador de UAV mediante un protocolo intrasujeto de medidas repetidas durante vuelos reales. El planteamiento de partida es que maniobras con distinto nivel de demanda pueden producir cambios observables en la carga percibida y en las señales fisiológicas del operador. Asimismo, se analiza si esas variaciones permiten diferenciar situaciones de baja y alta demanda en sesiones que no han sido utilizadas durante el entrenamiento del modelo. En concreto, se plantean tres hipótesis:

- **H1:** Las maniobras de vuelo generan niveles diferenciados de carga mental percibida.
- **H2:** Las señales fisiológicas varían en función de la condición experimental.
- **H3:** Un modelo personalizado permite discriminar entre estados de baja y alta demanda en sesiones no vistas durante el entrenamiento.

La contribución principal de este estudio es la validación preliminar de un protocolo personalizado, diseñado para evaluar la carga mental de un operador de UAV durante vuelos reales. Con este fin, se integran autoinformes, telemetría y señales fisiológicas de muñeca, proporcionando una caracterización multimodal de la tarea desde la dinámica del vuelo, la percepción subjetiva del operador y su respuesta fisiológica.

Este marco permite analizar la capacidad de discriminación entre condiciones de vuelo y evaluar una línea base personalizada mediante *Leave-One-Session-Out* (LOSO), donde cada sesión se reserva una vez para prueba y las restantes para entrenamiento. El estudio se plantea como una demostración inicial de viabilidad, no como una inferencia poblacional.

2. Trabajos relacionados

Como se mencionó en el anterior apartado, la estimación de la carga mental en operadores de UAV se ha estudiado mediante el uso de diferentes métricas (Russo et al., 2024; Singh et al., 2021; Dell'Agnola et al., 2022; Kostenko et al., 2022). Entre las herramientas más utilizadas se encuentran NASA-TLX y RSME, que permiten valorar la carga percibida y el esfuerzo mental una vez finalizada la tarea (Hart and Staveland, 1988; Zijlstra, 1993). Estas medidas son útiles para comprobar si las condiciones experimentales generan distintos niveles de demanda, aunque no permiten seguir de forma continua cómo evoluciona el estado del operador durante el vuelo.

Otras líneas de trabajo analizan las medidas oculares y faciales para estimar la atención, los patrones de exploración visual y la carga de trabajo en pilotos de UAV (Maciejewska, 2023; Rauffet et al., 2016). Estos indicadores pueden reflejar cambios asociados a la dificultad de la tarea, pero suelen depender de condiciones de captura relativamente controladas, como una iluminación adecuada, pocas oclusiones y una posición estable de la cámara. En este estudio se registra vídeo facial como modalidad complementaria, aunque el análisis se centra en señales fisiológicas de muñeca, por su carácter no invasivo y su potencial para monitorizar al operador.

Las señales de EDA, PPG, variabilidad de la frecuencia cardíaca y temperatura cutánea se han utilizado para estimar la carga cognitiva, el estrés y el estado funcional del operador en tareas de supervisión y teleoperación (Das Chakladar and Roy, 2024; Giorgi et al., 2021; Luzzani et al., 2024). En el ámbito UAV, estudios previos han utilizado características fisiológicas y aprendizaje automático en misiones con drones, pilotaje cooperativo o escenarios con enjambres (Dell'Agnola et al., 2022; Kostenko et al., 2022; Singh et al., 2021). No obstante, su interpretación sigue condicionada por el movimiento, la variabilidad interindividual, los cambios entre sesiones y la forma en que se induce la carga en el experimento.

La generalización en señales fisiológicas sigue siendo un reto. Aunque se han empleado modelos personalizados y generalistas en escenarios reales (Anders et al., 2025), los estudios N-de-1 muestran que las tendencias observadas a nivel de grupo no siempre describen bien la dinámica de una persona concreta (Fisher et al., 2018; Lillie et al., 2011). Esta idea ha motivado enfoques personalizados en detección de estrés (Bolpagni et al., 2024) y protocolos intrasujeto en conducción con sensores comerciales (Gruden et al., 2019). En esta línea, el presente trabajo combina vuelos reales, medidas repetidas, sincronización multimodal y evaluación entre sesiones para estudiar la carga mental de un operador objetivo.

Desde esta perspectiva, la contribución del artículo no se limita a aplicar características fisiológicas a una tarea UAV. Su objetivo es analizar si las señales mantienen capacidad discriminativa al registrarlas durante vuelos reales, en varios días de adquisición y con una separación estricta entre entrenamiento y prueba. Esta formulación permite estudiar un problema complementario al de la generalización poblacional: la viabilidad de una calibración personalizada para un operador concreto. De este modo, el trabajo aborda la diferenciación subjetiva de la carga (**H1**), la capacidad de separación fisiológica y conductual entre condiciones (**H2**) y la robustez de la generalización entre sesiones dentro del mismo individuo (**H3**).

3. Metodología

La metodología se diseñó para determinar, en un mismo operador objetivo, la solidez de los cambios en la carga percibida, la dinámica del vuelo y las señales fisiológicas bajo distintos escenarios, combinando adquisición multimodal en vuelos reales con evaluación personalizada entre sesiones.

3.1. Diseño experimental

Los experimentos se realizaron en un entorno interior con un UAV multirrotor DJI Air 2S. Se definieron tres escenarios de pilotaje manual: *Estacionario*, manteniendo la aeronave estable; *Fácil*, con un circuito sencillo sin obstáculos; y *Difícil*, con una trayectoria restringida con obstáculos estáticos.

Cada sesión incluyó seis bloques de vuelo, dos por escenario, presentados en orden aleatorio, evitando las repeticiones consecutivas. Cada bloque constó de 10 s de despegue, 120 s de tarea efectiva, 10 s de aterrizaje y 2 min de descanso. El protocolo se repitió en 10 sesiones realizadas en días diferentes, obteniendo 20 bloques por escenario. Esta estructura permite evaluar la consistencia de las diferencias entre repeticiones e incorporar variabilidad natural entre sesiones.

3.2. Sistema de adquisición y sensores

La adquisición integró telemetría del UAV, señales fisiológicas de muñeca y vídeo facial, como se muestra en la Figura 1. Se utilizó un DJI Air 2S para el pilotaje manual en interior. El rostro del operador se registró con una cámara Intel RealSense D435i situada sobre un trípode a aproximadamente 1.5 m, mientras el operador permanecía sentado. Las señales fisiológicas se adquirieron con un EmotiBit colocado en la muñeca no dominante, registrando PPG, EDA, temperatura cutánea e IMU a 25, 15, 7.5 y 25 Hz, respectivamente.

Todas las modalidades se sincronizaron mediante marcas temporales UTC y eventos durante el experimento. Aunque la Figura 1 incluye la posición prevista de una pulsera Empatica EmbracePlus, este dispositivo se excluyó del análisis por sus limitaciones en la monitorización en tiempo real, pudiéndose emplear en trabajos posteriores como referencia comparativa.

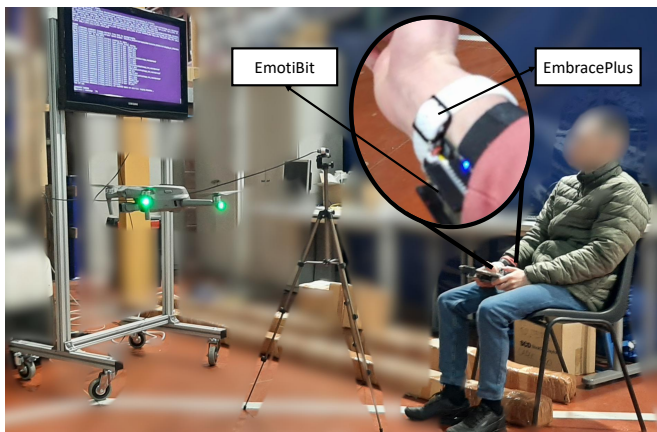


Figura 1: Configuración experimental para el pilotaje manual interior de UAV y la adquisición multimodal del operador.

3.3. Procedimiento experimental

Al inicio de cada sesión, el operador completó la escala Karolinska Sleepiness Scale (KSS) para estimar la somnolencia basal. En cada bloque, se verificó la calidad de la señal, se registraron la condición experimental y los marcadores temporales de inicio y fin. Tras finalizar cada bloque, el operador completó la escala RSME para medir el esfuerzo mental percibido y descansó 2 min. Al término de la sesión, se aplicó el cuestionario NASA-TLX no ponderado para cada condición. El flujo de adquisición se resume en el Algoritmo 1. Esta estructura permite segmentar los datos por sesión, bloque y condición, y facilita la reproducción del protocolo.

Algoritmo 1 Procedimiento de adquisición por sesión

```

1: Registrar KSS al inicio de la sesión
2: Iniciar adquisición multimodal y comprobar calidad de señal
3: for cada bloque de vuelo do
4:   Indicar la condición: Estacionario, Fácil o Difícil
5:   Registrar marcador temporal START
6:   Ejecutar despegue durante 10 s
7:   Ejecutar tarea de vuelo durante 120 s
8:   Ejecutar aterrizaje durante 10 s
9:   Registrar marcador temporal END
10:  Registrar RSME post-bloque
11:  Realizar descanso durante 2 min
12: end for
13: Registrar NASA-TLX no ponderado para las tres condiciones
14: Finalizar sesión

```

3.4. Procesamiento y evaluación

Los registros sincronizados se segmentaron por bloque experimental y se etiquetaron como *Estacionario*, *Fácil* o *Difícil*. La hipótesis **H1** se evaluó comparando las puntuaciones RSME y NASA-TLX, junto con perfiles cinemáticos de velocidad, aceleración y jerk derivados de la telemetría.

La hipótesis **H2** se analizó mediante las señales del EmotiBit y del vídeo facial. Las señales fisiológicas y de movimiento se segmentaron en ventanas de 10 s sin solape. La acelerometría se caracterizó mediante la magnitud vectorial, estadísticos robustos, el *jerk* y ráfagas de movimiento detectadas con un umbral definido como la mediana más tres veces la desviación absoluta mediana (MAD, *Median Absolute Deviation*). En EDA se estimaron las componentes tónica y fásica mediante una mediana móvil de 5 s, identificando las respuestas de conductancia como máximos locales superiores al mismo umbral robusto. La frecuencia cardíaca se resumió mediante estadísticos descriptivos, la señal PPG se analizó mediante la transformada rápida de Fourier en la banda de 0.7-3.0 Hz para estimar la componente cardíaca dominante, y la temperatura cutánea se caracterizó mediante estadísticos de nivel, dispersión y pendiente temporal.

El vídeo RGB de la cámara RealSense se procesó mediante MediaPipe FaceLandmarker. A partir de los *landmarks* faciales se obtuvieron indicadores de cierre ocular, apertura de la boca, pose de la cabeza, dirección de la mirada y calidad visual. Las características se agregaron por segmento mediante estadísticos descriptivos y tasas de eventos por minuto. Finalmente, se agruparon en familias de señal.

La hipótesis **H3** se evaluó mediante una línea base personalizada de clasificación entre *Estacionario* y *Difícil*, excluyendo *Fácil* por su carácter intermedio. Para este fin se empleó regresión logística con distintas combinaciones de EDA,

PPG y temperatura cutánea, siguiendo una evaluación LOSO. La imputación y la estandarización se ajustaron con los datos de entrenamiento para evitar fugas de datos. El rendimiento se resumió mediante exactitud balanceada, F1 y ROC-AUC, reportando media y desviación estándar entre pliegues válidos.

3.5. Participante, ética y disponibilidad

Los datos de este artículo proceden de un único participante de 25 años, con certificación UAS A1/A3 y cinco años de experiencia de pilotaje. Asimismo, el participante proporcionó el consentimiento informado. El tratamiento de los datos se realizó de acuerdo con el Reglamento General de Protección de Datos (GDPR) (European Parliament and Council of the European Union, 2016).

Dado que el estudio incluye vídeo facial y señales fisiológicas, los datos se consideran potencialmente sensibles y no se distribuyen públicamente. Para favorecer la reproducibilidad del estudio, el código utilizado para el procesamiento de los datos y la implementación del modelo podrá facilitarse previa solicitud razonada a los autores.

4. Resultados y discusión

Esta sección analiza los resultados en relación con las tres hipótesis. En primer lugar, se valida si los escenarios generaron distintos niveles de carga percibida y exigencia cinemática. A continuación, se estudia la separabilidad multimodal entre condiciones. Finalmente, se evalúa si las señales fisiológicas permiten una clasificación personalizada entre sesiones.

4.1. Validación de los escenarios experimentales

Los autoinformes representados en la Figura 2 muestran una progresión incremental de la carga percibida (*Estacionario* < *Fácil* < *Difícil*), especialmente marcada en la escala RSME y en las dimensiones de demanda mental y esfuerzo del NASA-TLX. Este comportamiento apoya la hipótesis **H1**, indicando que las maniobras diseñadas indujeron niveles diferenciados de demanda operativa percibida.

En consonancia, los datos de telemetría, detallados en la Figura 3, confirman que los escenarios indujeron dinámicas de vuelo diferenciadas, mediante un incremento consistente de los perfiles de velocidad, aceleración y jerk. Esta validación cinemática demuestra que las rutas complejas exigieron maniobras más intensas y variables, permitiendo vincular las respuestas fisiológicas posteriores a demandas operativas objetivamente distintas.

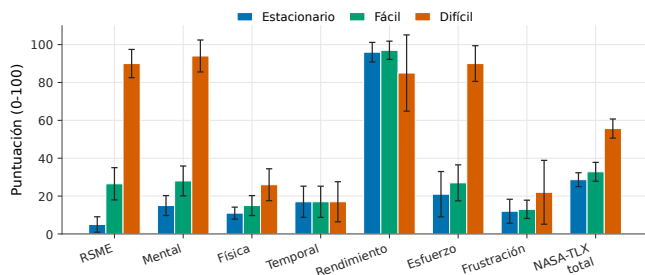


Figura 2: Carga de trabajo percibida por el operador en los escenarios *Estacionario*, *Fácil* y *Difícil*. Las barras muestran media $\pm$ desviación estándar entre sesiones. RSME se reescala linealmente de 0–150 a 0–100 para facilitar la comparación con NASA-TLX.

4.2. Separabilidad multimodal entre escenarios

En la Figura 4 se resume la capacidad de separación entre escenarios mediante el valor absoluto de la delta de Cliff ($|\delta|$), calculado a partir de las características extraídas y agrupadas por familia de señal. Esta medida no paramétrica compara todos los pares de valores entre dos condiciones. De forma intuitiva, valores cercanos a cero indican que ambas distribuciones se solapan ampliamente, mientras que valores de mayor magnitud reflejan una separación más clara. Se emplea $|\delta|$ por su capacidad de cuantificar la magnitud de la separabilidad, al margen de la dirección del cambio (Romano et al., 2006).

En el dispositivo EmotiBit, los mayores efectos se localizan en acelerometría (ACC), fundamentalmente en los contrastes que involucran la condición *Difícil*. Este resultado es coherente con la mayor exigencia motora de la ruta restringida y debe interpretarse como un indicador de interacción con el mando, no como una medida aislada de carga cognitiva. Las diferencias observadas en EDA y métricas cardíacas (CARDIO) sugieren, no obstante, que la separabilidad entre escenarios también aparece en señales fisiológicas no cinemáticas.

Por otro lado, los efectos derivados del análisis de las características de vídeo resultan más moderados. Las variaciones más relevantes aparecen en POSE y MIRADA, contrastando con los bajos valores obtenidos en CALIDAD. Este hecho sugiere que los cambios observados están más relacionados con orientación de la cabeza y atención visual que con iluminación, desenfoque o calidad de detección. En conjunto, los resultados apoyan parcialmente la hipótesis **H2**: las señales registradas muestran cambios dependientes del escenario, aunque con distinta sensibilidad entre modalidades.

Estos resultados también muestran la utilidad de mantener una adquisición multimodal. Aunque la línea base de clasificación se centra posteriormente en señales fisiológicas, la monitorización de vídeo y movimiento provee el contexto necesario para interpretar posibles fuentes de variabilidad. En particular, la presencia de efectos elevados en ACC advierte que la demanda de la tarea no puede separarse completamente de la exigencia motora del pilotaje manual. Esta observación es relevante para futuros diseños experimentales, ya que sugiere la necesidad de distinguir entre carga cognitiva, esfuerzo motor y artefactos de interacción con el mando.

Estas observaciones son coherentes con la literatura sobre carga mental en UAV, donde se han usado escalas subjetivas, indicadores oculares, señales fisiológicas y aprendizaje automático (Russo et al., 2024; Singh et al., 2021; Dell’Agnola et al., 2022; Kostenko et al., 2022). La diferencia principal de este trabajo es su enfoque intrasujeto, repetido en varios días y centrado en un operador objetivo. Esta perspectiva es relevante en biosensado, donde la variabilidad entre personas y sesiones puede limitar la utilidad de modelos generalistas (Fisher et al., 2018; Bolpagni et al., 2024).

4.3. Clasificación personalizada entre sesiones

La Tabla 1 presenta la clasificación binaria entre *Estacionario* y *Difícil*, excluyendo el escenario intermedio *Fácil*. La evaluación se realizó mediante LOSO, reservando una sesión completa para prueba en cada pliegue. Por tanto, los resultados reflejan generalización entre días de registro, no ajuste dentro de una misma sesión.

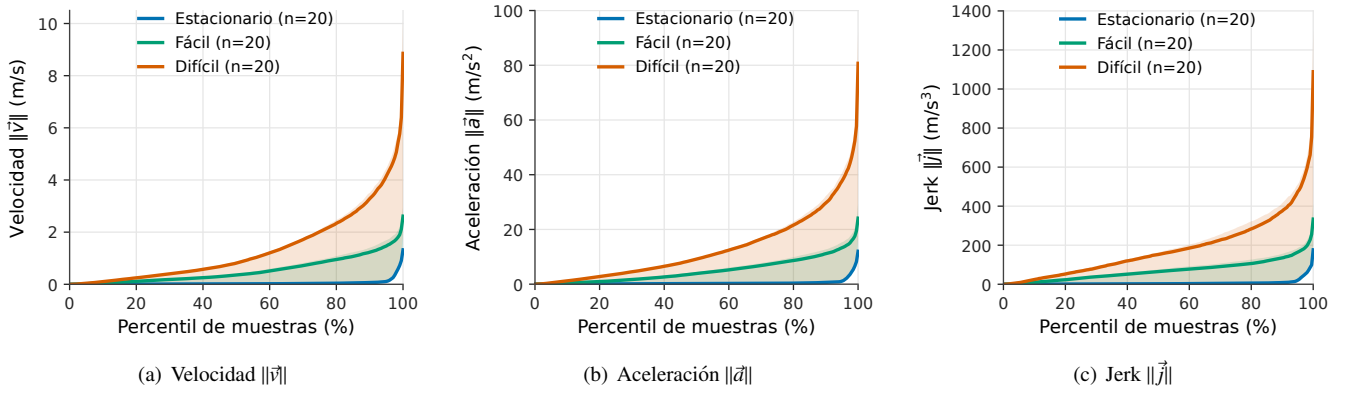


Figura 3: Perfiles percentílicos de variables cinemáticas del UAV por escenario: (a) velocidad, (b) aceleración y (c) jerk. Las curvas resumen la distribución intrasegmento y se agregan entre repeticiones mediante la mediana y el rango intercuartílico.

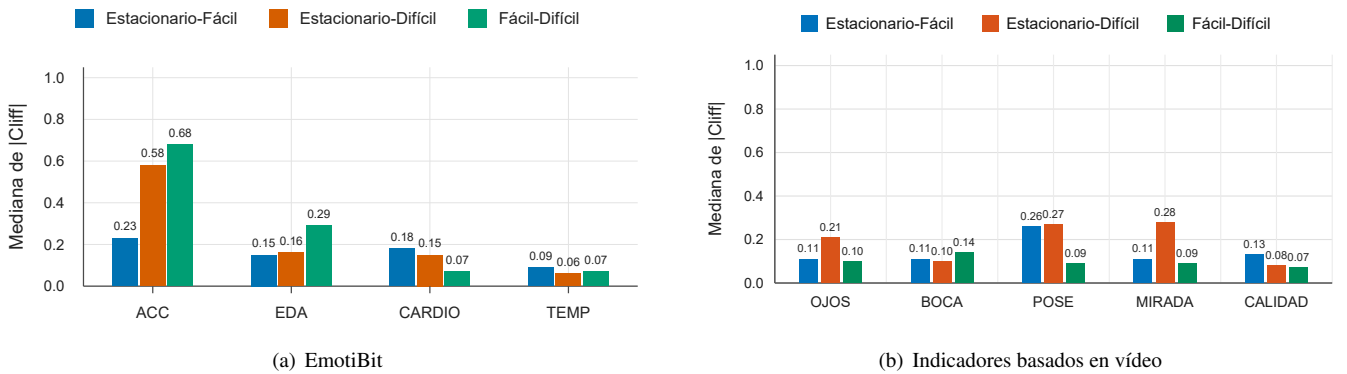


Figura 4: Separabilidad entre escenarios para las modalidades registradas. Cada panel muestra comparaciones binarias entre *Estacionario*, *Fácil* y *Difícil* mediante la mediana de $|\delta|$ de Cliff entre características de cada familia o grupo de señal.

Tabla 1: Clasificación binaria entre *Estacionario* y *Difícil* mediante regresión logística. Se muestra media $\pm$ desviación estándar en $n = 9$ pliegues LOSO válidos.

Config.	Bal. Acc	F1	ROC-AUC
EDA	0.733 $\pm$ 0.127	0.620 $\pm$ 0.286	0.872 $\pm$ 0.096
PPG	0.854 $\pm$ 0.133	0.809 $\pm$ 0.265	0.966 $\pm$ 0.026
TEMP	0.604 $\pm$ 0.162	0.611 $\pm$ 0.158	0.664 $\pm$ 0.181
EDA+PPG	0.910$\pm$0.061	0.906$\pm$0.070	0.985$\pm$0.019
EDA+TEMP	0.750 $\pm$ 0.107	0.746 $\pm$ 0.100	0.854 $\pm$ 0.120
PPG+TEMP	0.892 $\pm$ 0.061	0.883 $\pm$ 0.083	0.964 $\pm$ 0.027
EDA+PPG+TEMP	0.889 $\pm$ 0.078	0.884 $\pm$ 0.082	0.982 $\pm$ 0.020

Los mejores resultados se obtienen con **EDA+PPG**, con una exactitud balanceada de 0.910 ± 0.061 y un ROC-AUC de 0.985 ± 0.019 . PPG también muestra un rendimiento elevado individualmente, mientras que TEMP obtiene resultados más limitados, coherentes con su dinámica más lenta y periférica. La combinación de todas las señales no mejora a EDA+PPG, lo que indica que añadir modalidades no garantiza mayor rendimiento en este conjunto de datos.

Esta línea base no utiliza acelerometría ni canales IMU del dispositivo vestible, sino únicamente combinaciones de EDA, PPG y temperatura cutánea. Por tanto, la discriminación entre *Estacionario* y *Difícil* no depende directamente del movimiento de muñeca. El operador mantuvo la misma posición corporal durante todos los experimentos, reduciendo así la variabilidad asociada a movimientos globales. No obstante, la condición *Difícil* requirió una actuación más frecuente e in-

tensa sobre los *joysticks*, lo que pudo introducir diferencias residuales de origen motor en la calidad o variabilidad de las señales fisiológicas.

Más allá del rendimiento de la regresión logística como clasificador, el hallazgo fundamental es que un modelo de línea base simple es capaz de identificar patrones discriminadores en señales fisiológicas capturadas durante vuelos reales. Este hecho refuerza la idea de que los modelos personalizados pueden ser un paso intermedio útil antes de abordar modelos generalizables a múltiples operadores.

4.4. Interpretación y limitaciones

Las observaciones de este trabajo muestran convergencia entre la percepción subjetiva, la dinámica del vuelo y las respuestas de biosensado. Esta coherencia refuerza la interpretación de que las señales fisiológicas contienen información útil para la monitorización personalizada de operadores UAV. Frente a las aproximaciones tradicionales basadas en modelos generalistas o registros puntuales, este estudio aporta una validación intrasujeto, evaluada a lo largo de múltiples jornadas y orientada a un operador objetivo.

No obstante, estas evidencias constituyen un análisis de viabilidad preliminar que debe interpretarse con cautela. El estudio se limita a un único operador con licencia y a un entorno interior controlado, por lo que no permite extraer conclusiones poblacionales. Además, los bloques de 2 minutos evalúan

cambios de carga a corto plazo, pero no fatiga sostenida ni estrés prolongado. Finalmente, la condición *Difícil* combina mayor dificultad perceptiva, decisional y motora, por lo que no permite atribuir de forma causal las diferencias fisiológicas exclusivamente a carga cognitiva. En futuros protocolos será necesario introducir tareas que desacoplen parcialmente estos factores, por ejemplo comparando maniobras con demanda motora similar pero distinta complejidad decisional, o tareas de alta precisión de control con baja incertidumbre cognitiva.

5. Conclusiones y trabajo futuro

Este trabajo presenta una validación preliminar de un protocolo intrasujeto para estudiar la carga mental de un operador UAV mediante biosensado de muñeca durante vuelos reales. Los resultados muestran una progresión coherente de la carga percibida entre *Estacionario*, *Fácil* y *Difícil*, respaldada además por la telemetría, que confirma perfiles cinemáticos progresivamente más exigentes.

Las señales registradas también reflejan cambios dependientes de la condición, aunque con distinta sensibilidad entre modalidades. ACC captura principalmente diferencias asociadas a la exigencia motora, mientras que EDA y CARDIO sugieren variaciones fisiológicas relacionadas con la demanda. Además, la evaluación personalizada con LOSO muestra que EDA y PPG permiten discriminar entre *Estacionario* y *Difícil* en sesiones no vistas, alcanzando una exactitud balanceada de 0.910 ± 0.061 .

En conjunto, estas observaciones indican que este protocolo personalizado, repetido entre días y aplicado a vuelos reales, puede revelar patrones fisiológicos coherentes con la carga de la tarea. Como trabajo futuro, se ampliará el estudio a más operadores y se abordarán modelos de carga más completos, incluyendo enfoques multiclasa, continuos y de fusión multimodal robusta. Además, se valorará con mayor detalle la influencia de las demandas cognitiva y motora en la interpretación de las respuestas fisiológicas y los modelos desarrollados.

Agradecimientos

Esta investigación está financiada por el Gobierno de España a través de los proyectos PID2024-157191OB-C21 (financiado por MICIU/AEI/10.13039/501100011033 y FEDER/UE) y PID2022-140554OB-C32 (financiado por MCIN/AEI/10.13039/501100011033); por la Comunidad de Madrid a través de los proyectos SEGVAUTO 5Gen-CM (TEC-2024/ECO-277) y DIGIMATER-CM (TEC-2024/TEC-102); por el Ministerio de Ciencia, Innovación y Universidades (Gobierno de España) mediante la ayuda de Formación de Profesorado Universitario (FPU) FPU23/02872; y por la Universidad Carlos III de Madrid mediante el Programa de Investigación Predoctoral en Formación en Inteligencia Artificial (PIPF-IA).

Referencias

Anders, C., Bhaduri, I., Arnrich, B., 2025. Generalised machine learning models outperform personalised models for cognitive load classification in real-life settings. *Frontiers in Digital Health* 7, 1650085. DOI: 10.3389/fdgth.2025.1650085

Bainbridge, L., 1983. Ironies of automation. In: *Analysis, design and evaluation of man-machine systems*. Elsevier, pp. 129–135. DOI: 10.1016/B978-0-08-029348-6.50026-9

Bolpagni, M., Pardini, S., Dianti, M., Gabrielli, S., 2024. Personalized stress detection using biosignals from wearables: A scoping review. *Sensors* 24 (10), 3221. DOI: 10.3390/s24103221

Das Chakladar, D., Roy, P. P., 2024. Cognitive workload estimation using physiological measures: a review. *Cognitive Neurodynamics* 18 (4), 1445–1465. DOI: 10.1007/s11571-023-10051-3

Dell'Agnola, F., Jao, P.-K., Arza, A., Chavarriaga, R., Millan, J. d. R., Floreano, D., Atienza, D., 2022. Machine-learning based monitoring of cognitive workload in rescue missions with drones. *IEEE Journal of Biomedical and Health Informatics* 26 (9), 4751–4762. DOI: 10.1109/JBHI.2022.3186625

European Parliament, Council of the European Union, Apr. 2016. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 (general data protection regulation). OJ L 119, 4 May 2016, pp. 1–88. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

Fisher, A. J., Medaglia, J. D., Jeronimus, B. F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences* 115 (27), E6106–E6115. DOI: 10.1073/pnas.1711978115

Giorgi, A., Ronca, V., Vozzi, A., Sciaraffa, N., Di Florio, A., Tamborra, L., Simonetti, I., Aricò, P., Di Flumeri, G., Rossi, D., et al., 2021. Wearable technologies for mental workload, stress, and emotional state assessment during working-like tasks: A comparison with laboratory technologies. *Sensors* 21 (7), 2332. DOI: 10.3390/s21072332

Gruden, T., Stojmenova, K., Sodnik, J., Jakus, G., 2019. Assessing drivers' physiological responses using consumer grade devices. *Applied Sciences* 9 (24), 5353. DOI: 10.3390/app9245353

Hart, S. G., Staveland, L. E., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: *Advances in psychology*. Vol. 52. Elsevier, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9

Kostenko, A., Rauffet, P., Coppin, G., 2022. Supervised classification of operator functional state based on physiological data: Application to drones swarm piloting. *Frontiers in Psychology* 12, 770000. DOI: 10.3389/fpsyg.2021.770000

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., Schork, N. J., 2011. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized medicine* 8 (2), 161–173. DOI: 10.2217/pme.11.7

Luzzani, G., Buraioli, I., Guglieri, G., Demarchi, D., 2024. EDA, ppg and skin temperature as predictive signals for mental failure by a statistical analysis on stress and mental workload. *IEEE Open Journal of Engineering in Medicine and Biology* 6, 248–255. DOI: 10.1109/OJEMB.2024.3515473

Maciejewska, M., 2023. Uav operators' gaze behavior and workload in simulation flights—a preliminary study. *Transportation Research Procedia* 75, 134–141. DOI: 10.1016/j.trpro.2023.12.016

Parasuraman, R., Riley, V., 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39 (2), 230–253. DOI: 10.1518/001872097778543886

Rauffet, P., Botzer, A., Kostenko, A., Chauvin, C., Coppin, G., et al., 2016. Eye activity measures as indicators of drone operators' workload and task completion strategies. In: *Proc. Hum. Factors Ergonom. Soc. Eur. Ch.* pp. 211–227.

Romano, J., Kromrey, J. D., Coraggio, J., Skowronek, J., 2006. Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences on the nsse and other surveys. In: *annual meeting of the Florida Association of Institutional Research*. Vol. 177.

Russo, A. C., Junior, M. C., Villani, E., 2024. Eye-tracking analysis to assess the mental load of unmanned aerial system operators: systematic review and future directions. *The Aeronautical Journal*, 1–30. DOI: 10.1017/aer.2024.122

Singh, G., Chanel, C. P., Roy, R. N., 2021. Mental workload estimation based on physiological features for pilot-uav teaming applications. *Frontiers in Human Neuroscience* 15, 692878. DOI: 10.3389/fnhum.2021.692878

Zijlstra, F. R. H., 1993. Efficiency in work behaviour: A design approach for modern tools. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.