

Jornadas de Automática

Modelos de visión-lenguaje para optimizar el confort térmico en oficinas

María del Mar Castilla^{a,*}, Agustín Pérez-Castro^a, Juana López-Redondo^a, José Domingo Álvarez^a

^aDepartamento de Informática, CIESOL-CEINSA—ceiA3, Universidad de Almería, Ctra. Sacramento s/n, La Cañada de San Urbano, Almería, 04120, España

To cite this article: Castilla, M., Pérez-Castro, A., López-Redondo, J., Álvarez, J.D. 2026. Vision-language models for optimising thermal comfort in offices. *Jornadas de Automática*, 47.
<https://doi.org/10.17979/ja-cea.2026.47.13720>

Resumen

Para estimar el confort térmico de los usuarios de un recinto cerrado se recomienda hacer uso del índice del Voto Medio Predicho. Este índice se calcula usando dos variables que dependen exclusivamente de la componente humana, la tasa metabólica y el aislamiento de la ropa. Generalmente, estas variables se asumen estáticas en base a los estándares. Este trabajo presenta una metodología innovadora para la optimización del cálculo de este índice mediante la integración de Modelos de Visión-Lenguaje para estimar estos parámetros de forma dinámica a partir de imágenes en tiempo real. Concretamente, se ha evaluado el desempeño de cuatro modelos recientes con distintos tamaños y costes a través de la API unificada de *OpenRouter* frente a una verdad teórica (*Ground Truth*) obtenida mediante validación manual por parte de un humano. Los resultados obtenidos muestran como los modelos abiertos ofrecen un rendimiento aceptable apto para despliegues locales, con coste de operación cero y garantizando la privacidad de los usuarios del recinto.

Palabras clave: Confort térmico, Modelos de Visión y Lenguaje, Automatización en edificios, Inteligencia artificial, Optimización.

Vision-language models for optimising thermal comfort in offices

Abstract

To estimate occupants' thermal comfort sensation in an enclosed space, it is recommended to use the Predicted Mean Vote index. This index is calculated using, among other variables, two variables that depend exclusively on the human component: metabolic rate and the insulation provided by clothing. Generally, these variables are assumed to be static, based on tables included in the standards. In this work, an innovative methodology for optimising the calculation of this index by integrating Vision-Language Models to estimate these parameters dynamically from real-time images is presented. Specifically, the performance of four recent models of different sizes and costs has been systematically evaluated via the unified *OpenRouter* API against a theoretical ground truth obtained through manual validation by a human. The results show that open-source models offer an acceptable performance suitable for local deployments, with zero operational costs and guaranteeing the occupants' privacy.

Keywords: Thermal comfort, Vision-Language models, Building Automation, Artificial intelligence techniques, Optimisation.

1. Introducción

El control eficiente de los sistemas de climatización (HVAC, del inglés *Heating, Ventilation and Air Conditioning*) es un gran desafío para la reducción de la huella de carbono global. Según estudios recientes, los sistemas HVAC consumen más del 40 % de la energía final en edificios comerciales e institucionales (Borodinecs et al., 2024). Sin embargo, una

parte significativa de esa energía se desperdicia debido a estrategias de control rígidas que calientan o enfrían en exceso los distintos espacios afectando negativamente, a la productividad y al confort térmico de los ocupantes.

Para estimar la sensación térmica de los ocupantes en un recinto cerrado, el estándar internacional ISO 7730 (ISO Standard 7730, 2025) propone como método analítico el índice del

*Autor para correspondencia: mcastilla@ual.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Voto Medio Predicho (PMV, del inglés *Predicted Mean Vote*). Para calcular el índice PMV de manera precisa se requieren seis variables, cuatro de las cuales son de naturaleza física (temperatura del aire, temperatura radiante media, humedad relativa y velocidad del aire) y se pueden medir con exactitud en tiempo real de forma fácil mediante el uso de una red de sensores apropiada. Sin embargo, las dos variables restantes dependen exclusivamente de los ocupantes del recinto, la tasa metabólica y el aislamiento térmico proporcionado por la ropa. En la práctica, estas variables se consideran estáticas según tablas que dependen del tipo de actividad realizada por los usuarios y la estación del año. Sin embargo, esta aproximación ignora la diversidad de vestimentas y actividades en espacios compartidos, como una oficina o una sala de reuniones, así como la alta variabilidad que se produce a lo largo de un día (cambios de actividad, quitarse una chaqueta, etc.). Por tanto, considerar estas variables constantes introduce desviaciones en el cálculo del índice PMV que afectan directamente al confort térmico de los usuarios y a la eficiencia de los sistemas HVAC.

En la literatura se pueden encontrar varios trabajos que abordan la estimación dinámica no intrusiva de estos parámetros mediante el uso de técnicas de visión artificial clásica y redes neuronales convolucionales (CNN del inglés *Convolutional Neural Networks*) (Choi et al., 2022; Liu et al., 2022; Wei et al., 2024). No obstante, aunque estos trabajos logran precisiones aceptables en entornos controlados, necesitan grandes bases de datos etiquetadas para entrenamiento y carecen de robustez antes cambios rápidos en la moda o en la actividad realizada en el recinto. Los Modelos de Lenguaje y Visión (VLM del inglés *Vision-Language Models*) utilizados en un paradigma *zero-shot* (sin entrenamiento previo) permiten superar estas limitaciones al razonar directamente sobre estímulos visuales, infiriendo actividades y prendas de ropa con gran precisión (Zhang et al., 2025). En este trabajo se propone una arquitectura para la estimación dinámica del confort térmico mediante la integración de VLMs a través de la API unificada de OpenRouter. Concretamente, se incluye una comparación de cuatro VLMs: los comerciales *gemini-3.5-flash* y *gemini-3.1-flash-lite*, junto a los modelos abiertos *qwen/qwen3-vl-8b-instruct* y *google/gemma-4-26b-a4b-it*, que se evalúan frente a un *Ground Truth* de referencia establecido mediante validación manual por un ser humano. Para ello, se ha hecho uso de un dataset de 50 imágenes sintéticas realistas de oficinas en las que se ha ido variando la ocupación, la actividad y la vestimenta de los usuarios. La principal diferencia con respecto a los trabajos publicados radica en que se realiza una evaluación de los errores promedio agregados a nivel de oficina tras un emparejamiento persona a persona. Asimismo, se analiza el impacto de las discrepancias en el índice PMV y se estudia la viabilidad económica y técnica del despliegue de modelos abiertos frente a comerciales. Finalmente, se incluye una validación de la metodología propuesta utilizando datos e imágenes reales tomadas en el centro de investigación CIESOL ubicado en el campus de la Universidad de Almería (UAL).

El resto del artículo se estructura de la siguiente forma: la sección 2 detalla el procedimiento para calcular el índice PMV. En la sección 3 se explica el proceso de obtención del dataset sintético. La sección 4 está dedicada a explicar la ar-

quitectura propuesta. En la sección 5 se muestran y se analizan los resultados obtenidos. Finalmente, en la sección 6 se incluyen las principales conclusiones y futuros trabajos.

2. Confort térmico: índice PMV

El confort térmico hace referencia a la condición mental que expresa satisfacción con el ambiente térmico que rodea a una persona. Para determinar el confort térmico, uno de los índices más extendidos es el índice PMV propuesto por Fanger (ISO Standard 7730, 2025). Este índice se basa en un balance térmico del cuerpo humano, de forma que la sensación de confort se alcanza cuando la producción interna de calor iguala a las pérdidas hacia el entorno. Para ello, Fanger propuso una relación empírica entre la carga térmica corporal y la sensación térmica en una escala de 7 puntos (de muy frío a muy caliente), véase (1):

$$PMV = [0,303 \cdot \exp(-0,036 \cdot M) + 0,028] \cdot L \quad (1)$$

donde M es la tasa metabólica en W/m^2 ($1 \text{ met} = 58,15 \text{ W/m}^2$) y L representa la carga térmica corporal, es decir, la diferencia entre el calor producido y el calor disipado por sudoración, respiración, etc. Esta carga, L , se estima como una función no lineal que depende de cuatro variables de naturaleza física: temperatura del aire (T_a), temperatura radiante media (T_r), velocidad del aire (v_a) y la humedad relativa (RH); y dos variables que dependen del usuario: la tasa metabólica (M) y el aislamiento de la ropa (I_{cl}) ($1 \text{ clo} = 0,155 \text{ m}^2 K/W$), véase (2):

$$L = f(T_a, T_r, v_a, RH, M, I_{cl}) \quad (2)$$

El estándar ISO 7730 recomienda mantener el índice PMV entre $-0,5$ y $0,5$. Asimismo, para tareas típicas de oficina se suele establecer una actividad metabólica, M , en torno a $1,2 \text{ met}$, mientras que, el aislamiento de la ropa, I_{cl} , se fija a un valor estático que depende de la estación del año ($I_{cl} = 0,5 \text{ clo}$ para verano e $I_{cl} = 1,0 \text{ clo}$ para invierno). No obstante, como se mencionó anteriormente, en ambientes dinámicos reales, asumir variables estáticas ignora la variabilidad y diversidad de los distintos usuarios. Por ejemplo, en una oficina considerando condiciones ambientales típicas de verano, una persona sentada con ropa típica de la época ($M \approx 1 \text{ met}$ e $I_{cl} \approx 0,35 \text{ clo}$) tendría un $PMV \approx -0,5$. De la misma forma, una persona archivando o moviendo cajas de pie con ropa de verano pero más abrigada ($M \approx 1,4 \text{ met}$ e $I_{cl} \approx 0,5 \text{ clo}$) implicaría un $PMV \approx 0,25$. Como se puede ver, ambas personas experimentan sensaciones muy dispares por lo que la arquitectura propuesta en este trabajo permite calcular un índice PMV promedio para ajustar el sistema HVAC en base a la ocupación, el grado de vestimenta y la actividad real de los usuarios. Más información sobre el cálculo del índice PMV se puede encontrar en (Castilla et al., 2014).

3. Generación del dataset sintético de imágenes

Para evaluar la capacidad de los VLMs en un entorno diverso y ante condiciones diferentes, se ha generado un dataset compuesto por 50 imágenes realistas de oficinas mediante

el modelo *google/gemini-3.1-flash-image-preview* (Nano Banana 2). Concretamente, las imágenes seleccionadas simulan diferentes densidades de ocupación (entre 1 y 15 personas por imagen), con un 50 % de imágenes en las que los usuarios llevan vestimenta de verano y otro 50 % con vestimenta de invierno. Además, en las imágenes se abarca una mezcla de actividades comunes de oficina (teclear sentado, escribir en cuaderno, charlar de pie, escribir en pizarra, archivar en carpetas, etc.). Adicionalmente, se han considerado una serie de variaciones representando personas de distintas etnias y un 15 % de ocupantes con talla grande u obesidad. Para ello, se diseñaron *prompts* para el modelo empleado teniendo en consideración los requisitos anteriores. En la figura 1 se puede observar una muestra de las imágenes sintéticas utilizadas para la comparación de los VLMs.



Figura 1: Muestra imágenes sintéticas generadas con Nano Banana 2

4. Arquitectura basada en VLMs para el confort térmico

En este trabajo se propone una arquitectura que permite la optimización del cálculo del índice PMV mediante la estimación dinámica de los parámetros que dependen exclusivamente de los usuarios: la tasa metabólica y el aislamiento proporcionado por la ropa. Para ello, se realiza la integración de VLMs para, a partir de imágenes y un *prompt* estructurado, calcular esos parámetros. La arquitectura resultante está compuesta por cinco módulos principales que se pueden observar en la figura 2. El funcionamiento de cada uno de los módulos que componen la arquitectura se detalla a continuación:

1. **Adquisición de la imagen.** Mediante un dispositivo de monitorización se captura la escena y se envía al módulo de evaluación VLM. Recaltar que para la comparación inicial de los VLMs se ha utilizado el dataset sintético descrito en la sección 3. Posteriormente, se han validado los resultados obtenidos utilizando imágenes tomadas en un edificio real, el edificio CIESOL.
2. **Evaluación VLM.** La imagen proporcionada por el módulo de adquisición de la imagen se inyecta junto a un *prompt* estructurado, véase sección 4.1, para la estimación de M e I_{cl} a través de la API unificada de OpenRouter dirigiéndose al VLM seleccionado. Un resumen de los modelos seleccionados y sus características principales pueden observarse en la sección 4.2.
3. **Estructuración y normalización.** En algunos casos, es posible que el módulo evaluación VLM devuelva

JSONs sintácticamente corruptos (llaves desalineadas o comas omitidas). Para ello, se ha desarrollado este módulo que reconstruye la salida aplicando reglas de normalización y manteniendo estructuradas las descripciones en inglés. Concretamente, el JSON que devuelve el VLM se depura mediante una llamada síncrona a un gran modelo del lenguaje (LLM), el *gemini-3.5-flash*, con el objetivo de garantizar la generación de un JSON estructurado sintácticamente sin errores, traduciendo actividades y normalizando las variables asociadas a M e I_{cl} .

4. **Estimación del índice PMV.** A continuación, los parámetros determinados por el módulo de evaluación VLM, es decir, M e I_{cl} , se utilizan junto con los datos procedentes de los sensores (T_a , T_r , v_a y RH) para estimar el índice PMV.
5. **Control climático.** Finalmente, el valor del índice PMV se puede enviar al módulo que se encarga de realizar el control climático del edificio optimizando así el uso de los sistemas HVAC.

4.1. Diseño del prompt estructurado para la obtención de M e I_{cl}

A la hora de interactuar con los VLMs es de vital importancia diseñar un *prompt* capaz de reducir alucinaciones y mejorar la alineación con el estándar ISO 7730. Por tanto, para guiar el razonamiento *zero-shot* multimodal, se ha diseñado un *prompt* estructurado en inglés que se puede observar en el siguiente bloque:

Prompt de inferencia VLM utilizado

```
Analyze the image and estimate the thermal comfort of the occupants according to the UNE-EN ISO 7730:2025 standard. Return ONLY a valid JSON object, without any introductory or concluding text, and without Markdown code blocks (do not use '```json' or '```'). The JSON structure must be exactly like this:
{
  "total_people": 2,
  "people": [
    {
      "id": 1,
      "description": "Man sitting typing",
      "metabolic_rate_met": 1.2,
      "clothing_insulation_clo": 0.5
    },
    {
      "id": 2,
      "description": "Woman standing conversing",
      "metabolic_rate_met": 1.4,
      "clothing_insulation_clo": 0.4
    }
  ],
  "mean_met": 1.3,
  "mean_clo": 0.45
}
```

4.2. Modelos de Visión y Lenguaje

En este trabajo se han comparado cuatro modelos de última generación tanto comerciales como abiertos. Más concretamente, los modelos utilizados son: *gemini-3.5-flash*, *gemini-3.1-flash-lite*, *Qwen-3-vl-8b-instruct* y *Gemma-4-26b-it*. Sus principales características se observan en la tabla 1.

En general, la elección entre los distintos VLMs va a depender de la aplicación de los mismos. A grandes rasgos, *gemini-3.5-flash* y *gemini-3.1-flash-lite* son buenas opciones para procesar grandes volúmenes de datos multimedia en los

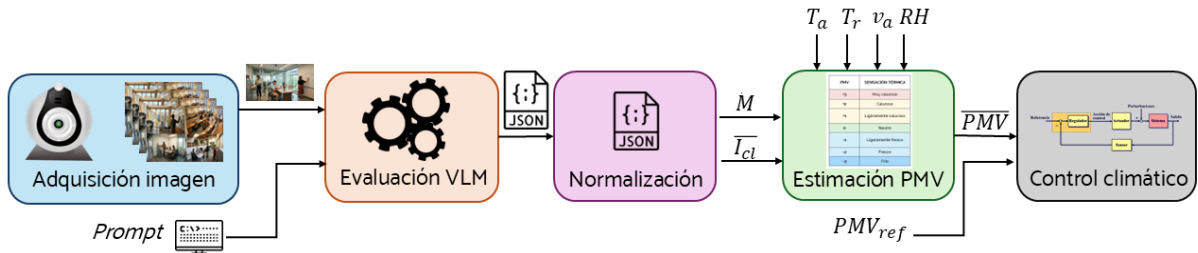


Figura 2: Arquitectura basada en VLMs para la optimización del cálculo del confort térmico

Tabla 1: Características principales de los VLMs utilizados para el cálculo de M e I_{cl} .

Modelo	Despliegue	Tamaño (Parám.)	Coste Entrada*	Coste Salida*
Modelos de API Comercial				
gemini-3.5-flash	Nube (Google)	Desconocido	\$0,075	\$0,30
gemini-3.1-flash-lite	Nube (Google)	Desconocido	\$0,030	\$0,100
Modelos Abiertos (Local)				
gemma-4-26b-it	Local (Libre)	26B	\$0,00 (Gratis)	\$0,00 (Gratis)
qwen-3-vl-8b	Local (Libre)	8B	\$0,00 (Gratis)	\$0,00 (Gratis)

* 1 millón tokens

que el primero busca la máxima velocidad y el segundo reducir costes. Por otro lado, si se prefiere un despliegue local y de código abierto, *Gemma-4-26b-it* ofrece un equilibrio entre la potencia de razonamiento de un modelo grande y la eficiencia de hardware gracias a su arquitectura de Mezcla de Expertos (MoE).

5. Resultados y discusión

Esta sección se encuentra organizada de la siguiente forma: en primer lugar, se ha incluido un ejemplo de respuesta del *prompt* estructurado para una de las imágenes sintéticas consideradas en el estudio evaluada con el modelo *gemini-3.1-flash-lite*. Posteriormente, se incluye un análisis estadístico y una comparación de los resultados obtenidos al evaluar las 50 imágenes sintéticas utilizando los cuatro modelos mencionados anteriormente. Finalmente, se muestran los resultados obtenidos al aplicar el modelo seleccionado a una imagen real capturada en la sala de reuniones del edificio CIESOL.



Figura 3: Imagen sintética que simula una oficina con ocupación media y vestimenta de verano

Cabe destacar que la influencia de las variables ambientales de la habitación no es objeto de estudio en este trabajo, por lo que se han asumido constantes en un escenario estandarizado de confort térmico Clase B, es decir, $T_a = T_r = 22,0^{\circ}\text{C}$,

$v_a = 0,1 \text{ m/s}$ y $RH = 50\%$ (ISO Standard 7730, 2025). Además, la estimación del índice PMV se ha incluido para evaluar la sensibilidad y propagación de errores de los VLMs.

5.1. Caso de Estudio: Imagen sintética evaluada con *gemini-3.1-flash-lite*

El primer paso ha sido analizar la validez y robustez del *prompt* estructurado que devuelve el módulo de estructuración y normalización. Para ilustrar el proceso, se ha incluido como ejemplo la respuesta asociada a una de las imágenes sintéticas, véase figura 3, evaluada con el VLM *gemini-3.1-flash-lite*. El resultado obtenido se puede observar a continuación:

Ejemplo de salida de *Prompt* estructurado

```
{
  "total_people": 3,
  "people": [
    {
      "id": 1,
      "description": "Woman standing at filing cabinet",
      "metabolic_rate_met": 1.6,
      "clothing_insulation_clo": 0.6
    },
    {
      "id": 2,
      "description": "Man sitting at table writing",
      "metabolic_rate_met": 1.2,
      "clothing_insulation_clo": 0.6
    },
    {
      "id": 3,
      "description": "Woman standing at whiteboard writing",
      "metabolic_rate_met": 1.7,
      "clothing_insulation_clo": 0.7
    }
  ],
  "mean_met": 1.5,
  "mean_clo": 0.633
}
```

5.2. Análisis de resultados obtenidos

Una vez realizada la inferencia del dataset total de imágenes con los cuatro modelos evaluados, los resultados obtenidos se han analizado haciendo principal hincapié en el rendimiento de la estimación de parámetros, M e I_{cl} , la precisión

Tabla 2: Comparativa experimental de rendimiento medio de los VLMs.

Modelo	Coste medio/Img	MAE (N_p)	MAE (M)	MAE (I_{cl})	MAE (PMV)
gemini-3.5-flash	0,00015\$	0,16	0,016 clo	0,007 met	0,037
gemini-3.1-flash-lite	0,00008\$	0,30	0,078 clo	0,109 met	0,219
gemma-4-26b-it	0\$	0,26	0,170 clo	0,103 met	0,488
qwen-3-vl-8b	0\$	0,66	0,156 clo	0,088 met	0,406

Tabla 3: Detalle comparativa imágenes.

Imagen	Ground Truth		gemini-3.5-flash		gemini-3.1-flash-lite		qwen-3-vl-8b		gemma-4-26b-it	
	N_p	PMV	N_p	PMV	N_p	PMV	N_p	PMV	N_p	PMV
Img01	1	0,40	1	0,40	1	0,42	1	0,16	1	-0,37
Img02	1	-0,12	1	-0,12	1	0,10	1	-0,57	1	-0,85
Img03	2	0,26	2	0,26	2	0,40	2	-0,47	2	-0,35
Img05	4	0,14	4	-0,08	4	-0,10	4	-0,45	4	-0,68
Img26	6	-0,59	6	-0,42	6	-0,07	6	-0,91	6	-0,41
Img29	12	-0,52	12	-0,52	11	-0,98	10	-0,87	12	-0,97

en el conteo de ocupantes y la viabilidad del despliegue local frente al uso de VLMs comerciales. La comparación de los modelos, se ha realizado frente a un *Ground Truth* obtenido mediante la validación manual del dataset por un humano.

Concretamente, se han incluido dos tablas que resumen los resultados obtenidos. Por un lado, la tabla 2 muestra un análisis global del desempeño de estimación de los parámetros: número de personas (N_p), tasa metabólica (M) y aislamiento de la ropa (I_{cl}) mediante el cálculo del error absoluto medio (MAE, del inglés *Mean Absolute Error*) obtenido al evaluar las 50 imágenes del dataset sintético. Además, también se incluye el MAE del índice PMV para analizar el error propagado por los VLMs. Finalmente, en esta tabla también se incluye el coste operativo por imagen. Por otro lado, la tabla 3 proporciona datos empíricos en una selección de imágenes críticas del dataset, abarcando diferentes densidades de ocupación, actividades y tipos de vestimenta. Más en detalle, las imágenes *Img01*, *Img02*, *Img03* e *Img05* se corresponden con el periodo de invierno con una densidad de ocupación baja y alternando distintas actividades (escribir en pizarra, en libreta) y ocupantes con talla grande. Las imágenes *Img26* e *Img29* presentan una alta densidad de ocupación para el periodo de verano y abarcan gran diversidad de actividades.

Lo primero es verificar que los VLMs comparados no omitan ocupantes, ya que, ese dato es de vital importancia para garantizar que el cálculo del PMV promedio de un recinto sea representativo. Si un VLM no es capaz de detectar visualmente a todos los individuos, el cálculo del PMV de la zona estaría sesgado hacia el subconjunto visible de usuarios ignorando las necesidades del resto. Como se puede observar en el resumen global de la tabla 2, la precisión en el conteo de personas es excelente en los modelos de mayor capacidad. El modelo *gemini-3.5-flash* presenta un error residual muy bajo frente al *Ground Truth*, $MAE(N_p) = 0,16$, lo que corrobora su excelente consistencia de razonamiento espacial. Las discrepancias para los otros modelos se mantienen también bajas (con un $MAE(N_p)$ de 0,30 y 0,26 para *gemini-3.1-flash-lite* y *gemma-4-26b-it*, respectivamente, mientras que el modelo *qwen-3-vl-8b* presenta un MAE de valor 0,66 personas, lo que asegura que prácticamente ningún ocupante quede excluido del cálculo del confort térmico dinámico.

En resumen, los resultados obtenidos demuestran la viabilidad de emplear VLMs bajo un paradigma *zero-shot* para calcular M e I_{cl} . Concretamente, el modelo *gemini-3.5-flash* demuestra una excelente concordancia con el *Ground Truth* logrando un MAE asociado al índice PMV mínimo posicionándose como una alternativa comercial eficiente y con bajo coste. Por ejemplo, suponiéndose que se toman imágenes cada 15 minutos y que la jornada laboral es de 8 horas durante 5 días a la semana, eso implicaría que una oficina realizaría 8320 peticiones a lo largo de un año, lo que supone un coste total aproximado de 1,25\$. No obstante, este modelo tiene la desventaja de tener que subir a la nube imágenes reales de una oficina lo que influye directamente en la privacidad de dichas imágenes. Los modelos de código abierto, *gemma-4-26b-it* y *qwen-3-vl-8b*, logran un desempeño aceptable aunque sin llegar a la exactitud del modelo comercial *gemini-3.5-flash*. Sin embargo, el coste es cero y, al poder realizar un despliegue local, se puede garantizar la soberanía de las imágenes.

Finalmente, la figura 4 muestra una comparación de estimaciones del índice PMV promedio para cada una de las 50 imágenes que componen el dataset sintético. A partir de esa imagen se puede inferir como el modelo que proporciona mejores resultados para todas las imágenes es el *gemini-3.5-flash*. Por otro lado, también se puede extraer como conclusión de esa figura que los modelos de código abierto proporcionan mejores resultados durante el periodo de verano y tienen mayores discrepancias durante el periodo de invierno.

5.3. Caso de Estudio: Imagen real capturada en el edificio CIESOL

Por último, se muestra la evaluación de los VLMs comparados utilizando una imagen real capturada en el edificio CIESOL, un centro de investigación localizado en el campus de la UAL (<http://www.ciesol.es>). Concretamente, la imagen fue tomada a las 11:31 horas de un día del periodo de verano en la sala de reuniones del edificio, véase figura 5. Indicar, que en el artículo se ha introducido la imagen pixelada por cuestiones de privacidad. Las condiciones ambientales en ese instante eran: $T_a = T_r = 25,8^\circ\text{C}$, $RH = 38,7\%$ y $v_a = 0,15\text{ m/s}$. Los resultados cuantitativos obtenidos se muestran en la tabla 4. Se

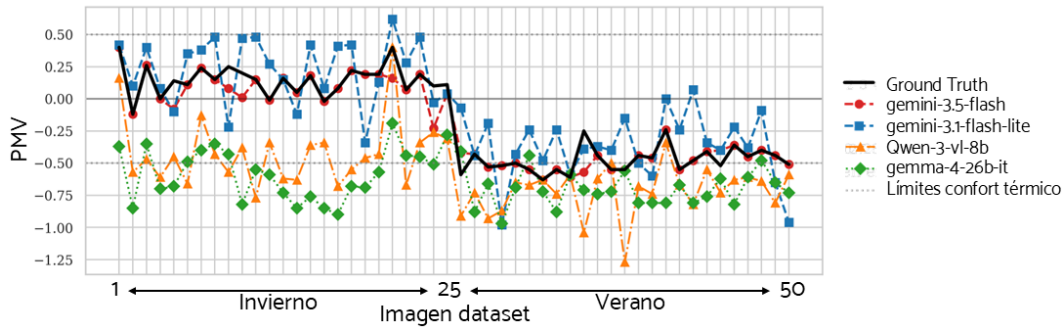


Figura 4: Comparación de estimaciones PMV dataset sintético

puede observar como, en este caso, todos los modelos proporcionan resultados similares que se ajustan con bastante precisión a la realidad.



Figura 5: Imagen real capturada en el edificio CIESOL

Tabla 4: Estimación experimental de parámetros de confort en imagen real

Modelo	N_p Est.	$\bar{M}$ (met)	$\bar{I}_{cl}$ (clo)	PMV
gemini-3.5-flash	6	1.200	0.492	0.090
gemini-3.1-flash-lite	6	1.200	0.517	0.140
qwen-3-vl-8b	6	1.167	0.517	0.080
gemma-4-26b-it	6	1.183	0.500	0.070

6. Conclusiones

En este trabajo se ha presentado un enfoque novedoso para subsanar una deficiencia a la hora de calcular el índice PMV: la consideración estática de la tasa de actividad metabólica (M) y el aislamiento proporcionado por la ropa (I_{cl}). Mediante el uso de VLMs aplicados siguiendo un paradigma *zero-shot* se ha podido demostrar la viabilidad técnica de estimar estas variables dinámicamente en tiempo real para capturar la diversidad y variabilidad de los ocupantes a lo largo de un día a partir de imágenes. Concretamente, se han comparado cuatro modelos utilizando tanto comerciales como de código abierto. Los resultados obtenidos muestran como el modelo comercial *gemini-3.5-flash* es el más preciso a un coste bajo. Sin embargo, el desempeño de los modelos de código abierto se puede considerar aceptable, siendo posible desplegar los mismos de forma local y resolviendo simultáneamente uno de los mayores problemas en este campo, la protección de la privacidad visual de los ocupantes al procesar las imágenes de forma local.

Como trabajo futuro, se proponen principalmente dos líneas diferentes: el diseño de un índice personalizado de confort térmico en tiempo real de forma descentralizada y respetando la privacidad de los usuarios. Posteriormente, se integraría ese índice personalizado de confort térmico para realizar un control multizona dinámico del sistema de climatización del edificio.

Agradecimientos

Este trabajo ha recibido apoyo financiero del proyecto PID2024-155521OB-I00 financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

Referencias

- Borodinecs, A., Palcikovskis, A., Krumins, A., Zajecs, D., Lebedeva, K., 2024. Assessment of hvac performance and savings in office buildings using data-driven method. *Clean Technologies* 6 (2), 802–813. DOI: 10.3390/cleantechno16020041
- Castilla, M., Álvarez, J. D., Rodríguez, F., Berenguel, M., 2014. *Comfort control in buildings*. Springer.
- Choi, H., Jeong, B., Lee, J., Na, H., Kang, K., Kim, T., 2022. Deep-vision-based metabolic rate and clothing insulation estimation for occupant-centric control. *Building and Environment* 221, 109345. URL: <https://doi.org/10.1016/j.buildenv.2022.109345> DOI: 10.1016/j.buildenv.2022.109345
- ISO Standard 7730, 2025. *Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*. International Organization for Standardization, Geneva, Switzerland.
- Liu, J., Foged, I. W., Moeslund, T. B., 2022. Clothing insulation rate and metabolic rate estimation for individual thermal comfort assessment in real life. *Sensors* 22 (2), 619. URL: <https://www.mdpi.com/1424-8220/22/2/619> DOI: 10.3390/s22020619
- Wei, Z., Calautit, J. K. S., Wei, S., Tien, P. W., 2024. Real-time clothing insulation level classification based on model transfer learning and computer vision for pmv-based heating system optimization through piecewise linearization. *Building and Environment* 253, 111277. URL: <https://doi.org/10.1016/j.buildenv.2024.111277> DOI: 10.1016/j.buildenv.2024.111277
- Zhang, D., Xiong, Z., Zhu, X., 2025. Evaluation of thermal comfort in urban commercial space with vision-language-model-based agent model. *Land* 14 (4), 786. URL: <https://www.mdpi.com/2073-445X/14/4/786> DOI: 10.3390/land14040786