
This is the **published version** of the bachelor thesis:

Garcia Sanchez, Carlos. *Complejidad cromático-espacial y preferencia estética visual : análisis mediante categorización cromática*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/334455>

under the terms of the  license.

Complejidad cromático-espacial y preferencia estética visual: análisis mediante categorización cromática

Carlos Garcia Sanchez

24 de junio de 2026

Resumen— Este documento presenta un análisis sobre la relación entre características cromático-espaciales y preferencia estética visual. El trabajo no busca construir un predictor universal de belleza, sino evaluar si descriptores interpretables derivados de mapas de categorización cromática capturan parte de la variación de valoraciones humanas. Se analizan tres conjuntos de datos de estética visual —AVA (Aesthetic Visual Analysis), AADB (Aesthetics and Attributes Database) y MSC (Minimum Semantic Content)— mediante cuatro variantes de categorización cromática. A partir de los mapas de etiquetas resultantes, se calculan métricas de diversidad, dominancia, fragmentación y complejidad multiescala, y se estudian relaciones lineales, monotónicas y cuadráticas con las valoraciones estéticas humanas. Los resultados muestran señales débiles en AVA/AADB y asociaciones más claras en MSC, especialmente en el conjunto completo y en subconjuntos transformados.

Palabras clave— Estética computacional; preferencia estética visual; categorización cromática; complejidad cromático-espacial; clutter visual; análisis de correlación; relaciones no lineales.

Abstract— This document presents an analysis of the relationship between chromatic-spatial com-

plexity and visual aesthetic preference. The work does not aim to build a universal beauty predictor. Instead, it evaluates whether interpretable descriptors computed from colour-category maps explain part of the variability in human aesthetic ratings. AVA, AADB and MSC are processed with four colour-categorization variants; diversity, dominance, fragmentation and multiscale complexity metrics are extracted; and linear, monotonic and quadratic relationships are analysed. Results show weak signals in AVA/AADB and clearer associations in MSC, especially in the complete dataset and transformed subsets.

Keywords— Computational aesthetics; colour categorization; visual clutter; MSC; AVA; AADB; Spearman; visual complexity.



1 INTRODUCCIÓN

LA evaluación estética de imágenes es un problema complejo porque no depende únicamente de propiedades físicas de la imagen, sino de la interacción entre información visual de bajo nivel, contenido representado, contexto y variabilidad individual. En estética computacional se han propuesto enfoques basados en características de color, composición, textura, saliencia y modelos de aprendizaje automático [2, 4, 10, 11]. Sin embargo, cuando el objetivo no es solo predecir una puntuación estética, sino interpretar el papel de propiedades visuales concretas, los

modelos puramente predictivos pueden resultar poco transparentes.

Este trabajo se centra en una pregunta más acotada: hasta qué punto determinadas propiedades cromático-espaciales, obtenidas a partir de algoritmos de categorización cromática, se relacionan con puntuaciones estéticas humanas. El punto de partida es que el color no solo puede describirse de forma continua, sino también mediante categorías perceptuales. En este sentido, el modelo de categorización cromática NICE propuesto por Parraga y Akbarinia conecta percepción y categorización del color mediante regiones cromáticas perceptualmente motivadas [7]. A partir de los mapas de etiquetas cromáticas resultantes es posible construir métricas interpretables de diversidad, dominancia, fragmentación y complejidad cromático-espacial.

Para evaluar estas métricas se utilizan tres conjuntos de datos de estética visual: AVA (*Aesthetic Visual Analysis*) [6], AADB (*Aesthetics and Attributes Database*) [5] y MSC (*Minimum Semantic Content*) [8, 9]. AVA y AADB propor-

- E-mail de contacte: Carlos Garcia Sanchez.
- Menció: Computació.
- Treball tutoritzat per: C. Alejandro Parraga.
- Curs 2025/26.

cionan un contraste con bases de datos fotográficas ampliamente utilizadas en estética computacional, mientras que MSC resulta especialmente relevante para este trabajo porque reduce y controla parcialmente la carga semántica de las imágenes. Esta diferencia permite estudiar si las métricas cromático-espaciales muestran una señal más clara cuando la puntuación estética depende menos del reconocimiento de objetos, escenas o contenido narrativo.

La contribución principal del trabajo es un análisis experimental reproducible de descriptores cromático-espaciales derivados de mapas de categorización cromática. En particular, se comparan varias variantes de categorización, se calculan métricas de diversidad, dominancia, fragmentación, frontera y complejidad multiescala, y se estudia su relación con puntuaciones estéticas humanas mediante correlaciones lineales, monotónicas y modelos cuadráticos simples. El objetivo no es defender una métrica universal de belleza, sino determinar en qué condiciones estos descriptores aportan información interpretable sobre la preferencia estética visual.

2 OBJETIVOS Y ALCANCE

El objetivo general del trabajo es analizar si la complejidad cromático-espacial puede aportar información interpretable sobre la preferencia estética visual. Este objetivo general se descompone en cinco objetivos específicos:

1. Implementar un flujo reproducible de preparación de datos, categorización cromática y extracción de métricas por imagen.
2. Comparar variantes de categorización cromática, manteniendo una nomenclatura interna clara.
3. Definir y calcular métricas de diversidad, dominancia, fragmentación, densidad de frontera y complejidad multiescala sobre los mapas de etiquetas cromáticas.
4. Evaluar cada métrica frente a las valoraciones estéticas humanas disponibles en los conjuntos de datos AVA, AADB y MSC, utilizando correlaciones de Pearson y Spearman.
5. Analizar si algunas métricas muestran relaciones no lineales con la valoración estética, mediante modelos cuadráticos simples que permitan detectar posibles patrones de U o U invertida.

El alcance del proyecto queda limitado deliberadamente. No se entrena una red neuronal para maximizar la precisión predictiva; tampoco se pretende estimar todos los componentes cognitivos de la estética. La aportación se sitúa en el análisis experimental de descriptores simples, trazables y conectados con procesos perceptuales. Esta decisión facilita la interpretación de los resultados y orienta el trabajo hacia la comprensión del fenómeno, más que hacia la competición en un benchmark predictivo.

3 TRABAJOS RELACIONADOS

La estética computacional ha estudiado la predicción y el análisis de la calidad estética a partir de rasgos visuales y

modelos aprendidos. Datta et al. [2] plantearon uno de los primeros enfoques sistemáticos para estudiar estética fotográfica mediante características computacionales. Joshi et al. [4] revisaron posteriormente la relación entre estética, emoción y análisis de imágenes desde una perspectiva de señal y visión por computador. Redies [10] defendió que la experiencia estética visual puede relacionarse con regularidades estadísticas presentes en imágenes naturales y obras artísticas, integrando componentes universales y contextuales. Más recientemente, Redies et al. [11] han publicado una herramienta para calcular propiedades cuantitativas de imagen en investigación estética, reforzando la importancia de utilizar métricas reproducibles e interpretables.

Los conjuntos de datos son un elemento crítico en este tipo de estudios, porque condicionan qué aspectos de la valoración estética pueden observarse. AVA (*Aesthetic Visual Analysis*) proporciona una base de datos grande con puntuaciones estéticas y metadatos asociados a las imágenes [6]. AADB (*Aesthetics and Attributes Database*) incorpora además atributos estéticos y adaptación al contenido para tareas de ranking estético [5]. Sin embargo, ambos conjuntos están formados por fotografías en las que la valoración humana puede depender no solo de propiedades visuales de bajo nivel, sino también del contenido representado, la escena, los objetos, las personas o la composición fotográfica. Esto es relevante porque dichos factores pueden enmascarar o diluir la relación entre métricas cromático-espaciales simples y preferencia estética. MSC (*Minimum Semantic Content*) aborda esta limitación desde una perspectiva distinta: utiliza imágenes con contenido semántico reducido o controlado para estudiar la estética visual minimizando, en la medida de lo posible, la influencia del reconocimiento de objetos o escenas [8, 9].

El concepto de *clutter* visual resulta relevante porque relaciona la cantidad y organización de información visual con la carga perceptual que una imagen puede imponer al observador. Rosenholtz et al. [12] proponen medidas de *clutter* orientadas a cuantificar la congestión de rasgos visuales y su impacto perceptual. En este trabajo no se replica directamente esa métrica porque el objetivo no es medir congestión visual a partir de los mismos rasgos o del mismo modelo perceptual, sino estudiar una representación más específica basada en categorías cromáticas. La conexión conceptual es, no obstante, útil: si la organización de rasgos visuales puede afectar a la percepción de una imagen, entonces la distribución y fragmentación espacial de sus categorías cromáticas también puede aportar información sobre su complejidad perceptual.

En este enfoque, una imagen no se describe mediante objetos semánticos, sino mediante un mapa de etiquetas cromáticas. Cada píxel o región queda asignado a una categoría de color, y las métricas se calculan sobre la distribución y organización espacial de esas etiquetas. Así, medidas como la entropía, la dominancia, la densidad de frontera o la fragmentación no indican cuántos objetos contiene una escena, sino cómo se distribuyen y se agrupan las categorías cromáticas dentro de la imagen. Esta distinción es importante porque permite analizar propiedades visuales de bajo nivel sin depender directamente del reconocimiento semántico de la escena.

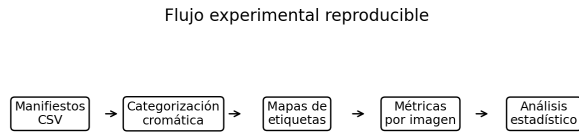


Fig. 1: Flujo experimental reproducible empleado en el análisis.

4 METODOLOGÍA

La metodología se organiza en cuatro niveles: preparación de datasets, categorización cromática, extracción de métricas y análisis estadístico. La Fig. 1 resume el flujo experimental.

La Tabla 1 resume los conjuntos de datos utilizados en el análisis experimental. Cada fila corresponde a un conjunto o subconjunto de imágenes procesado de forma independiente. La columna *Run / manifiesto* identifica el fichero o ejecución reproducible utilizada en el código; *N* indica el número final de imágenes analizadas tras el preprocesado; *Score* describe la puntuación estética empleada como variable de referencia; y la última columna resume el papel de cada conjunto dentro del diseño experimental.

4.1 Conjuntos de datos

El análisis experimental se realiza sobre tres conjuntos de datos de estética visual: AVA, AADB y MSC. En todos los casos, cada imagen tiene asociada una puntuación estética humana que se utiliza como variable de referencia. El objetivo no es clasificar las imágenes como “bonitas” o “feas”, sino estudiar si las métricas cromático-espaciales calculadas a partir de los mapas de categorías cromáticas muestran alguna relación cuantitativa con esas puntuaciones humanas.

AVA y AADB se utilizan como conjuntos fotográficos de referencia. Están formados por imágenes naturales o fotográficas, por lo que la valoración estética puede depender de muchos factores simultáneos: el contenido representado, la composición, la iluminación, el sujeto principal, la calidad técnica de la fotografía o las preferencias individuales de los evaluadores. Esta complejidad los convierte en conjuntos útiles para comprobar si las métricas propuestas conservan señal en escenarios realistas, pero también hace esperable que las correlaciones con descriptores de bajo nivel sean limitadas.

MSC desempeña un papel diferente dentro del análisis. A diferencia de AVA y AADB, está diseñado para estudiar imágenes con contenido semántico reducido o controlado. Esto permite disminuir, aunque no eliminar por completo, la influencia del reconocimiento de objetos o escenas en la valoración estética. Por este motivo, MSC resulta especialmente relevante para evaluar métricas cromático-espaciales: si una métrica mide propiedades de color, fragmentación o complejidad visual, es razonable esperar que su señal sea más observable cuando la valoración depende menos del significado de la imagen y más de su apariencia visual.

También es importante considerar el rango de puntuacio-

nes disponible en cada conjunto. Una correlación mide hasta qué punto dos variables varían conjuntamente. Si las puntuaciones estéticas están concentradas en un intervalo estrecho, la variable de referencia cambia poco y resulta más difícil detectar asociaciones con las métricas visuales, aunque estas contengan información relevante. Por este motivo se utilizaron subconjuntos balanceados en AVA y AADB, y se dio un papel central a MSC, que proporciona un rango de valoración más amplio y subconjuntos con transformaciones controladas.

Dentro de MSC se distinguen varias familias de imágenes. El subconjunto *unmodified* contiene imágenes no modificadas. Los subconjuntos *uglified* y *auto-uglified* contienen imágenes transformadas para reducir su valoración estética, manual o automáticamente según el caso. El subconjunto *beautified* contiene imágenes transformadas para aumentar su valoración estética. Esta división permite analizar si las métricas responden de forma diferente a imágenes originales y a imágenes modificadas con una intención estética concreta.

Durante la preparación de MSC se detectó una incidencia de cobertura en el procesamiento. El conjunto completo contenía 10.426 imágenes, pero la primera ejecución generó resultados para 8.954 imágenes. Las 1.472 imágenes no procesadas pertenecían precisamente a subconjuntos transformados: 827 del grupo *uglified* y 645 del grupo *beautified*. Esto era relevante porque esos subconjuntos son necesarios para estudiar cómo cambian las métricas cuando la imagen ha sido modificada para alterar su valoración estética.

Para corregir la incidencia se verificó que las imágenes faltantes existían en disco y que podían leerse correctamente como imágenes RGB. A partir de esa comprobación se generó un manifiesto corregido, denominado *msc-full-rgbfix*, que incluía las rutas adecuadas para relanzar el procesamiento de las imágenes pendientes. Tras esta corrección, cada método de categorización cromática produjo una fila de resultados por imagen, es decir, 10.426 filas por método. Cada fila contiene las métricas calculadas para una imagen concreta, junto con su puntuación estética asociada.

El sufijo *uglifier* identifica imágenes de MSC vinculadas a transformaciones de degradación estética. En la versión corregida se conservaron 1.791 imágenes con este identificador, lo que permitió mantener la trazabilidad entre imagen original, imagen transformada y puntuación humana. Esta corrección evita que el análisis de MSC quede sesgado por la ausencia de una parte importante de las imágenes transformadas.

A diferencia de AVA y AADB, que se utilizan como conjuntos fotográficos de referencia sin subdividirlos por tipo de transformación, MSC contiene varias familias de imágenes con significado experimental distinto. Por este motivo, además de indicar su tamaño total en la Tabla 1, se muestra por separado su composición interna en la Fig. 2. Las etiquetas de la figura no representan intervalos de puntuación, sino tipos de imagen dentro de MSC: *unmodified* corresponde a imágenes no modificadas, *auto-uglified* a imágenes degradadas automáticamente, *uglified* a imágenes degradadas para reducir su valoración estética y *beautified* a imágenes modificadas para aumentar su valoración estética.

TABLA 1: RESUMEN DE LOS CONJUNTOS DE DATOS Y SUBCONJUNTOS UTILIZADOS EN EL ANÁLISIS EXPERIMENTAL. CADA FILA REPRESENTA UN CONJUNTO DE IMÁGENES PROCESADO DE FORMA INDEPENDIENTE; LAS COLUMNAS INDICAN EL IDENTIFICADOR REPRODUCIBLE DEL EXPERIMENTO, EL NÚMERO FINAL DE IMÁGENES, LA PUNTUACIÓN ESTÉTICA UTILIZADA Y EL PAPEL DE ESE CONJUNTO EN LA COMPARACIÓN.

Conjunto	Run / manifiesto	N	Score	Papel en el análisis
AVA	ava_balanced_3_8_seed01	4400	Media AVA filtrada 3–8	Contraste con una base fotográfica amplia y rango de puntuaciones controlado.
AADB	aadb_balanced_0_1_0_9_seed01	2786	Score estético 0.1–0.9	Contraste con una base de estética fotográfica que incorpora atributos asociados a la imagen.
MSC completo	msc_full_rgbfix	10426	Media de 100 votos, escala 1–5	Conjunto principal para analizar imágenes con contenido semántico reducido o controlado.
MSC unmodified	msc_unmodified	5684	Media de votos	Subconjunto de imágenes no modificadas; permite una lectura más conservadora de MSC.
MSC auto_uglified	msc_auto_uglified	2951	Media de votos	Subconjunto de imágenes transformadas automáticamente hacia menor valoración estética.
MSC uglified	msc_uglified	872	Media de votos	Subconjunto de imágenes transformadas hacia menor valoración estética.
MSC beautified	msc_beautified	919	Media de votos	Subconjunto de imágenes transformadas hacia mayor valoración estética.

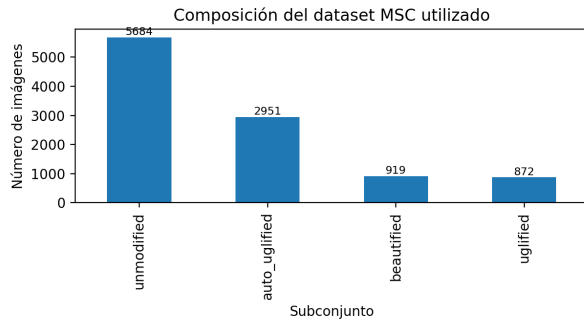


Fig. 2: Composición interna del conjunto MSC utilizado en los experimentos. Cada barra representa el número de imágenes de una familia del dataset: imágenes no modificadas (*unmodified*), imágenes degradadas automáticamente (*auto-uglified*), imágenes degradadas para reducir su valoración estética (*uglified*) e imágenes modificadas para aumentar su valoración estética (*beautified*). AVA y AADB no se incluyen en esta figura porque se utilizan como conjuntos fotográficos de referencia y no disponen de esta misma división por transformaciones.

4.2 Categorización cromática

La categorización cromática transforma cada imagen en un mapa de etiquetas de color. En lugar de trabajar directamente con los valores RGB originales, cada píxel o región se asigna a una categoría cromática discreta. Esta representación conserva la organización espacial de la imagen, pero sustituye los valores continuos de color por etiquetas interpretables. Sobre esos mapas se calculan posteriormente las métricas de diversidad, dominancia, fragmentación, densidad de frontera y complejidad multiescala descritas en la Sección 4.3.

El punto de partida metodológico es el modelo NICE, que propone una solución computacional para relacionar percepción y categorización del color [7]. En este trabajo no se evalúa únicamente una implementación, sino varias variantes de categorización cromática. La razón es que las métricas posteriores dependen directamente del mapa de etiquetas generado: si un método agrupa los colores de forma distinta, también cambiarán la entropía, la fragmentación o la densidad de fronteras calculadas sobre ese mapa.

Se comparan cuatro variantes experimentales. La varian-

te *ourlab1000* se utiliza como referencia propia principal y aplica la categorización cromática sobre una representación basada en el espacio Lab, con una parametrización ajustada de las regiones de categoría. La variante *ourlsy* mantiene la misma lógica general de asignación a categorías, pero cambia el espacio cromático utilizado, lo que permite evaluar si la representación del color afecta a la señal estética observada. Las variantes *joost* y *robert* se incluyen como comparadores externos disponibles en el entorno experimental. Su función es comprobar si los resultados dependen de una única implementación o si se mantienen, al menos parcialmente, al utilizar mapas de etiquetas generados por variantes alternativas.

Todas las variantes siguen la misma estructura experimental: reciben una imagen como entrada, generan un mapa de categorías cromáticas como salida y permiten calcular el mismo conjunto de métricas sobre dicho mapa. De este modo, las diferencias observadas en los resultados pueden atribuirse al efecto de la categorización cromática utilizada, manteniendo constante el protocolo de extracción de métricas y análisis estadístico. La descripción técnica ampliada de cada variante se incluye en la Sección A.1 del apéndice técnico.

4.3 Métricas extraídas

Una vez obtenido el mapa de etiquetas cromáticas de cada imagen, se calculan métricas que resumen diferentes propiedades de esa representación. Estas métricas no se aplican directamente sobre los valores RGB originales, sino sobre la distribución y organización espacial de las categorías cromáticas asignadas por cada algoritmo.

Las métricas se agrupan en tres tipos. En primer lugar, las métricas de diversidad y distribución describen cuántas categorías cromáticas aparecen y cómo se reparte la imagen entre ellas. En segundo lugar, las métricas espaciales describen cómo se organizan esas categorías en la imagen, por ejemplo si aparecen agrupadas o fragmentadas. Finalmente, las métricas multiescala evalúan si esa organización cambia al analizar la imagen con diferentes tamaños de bloque.

num-categories. Número de categorías cromáticas presentes en el mapa de etiquetas. Se interpreta como una medida discreta de diversidad cromática.

entropy. Entropía de Shannon de la distribución de etiquetas cromáticas. Resume el grado de dispersión de la distribución: valores altos indican que la imagen reparte su superficie entre varias categorías, mientras que valores bajos indican mayor concentración en pocas categorías.

effective-categories. Número efectivo de categorías, calculado a partir de la entropía. Permite expresar la diversidad cromática en una escala más interpretable, equivalente al número de categorías que contribuirían de forma equilibrada.

dominant-category-fraction. Proporción de la imagen ocupada por la categoría cromática más frecuente. Se interpreta como una medida de concentración o dominancia: valores altos indican que una categoría cromática ocupa gran parte de la imagen.

fragmentation. Promedio de diversidad local de etiquetas en bloques espaciales. Describe hasta qué punto las categorías cromáticas aparecen mezcladas localmente en la imagen.

boundary-density. Densidad de cambios entre etiquetas cromáticas vecinas. Se interpreta como una medida de complejidad de frontera: valores altos indican mayor cantidad de transiciones entre regiones cromáticas.

multiscale-entropy. Variación de la entropía al cambiar la escala de muestreo. Permite evaluar si la distribución cromática es estable o cambia al observar la imagen con distintos niveles de detalle.

multiscale-fragmentation. Variación de la fragmentación al utilizar distintos tamaños de bloque. Mide la sensibilidad de la organización espacial de las categorías cromáticas a la escala de análisis.

MCCI. Índice compuesto de complejidad cromático-espacial. Combina diversidad, fragmentación, variación multiescala, densidad de frontera y penalización por dominancia en un único descriptor reproducible.

4.4 Métrica compuesta MCCI

Además de las métricas individuales, se define un índice compuesto denominado MCCI. En este trabajo, MCCI se utiliza como abreviatura operativa de *Multiscale Chromatic Complexity Index*, es decir, un índice de complejidad cromática multiescala. Su objetivo es resumir en un único descriptor varias propiedades complementarias del mapa de etiquetas cromáticas: diversidad global, fragmentación espacial, variación multiescala, densidad de frontera y dominancia de la categoría principal.

El índice se calcula como una combinación ponderada de componentes normalizados:

$$\text{MCCI} = 0,25e + 0,30f + 0,20m + 0,20b + 0,05(1-d), \quad (1)$$

donde e representa la entropía normalizada, f la fragmentación normalizada, m la variación multiescala normalizada, b la densidad de frontera normalizada y d la fracción ocupada por la categoría cromática dominante. El término $(1-d)$ actúa como penalización inversa de la dominancia: aumenta cuando la imagen no está concentrada en una única categoría cromática.

Los pesos se fijaron de forma heurística antes del análisis estadístico, no mediante entrenamiento ni optimización sobre las puntuaciones estéticas humanas. Se dio mayor peso a la fragmentación (0,30) y a la entropía (0,25) porque son los componentes más directamente relacionados con la idea de complejidad cromático-espacial: cuántas categorías aparecen y cómo se distribuyen localmente. La variación multiescala (0,20) y la densidad de frontera (0,20) reciben pesos intermedios porque aportan información espacial complementaria. La penalización por dominancia (0,05) se incluye con menor peso para evitar que el índice dependa en exceso de imágenes dominadas por una sola categoría.

Por tanto, MCCI debe interpretarse como un descriptor compuesto exploratorio y reproducible, no como una formulación teórica cerrada ni como un modelo predictivo entrenado. Su utilidad se evalúa comparando su relación con las puntuaciones estéticas frente a la de las métricas individuales.

4.5 Protocolo estadístico

Para cada combinación de conjunto de datos, método de categorización cromática y métrica visual, se analiza la relación entre la métrica calculada y la puntuación estética humana asociada a cada imagen. A lo largo del trabajo se utiliza el término *puntuación estética* para referirse a esta valoración humana de referencia, independientemente de que cada conjunto de datos la almacene en una escala distinta.

El análisis principal se basa en correlaciones de Spearman y Pearson. La correlación de Spearman se utiliza como medida principal porque evalúa relaciones monotónicas y no exige que la asociación entre una métrica y la puntuación estética sea estrictamente lineal. Esto resulta adecuado para métricas discretas o con distribuciones no gaussianas, como el número de categorías cromáticas. La correlación de Pearson se calcula de forma complementaria porque mide la asociación lineal entre ambas variables y permite comprobar si una relación monotónica observada también se aproxima a una tendencia lineal.

Junto con cada correlación se reporta su valor p . Este valor indica la probabilidad de obtener una asociación igual o más extrema que la observada bajo la hipótesis nula de ausencia de relación. Sin embargo, los valores p se interpretan con cautela: en muestras grandes, asociaciones de magnitud pequeña pueden resultar estadísticamente significativas. Por este motivo, la interpretación se basa principalmente en el tamaño del efecto, es decir, en la magnitud de los coeficientes de correlación, y no únicamente en la significación estadística [13].

En los análisis extendidos se calculan intervalos de confianza mediante *bootstrapping* y se aplica corrección por comparaciones múltiples mediante FDR (*False Discovery Rate*) [3, 1]. Los intervalos de confianza permiten estimar la estabilidad de las correlaciones observadas, mientras que

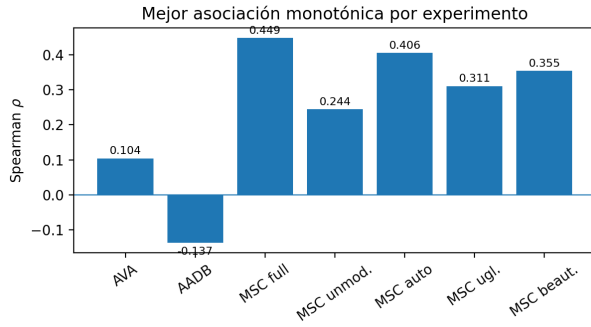


Fig. 3: Mejor asociación monotónica observada en cada experimento, medida mediante el coeficiente de Spearman ρ . El eje horizontal identifica el conjunto o subconjunto analizado: AVA balanceado, AADB balanceado, MSC completo, MSC *unmodified*, MSC *auto-uglified*, MSC *uglified* y MSC *beautified*. El eje vertical muestra el valor de ρ correspondiente a la combinación método-métrica con mayor $|\rho|$ en cada caso. Todas las correlaciones representadas son estadísticamente significativas con $p < 0,05$; los valores exactos de p , los intervalos de confianza y las correlaciones de Pearson se detallan en la Tabla 2.

la corrección FDR reduce el riesgo de interpretar como relevantes asociaciones que podrían aparecer por azar al evaluar muchas métricas, métodos y subconjuntos.

El análisis se realiza sobre el conjunto completo de cada experimento, denominado *full*, y también sobre subconjuntos definidos por percentiles de puntuación estética. El subconjunto *top30* contiene el 30 % de imágenes con mayor puntuación estética; *bottom30* contiene el 30 % con menor puntuación; y *mid40*, cuando se utiliza, contiene el 40 % central. Esta división permite comparar el comportamiento de las métricas en imágenes mejor valoradas, peor valoradas y de valoración intermedia. Se eligió el umbral del 30 % como compromiso entre contraste estético y tamaño muestral: porcentajes más extremos, como 10 % o 20 %, producirían grupos más diferenciados, pero también más pequeños y menos estables, especialmente en subconjuntos reducidos como *uglified* o *beautified*.

Además de las correlaciones, se ajustan modelos cuadráticos simples para explorar posibles relaciones no lineales entre las métricas y la puntuación estética. Este análisis permite detectar patrones de U o U invertida, es decir, casos en los que los valores intermedios de una métrica se asocian con puntuaciones distintas a las observadas en los extremos. Esta exploración es coherente con trabajos que relacionan la experiencia estética con propiedades estadísticas, complejidad visual y regularidades perceptuales de la imagen [10, 11]. En este contexto, los modelos cuadráticos no se utilizan como predictores finales, sino como una herramienta descriptiva para comprobar si añadir un término no lineal mejora la explicación respecto a un ajuste lineal simple.

5 RESULTADOS

La Tabla 2 muestra la mejor asociación monotónica observada en cada experimento. La Fig. 3 resume visualmente estas correlaciones.

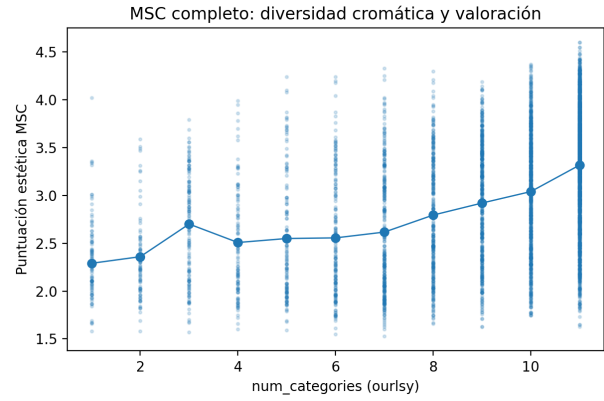


Fig. 4: Relación entre el número de categorías cromáticas y la puntuación estética en MSC completo para la variante *ourlsy*. La línea resume la tendencia media por intervalos de la métrica.

5.1 AVA y AADB como contrastes

Los resultados en AVA y AADB muestran asociaciones de baja magnitud en comparación con MSC. En AVA balanceado, la mayor asociación monotónica se observa con la combinación *ourlsy / num-categories* en el conjunto completo, con $\rho = 0,104$ y $p < 10^{-3}$. Aunque la asociación es estadísticamente significativa, el tamaño del efecto es reducido.

En AADB balanceado, la mayor asociación en valor absoluto aparece en el subconjunto *top30* con la combinación *ourlsy / effective-categories*, con $\rho = -0,137$ y $p < 10^{-3}$. El signo negativo indica una asociación monotónica inversa en ese subconjunto concreto. No obstante, la magnitud del efecto sigue siendo baja, por lo que este resultado debe interpretarse con prudencia. Los valores exactos de p , junto con los intervalos de confianza y las correlaciones de Pearson, se muestran en la Tabla 2.

5.2 MSC completo

MSC completo muestra asociaciones más altas que AVA y AADB. La mayor asociación corresponde a la combinación *ourlsy / num-categories*, con Spearman rho igual a 0.449, Pearson r igual a 0.433 y valor p menor que 0.001. También destacan *ourlab1000 / entropy* y *ourlab1000 / effective-categories*, ambas con Spearman rho igual a 0.417. La fracción dominante aparece con signo negativo en varias variantes, lo que indica que una mayor concentración de la imagen en una sola categoría cromática tiende a asociarse con puntuaciones estéticas más bajas.

La Fig. 4 ilustra el caso de *ourlsy / num-categories*. La nube de puntos muestra una dispersión amplia, como es habitual en datos perceptuales, pero la tendencia agregada indica una asociación positiva entre el número de categorías cromáticas y la puntuación estética en MSC completo. Los valores exactos de correlación y significación estadística se resumen en la Tabla 2.

5.3 Subconjuntos de MSC

La división de MSC por familias de imagen permite observar diferencias entre imágenes no modificadas y trans-

TABLA 2: MEJOR ASOCIACIÓN MONOTÓNICA OBSERVADA EN CADA EXPERIMENTO SEGÚN $|\rho|$ DE SPEARMAN. SE INCLUYEN VALORES p Y, CUANDO ESTABAN DISPONIBLES, INTERVALOS DE CONFIANZA BOOTSTRAP PARA ρ .

Experimento	Método / métrica	Subset	N	ρ	p_ρ	IC 95 % ρ	r	p_r
AVA balanceado	ourlsy / num_categories	full	4400	0.104	4.3e-12	[0.075, 0.133]	0.079	1.4e-07
AADB balanceado	ourlsy / effective_categories	top30	844	-0.137	6.6e-05	[-0.202, -0.074]	-0.140	4.7e-05
MSC completo	ourlsy / num_categories	full	10426	0.449	$< 10^{-300}$	[0.433, 0.465]	0.433	$< 10^{-300}$
MSC unmodified	joost / entropy	full	5684	0.244	5.0e-78	[0.221, 0.270]	0.240	4.7e-75
MSC auto_uglified	ourlsy / num_categories	full	2951	0.406	2.7e-117	[0.376, 0.435]	0.384	5.0e-104
MSC uglified	ourlab1000 / effective_categories	full	872	0.311	4.9e-21	[0.248, 0.372]	0.290	2.4e-18
MSC beautified	ourlsy / num_categories	bottom30	280	0.355	9.9e-10	[0.250, 0.454]	0.387	1.9e-11

formadas. En el subconjunto *unmodified*, la mayor asociación se obtiene con *joost / entropy*, con $\rho = 0,244$. En *auto-uglified*, *ourlsy / num-categories* alcanza $\rho = 0,406$, mientras que *ourlsy / multiscale-fragmentation* alcanza $\rho = 0,387$. En *uglified*, la mayor asociación corresponde a *ourlab1000 / effective-categories*, con $\rho = 0,311$. Finalmente, en *beautified*, el valor máximo aparece en el subconjunto *bottom30*, con *ourlsy / num-categories* y $\rho = 0,355$.

Estos resultados muestran que las asociaciones no son homogéneas en todas las familias de MSC. Las correlaciones son más altas en algunos subconjuntos transformados que en *unmodified*, donde la magnitud de la asociación disminuye. Los valores exactos de correlación, significación estadística y tamaño muestral se recogen en la Tabla 2.

5.4 Relaciones no lineales

Además de las correlaciones de Spearman y Pearson, se ajustaron modelos de regresión simples para explorar si algunas métricas presentan relaciones no lineales con la puntuación estética. En concreto, se comparó un modelo lineal, que solo permite una relación creciente o decreciente, con un modelo cuadrático, que permite detectar patrones de U o U invertida. Este análisis no sustituye a las correlaciones: responde a una pregunta diferente. Mientras que Spearman mide asociación monotónica, el coeficiente R^2 mide la proporción de variabilidad de la puntuación estética explicada por un modelo de regresión.

La Tabla 3 y la Fig. 5 resumen la mejora obtenida al añadir el término cuadrático. El valor ΔR^2 representa la diferencia entre el R^2 del modelo cuadrático y el R^2 del modelo lineal. Por tanto, valores positivos de ΔR^2 indican que el modelo cuadrático describe mejor la relación entre la métrica y la puntuación estética que una recta simple.

El patrón más claro aparece en MSC completo y en MSC *auto-uglified*. En MSC completo, la combinación *joost / boundary-density* alcanza $\Delta R^2 = 0,076$ con forma de U invertida. En MSC *auto-uglified*, la misma métrica alcanza $\Delta R^2 = 0,071$. En AVA y AADB, las mejoras máximas son mucho menores, alrededor de 0,008–0,009. La Fig. 6 muestra un ejemplo de ajuste cuadrático en MSC completo.

6 DISCUSIÓN

Los resultados sostienen una conclusión intermedia. Las métricas cromático-espaciales no explican de forma uniforme la estética visual en todos los conjuntos de datos, pero sí capturan variaciones perceptuales que se hacen más visibles cuando el conjunto contiene suficiente rango de puntuaciones y menor carga semántica. Por tanto, el conjunto de datos no es un simple soporte experimental, sino una parte activa de la interpretación.

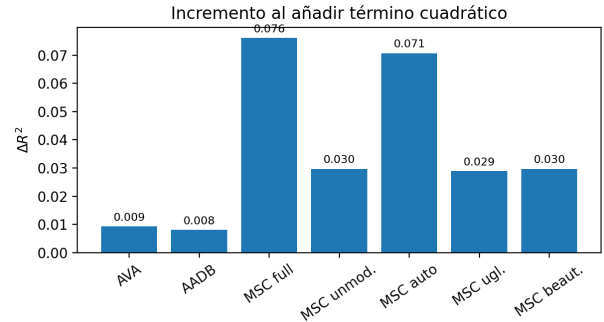


Fig. 5: Mejora máxima observada al añadir un término cuadrático en cada experimento.

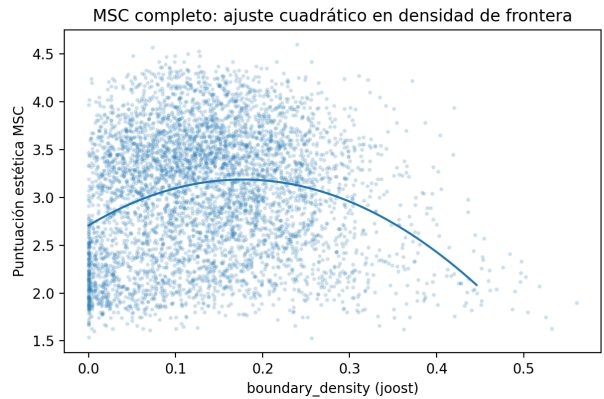


Fig. 6: Ejemplo de ajuste cuadrático entre la densidad de frontera y la valoración estética en MSC completo para la variante *joost*.

Los resultados de AVA y AADB son útiles como contraste frente a MSC. El hecho de que las métricas cromático-espaciales no produzcan correlaciones altas en todos los conjuntos indica que la señal observada no depende únicamente de la definición matemática de las métricas, sino también de la composición del conjunto de datos, del rango de puntuaciones estéticas y del peso relativo del contenido semántico. En AVA y AADB, el contenido representado, la composición fotográfica y la heterogeneidad de las imágenes pueden diluir el papel de propiedades visuales de bajo nivel. En MSC, en cambio, la reducción parcial de la carga semántica y la presencia de transformaciones que modifican la apariencia visual hacen que la señal cromático-espacial sea más detectable.

En particular, el resultado de *ourlsy / num-categories* en MSC completo no debe interpretarse como que “más categorías cromáticas siempre equivalen a mayor belleza”. La lectura más prudente es que, en este conjunto de datos, la di-

TABLA 3: MAYOR MEJORA OBSERVADA AL COMPARAR UN AJUSTE LINEAL SIMPLE CON UN AJUSTE CUADRÁTICO.

Experimento	Método / métrica	Subset	N	R^2 lin.	R^2 cuad.	ΔR^2	Forma
AVA balanceado	ourlab1000 / entropy	full	4400	0.001	0.010	0.009	U invertida
AADB balanceado	ourlab1000 / effective_categories	bottom30	868	0.004	0.012	0.008	U invertida
MSC completo	joost / boundary_density	full	10426	0.001	0.077	0.076	U invertida
MSC unmodified	ourlsy / num_categories	bottom30	1732	0.005	0.035	0.030	U
MSC auto_uglified	joost / boundary_density	full	2951	0.022	0.093	0.071	U invertida
MSC uglified	joost / boundary_density	full	872	0.001	0.029	0.029	U invertida
MSC beautified	robert / multiscale_entropy	bottom30	280	0.008	0.037	0.030	U invertida

versidad cromática capturada por el algoritmo se asocia de forma clara con la puntuación estética humana. Esta asociación puede depender de la composición específica de MSC y de la presencia de imágenes transformadas, por lo que debe interpretarse junto con los análisis por subconjuntos.

El análisis por subconjuntos refuerza esta cautela. Parte de la correlación global de MSC procede de analizar conjuntamente imágenes originales e imágenes transformadas. Al aislar el subconjunto *unmodified*, la asociación disminuye, aunque no desaparece. Esto sugiere que las métricas cromático-espaciales son especialmente sensibles a cambios inducidos en la organización cromática y espacial, mientras que su relación con imágenes no modificadas es más moderada.

Esta observación es coherente con el planteamiento de MSC y con la advertencia de que las bases de datos de estética pueden estar sesgadas hacia determinados rangos de valoración o tipos de imagen [8, 9].

La segunda aportación es metodológica. El flujo implementado permite comparar de forma sistemática cuatro variantes de categorización cromática. Los resultados no seleccionan un ganador absoluto: *ourlsy* domina algunas correlaciones globales, *ourlab1000* obtiene buenos resultados en entropía y subconjuntos transformados, y *joost* aparece con fuerza en relaciones cuadráticas de densidad de frontera. Esta diversidad de resultados sugiere que la elección del espacio cromático y de las fronteras categoriales afecta a la señal estética observada.

La tercera aportación es la lectura no lineal. Métricas como la densidad de frontera o MCCI no son siempre las mejores en correlación monotónica, pero sí muestran mejoras cuadráticas relevantes. Esto indica que algunas relaciones entre complejidad cromático-espacial y puntuación estética no se describen bien mediante una tendencia estrictamente creciente o decreciente.

Una interpretación posible es que ciertas métricas de frontera o complejidad compuesta capturan mejor niveles intermedios de organización visual. Esta idea conecta con la noción de *clutter* visual: una imagen con poca estructura puede resultar pobre desde el punto de vista visual, mientras que una fragmentación excesiva también puede penalizar la preferencia. En consecuencia, la preferencia estética podría depender no solo de aumentar la diversidad cromática, sino de mantener un equilibrio entre variedad y organización espacial. Esta hipótesis debe interpretarse de forma prudente y reforzarse con ejemplos visuales extremos y análisis cualitativos complementarios.

7 CONCLUSIONES Y TRABAJO FUTURO

Los resultados obtenidos permiten extraer las siguientes conclusiones principales:

- Las métricas derivadas de mapas de categorización cromática muestran una relación limitada con la puntuación estética en AVA y AADB. Esto indica que, en conjuntos fotográficos complejos y semánticamente ricos, los descriptores cromático-espaciales de bajo nivel explican solo una parte reducida de la preferencia estética visual.
- MSC ofrece una señal más clara que AVA y AADB. En el conjunto completo, la combinación *ourlsy* / *num-categories* alcanza una correlación de Spearman de 0.449, y varias métricas relacionadas con entropía y diversidad efectiva presentan asociaciones superiores a 0.39. Este resultado sugiere que las propiedades cromático-espaciales son más informativas cuando se reduce parcialmente la influencia del contenido semántico.
- La señal observada en MSC depende de la composición del conjunto. Al separar imágenes no modificadas de imágenes transformadas, la magnitud de las correlaciones cambia de forma apreciable. Por tanto, los resultados globales de MSC deben interpretarse junto con el análisis por subconjuntos.
- Las métricas de densidad de frontera y complejidad compuesta muestran relaciones no lineales relevantes, especialmente en MSC completo y en MSC *auto-uglified*. Esto indica que algunas propiedades cromático-espaciales pueden relacionarse con la puntuación estética mediante patrones de U o U invertida, y no solo mediante relaciones crecientes o decrecientes.
- MCCI se ha implementado como un descriptor compuesto y reproducible de complejidad cromático-espacial. Su utilidad debe interpretarse de forma exploratoria: no sustituye a las métricas individuales, pero permite resumir varias dimensiones del mapa de etiquetas cromáticas en un único índice comparable entre métodos y conjuntos de datos.

Como trabajo futuro, una primera línea de continuación consiste en ampliar el análisis visual cualitativo de casos extremos, comparando imágenes con valores altos y bajos de cada métrica para comprobar si las diferencias numéricas se corresponden con diferencias perceptualmente interpretables. También sería útil estudiar variantes alternativas de MCCI, incluyendo otros pesos o estrategias de normalización, y evaluar si el índice mantiene su comportamiento en nuevos conjuntos de datos. Finalmente, el enfoque podría extenderse combinando las métricas cromático-espaciales con otros descriptores de composición, textura o saliencia, con el objetivo de analizar cómo interactúan distintos factores de bajo nivel en la preferencia estética visual.

USO DE LA IA GENERATIVA

Durante la preparación de este documento se han utilizado herramientas de IA generativa como apoyo para reorganizar el material, revisar la estructura académica del documento, transformar salidas experimentales en texto explicativo y detectar puntos débiles de presentación. La ejecución de código, generación de manifiestos, validación de rutas, procesamiento de datasets y obtención de resultados numéricos procede de los scripts y ficheros del proyecto. Las cifras incluidas en tablas y figuras se han tomado de `features_final.csv`, `correlations_summary.csv`, `correlations_extended.csv` y `quadratic_models.csv`. La IA no se ha usado como fuente de resultados experimentales ni sustituye la verificación de las salidas.

AGRADECIMIENTOS

Agradezco al tutor, C. Alejandro Parraga, el seguimiento del trabajo, las indicaciones metodológicas y los comentarios sobre la organización académica del informe. También se agradece el acceso a referencias y materiales de investigación que han permitido orientar el TFG hacia una evaluación más rigurosa de la relación entre propiedades visuales de bajo nivel y preferencia estética.

REFERENCIAS

- [1] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, pp. 289–300, 1995.
- [2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *European Conference on Computer Vision*, vol. 3953, pp. 288–301, 2006.
- [3] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994.
- [4] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo, "Aesthetics and emotions in images: A computational perspective," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 94–115, 2011.
- [5] S. Kong, X. Shen, Z. Lin, R. Mech, and C. Fowlkes, "Photo aesthetics ranking network with attributes and content adaptation," in *European Conference on Computer Vision*, 2016.
- [6] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2408–2415, 2012.
- [7] C. A. Parraga and A. Akbarinia, "NICE: A computational solution to close the gap from colour perception to colour categorization," *PLOS ONE*, vol. 11, e0149538, 2016. doi: 10.1371/journal.pone.0149538.
- [8] C. A. Parraga, O. Penacchio, M. Muñoz Gonzalez, B. Raducanu, and X. Otazu, "Aesthetics Without Semantics," arXiv:2505.05331, 2025.
- [9] O. Penacchio, A. Javed, B. Raducanu, X. Otazu, and C. A. Parraga, "The Minimum Semantic Content (MSC) Dataset: A Large, Balanced Resource for Computational Aesthetics Research," *Scientific Data*, vol. 13, article 470, 2026. doi: 10.1038/s41597-026-06816-0.
- [10] C. Redies, "Combining universal beauty and cultural context in a unifying model of visual aesthetic experience," *Frontiers in Human Neuroscience*, vol. 9, article 218, 2015.
- [11] C. Redies, R. Bartho, L. Koßmann, B. Spehar, R. Hübner, J. Wagemans, and G. U. Hayn-Leichsenring, "A toolbox for calculating quantitative image properties in aesthetics research," *Behavior Research Methods*, vol. 57, article 117, 2025. doi: 10.3758/s13428-025-02632-3.
- [12] R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *Journal of Vision*, vol. 7, no. 2, article 17, 2007.
- [13] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.

A APÉNDICE TÉCNICO

A.1 Variantes de categorización cromática

En esta sección se describen con más detalle las cuatro variantes de categorización cromática utilizadas en el análisis experimental. Todas ellas tienen la misma función dentro del flujo de trabajo: reciben una imagen como entrada y producen un mapa de etiquetas cromáticas como salida. Sobre esos mapas se calculan posteriormente las métricas de diversidad, dominancia, fragmentación, densidad de frontera y complejidad multiescala. La diferencia entre variantes está en cómo representan el color y cómo asignan cada píxel o región a una categoría cromática.

ourlab1000. Esta variante se utiliza como referencia principal del trabajo. Parte de una representación del color basada en el espacio Lab y aplica una parametrización ajustada de las regiones de categoría. Su papel es proporcionar una implementación controlada y reproducible sobre la que evaluar la hipótesis central: si la distribución espacial de categorías cromáticas permite obtener descriptores relacionados con la valoración estética humana.

ourlsy. Esta variante mantiene la lógica general de categorización, pero cambia el espacio cromático utilizado antes de asignar las categorías. Su inclusión permite estudiar si la señal obtenida depende del espacio de representación del color. Si dos espacios cromáticos producen mapas de etiquetas distintos, también pueden cambiar métricas como la entropía, la fragmentación o la densidad de frontera.

joost. Esta variante se incluye como comparador externo dentro del entorno experimental. Genera mapas de categorías cromáticas mediante una definición alternativa de las asignaciones de color. Su función no es sustituir a la variante principal, sino comprobar si los patrones observados aparecen únicamente con la implementación propia o si se mantienen parcialmente al utilizar otro procedimiento de categorización.

robert. Esta variante actúa como segundo comparador externo. Al igual que *joost*, produce mapas de etiquetas cromáticas compatibles con el mismo módulo de extracción de métricas. Su inclusión permite analizar la estabilidad de los resultados ante cambios en el método de categorización y evita que las conclusiones dependan de una única implementación.

En conjunto, estas cuatro variantes permiten separar dos niveles del análisis: por un lado, la generación del mapa de categorías cromáticas; por otro, el cálculo de métricas sobre ese mapa. Mantener constante el segundo nivel y variar el primero permite estudiar hasta qué punto los resultados dependen del método de categorización utilizado.

A.2 Trazabilidad de resultados

Los resultados principales se apoyan en los ficheros generados automáticamente dentro de `results/`. Para cada run y método se conserva `features_final.csv`. Las correlaciones principales se guardan en `correlations_summary.csv`; las salidas con intervalos de confianza y corrección FDR se guardan en `nonlinear_analysis/correlations_extended.csv`; y los modelos cuadráticos se guardan en `nonlinear_analysis/quadratic_models.csv`. Esta organización permite reconstruir las tablas de la memoria y distinguir resultados seleccionados de material exploratorio.