CIENCIA INTERDISCIPLINARIA INTERNACIONAL

ISSN: 3121-2948

Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

DOI: https://doi.org/10.70577/cieninter.v4i4.162

Descriptor multivista de activaciones neuronales, mediante la fusión de grafos funcionales y homología persistente

Multiview descriptor of neural activations, through the fusion of functional graphs and persistent homology

César Daniel Reinoso Reinoso[1]
Departamento de Posgrados, Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo (ESPOCH)
Riobamba, Ecuador
cesard.reinoso@espoch.edu.ec
https://orcid.org/0000-0002-1233-517X

Amalia Isabel Escudero Villa[2]
Facultad de Ciencias, Escuela Superior Politécnica de Chimborazo (ESPOCH)
Riobamba, Ecuador
aescudero@espoch.edu.ec
https://orcid.org/0000-0003-4506-932X

Cómo citar

Reinoso Reinoso, C. D., & Escudero Villa, A. I. (2026). Descriptor multivista de activaciones neuronales, mediante la fusión de grafos funcionales y homología persistente. *Ciencia Interdisciplinaria Internacional*, *4*(4), 234–252. https://doi.org/10.70577/cieninter.v4i4.162

Fecha de recepción: 2026-10-15

Fecha de aceptación: 2026-10-29

Fecha de publicación: 2026-11-11

CIENCIA INTERDISCIPLINARIA INTERNACIONAL

ISSN: 3121-2948

Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

RESUMEN

La presente investigación propone un descriptor multivista de las activaciones internas de las redes neuronales, es decir, un conjunto fijo de números que resume una red y combina dos descripciones distintas del mismo objeto, una basada en métricas de teoría de grafos sobre la conectividad entre neuronas y otra en la homología persistente, que resume la forma de la nube de activaciones. Para su validación se emplearon cuarenta redes neuronales de dos tipos, perceptrones multicapa y convolucionales, distribuidas en cinco grupos y entrenadas en tres conjuntos de imágenes. La variable independiente es el grupo de cada red, y las dependientes son los números que el descriptor calcula para ella (diez medidas del grafo, doce estadísticos de persistencia y su fusión), sin usar la exactitud. La fusión proporcionó resultados altamente significativos en cuanto a funcionamiento y predicción, coincidencia y rendimiento, mediante validación cruzada (0,949, 0,957 y 0,972 en los tres conjuntos), contrastes no paramétricos (29 de 30 pruebas significativas) y el análisis ROC (0,949 con la fusión, 0,854 y 0,869 con cada vista y 0,754 con una línea base), respectivamente. Como los dos tipos de red difieren mucho en exactitud, dos comprobaciones adicionales, una restringida a los perceptrones multicapa y otra que igualó la exactitud dentro de cada tipo, mostraron que la ventaja del descriptor no depende de esa diferencia.

Palabras clave: Análisis topológico de datos, Aprendizaje profundo, Aprendizaje multivista, Homología persistente, Interpretabilidad.

ABSTRACT

This study proposes a multiview descriptor of the internal activations of neural networks: a fixed-length set of numbers that summarizes a network by combining two different descriptions of the same object, one drawn from graph-theoretic measures of the connectivity among neurons and the other from persistent homology, which captures the shape of the activation cloud. Its validation used forty neural networks of two types, multilayer perceptrons and convolutional networks, arranged in five groups and trained on three image datasets. The independent variable is the group of each network; the dependent variables are the numbers the descriptor computes for it (ten graph measures, twelve persistence statistics and their fusion), none of which uses accuracy. The fusion yielded highly significant results for operation and prediction, agreement, and performance, as assessed by cross-validation (0.949, 0.957 and 0.972 on the three datasets), non-parametric tests (29 of 30 tests significant) and ROC analysis (0.949 with the fusion, 0.854 and 0.869 with each view and 0.754 with a baseline), respectively. Because the two network types differ widely in accuracy, two further checks, one restricted to the multilayer perceptrons and one that equalized accuracy within each type, showed that the descriptor's advantage does not rest on that difference.

Keywords: Deep learning, Interpretability, Multi-view learning, Persistent homology, Topological data analysis.

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

INTRODUCCIÓN

Las redes neuronales artificiales sostienen aplicaciones tan diversas como el diagnóstico por imagen, la conducción autónoma y el reconocimiento de objetos. Su calidad suele juzgarse con una sola cifra, la exactitud, entendida como la proporción de ejemplos nuevos que el sistema clasifica correctamente. Esa cifra, sin embargo, no revela cómo funciona la red por dentro, ya que dos redes con exactitudes similares pueden comportarse como sistemas internamente distintos (Recht et al., 2019; Raghu et al., 2017; Klabunde et al., 2025). Este desconocimiento del funcionamiento interno dificulta diseñar, comparar y auditar modelos, por lo que, comprender la organización interna de una red entrenada, y no solo su desempeño final, se ha convertido en una necesidad central del aprendizaje profundo.

Para comprender esa organización, el interior de una red se estudia a través de sus activaciones, es decir, los valores que cada neurona produce ante una entrada, y para examinarlas se han propuesto herramientas complementarias. Entre ellas, destacan los mapas de saliencia (*saliency maps*), imágenes de calor con las cuales Simonyan et al. (2013) obtuvieron que las zonas de mayor peso en la decisión coinciden con el objeto reconocido, hallazgo extendido a cualquier arquitectura por Selvaraju et al. (2017). A ellas se suman los clasificadores de sondeo, modelos sencillos ajustados sobre las activaciones de una capa con los que Alain y Bengio (2016) demostraron que la información útil para clasificar crece hacia las capas finales. De manera análoga, mediante el análisis de similitud representacional, Raghu et al. (2017) comprobaron que las capas cercanas a la entrada culminan su aprendizaje antes que las profundas. Sin embargo, estas técnicas captan de forma limitada la dependencia mutua entre neuronas, de modo que la organización colectiva de la red queda descrita solo parcialmente.

Esa organización colectiva sí ha sido abordada desde la neurociencia, donde la actividad conjunta del cerebro se representa como un grafo funcional, una red en la que cada unidad constituye un nodo y dos unidades se conectan cuando se activan a la vez. Con tal representación, Bullmore y Sporns (2009) obtuvieron que el cerebro se organiza en módulos comunicados por nodos concentradores, principio que resultaba invisible al estudiar cada unidad por separado. Esta idea fue trasladada a las redes artificiales por Zhang et al. (2023), quienes conectaron las neuronas que se activan de forma conjunta y obtuvieron que las medidas del grafo resumen la organización interna del modelo; esta estrategia ya se había empleado en otros tipos de redes neuronales (Horta et al., 2021; Corneanu et al., 2020).Junto al grafo, una segunda vía trata las activaciones como una nube de puntos cuya forma global se describe con la homología persistente, herramienta de la topología que cuenta los grupos y huecos de la nube a distintas escalas (Ballester et al., 2026). Con ella, Naitzat et al. (2020) obtuvieron que las redes simplifican la forma de los datos capa tras capa hasta volverlos separables, mientras que Moor et al. (2020) lograron representaciones comprimidas que conservan la forma original. En el antecedente más próximo a este trabajo, Ballester et al. (2024) obtuvieron que los resúmenes de persistencia (estadísticos de cuánto perduran esos grupos y huecos) de grafos de correlación entre neuronas predicen la brecha de generalización, es decir, la diferencia de desempeño entre entrenamiento y prueba. Ambos enfoques proporcionan dos vistas complementarias, es decir, dos representaciones distintas de un mismo objeto; integrarlas es

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

prometedor, ya que el aprendizaje multivista ha demostrado que combinarlas supera al uso de una sola, como al unir audio y video para reconocer el habla (Baltrusaitis et al., 2019; Sun, 2013).

Con base en estos antecedentes, los propios autores señalan problemas pendientes. Demsar (2006) advirtió que los métodos de aprendizaje suelen compararse sin pruebas estadísticas adecuadas y propuso pruebas no paramétricas, es decir, libres de supuestos sobre la distribución de los datos. Esta advertencia fue reforzada por García et al. (2010), quienes insistieron en corregir el error acumulado en las comparaciones múltiples. A ello se suma el problema de la integración, ya que fusionar vistas de naturaleza distinta continúa siendo un problema abierto (Zhao et al., 2017; Liang et al., 2024); de hecho, los esfuerzos más cercanos aplicaron la homología persistente sobre el propio grafo funcional, sin fusionar las medidas del grafo con los estadísticos de persistencia como vistas separadas (Corneanu et al., 2020; Ballester et al., 2024; Zhang y Lin, 2024; Songdechakraiwut y Wu, 2025).

Finalmente, la práctica habitual reúne en una misma muestra redes de familias arquitectónicas distintas, entendidas como los tipos de diseño interno de una red, mezcla que puede aumentar artificialmente la separación observada. Así ocurre al agrupar perceptrones multicapa y redes convolucionales sin controlar el factor de confusión, la variable oculta que distorsiona una comparación (Ballester et al., 2024). Estas limitaciones convergen, por tanto, en un problema central, la limitada disponibilidad de una caracterización unificada de las activaciones validada bajo control estadístico.

Atender ese problema permitiría comparar y auditar redes más allá de su exactitud, razón por la cual en el presente trabajo se fusionaron ambas vistas, el descriptor de grafo funcional $\boldsymbol{X}_1$ (diez medidas del grafo) y el descriptor de homología persistente $\boldsymbol{X}_2$ (doce estadísticos de la forma de la nube), en un solo descriptor multivista, denotado $\boldsymbol{X}_3$. Su validación empleó cuarenta redes de las familias perceptrón multicapa (MLP), que apila capas totalmente conectadas, y red convolucional (CNN), que aplica filtros locales sobre la imagen, entrenadas en tres bancos de imágenes de referencia (CIFAR-10, CIFAR-100 y SVHN). De esta experimentación se obtuvo que el descriptor fusionado reconoce la familia de una red mejor que cada vista por separado; los valores del área bajo la curva ROC (AUC), que valora ese reconocimiento entre 0.5, que es equivalente al azar, y 1, la separación perfecta, fueron de 0.854 y 0.869 para cada vista por separado y de 0.949 con la fusión. Esta ventaja superó además a una línea base, es decir, un modelo de referencia que solo usa el tamaño y la profundidad de la red, en los tres bancos de imágenes. También se mantuvo dentro de un mismo tipo de red, condición en la cual el factor de confusión desaparece.

Con el fin de superar estas limitaciones, esta investigación se propone validar la caracterización multivista de las activaciones de una red neuronal mediante un protocolo estadístico no paramétrico con control del factor de confusión arquitectónico. De esta manera, el reconocimiento de la familia de una red se sustenta en evidencia rigurosa y reproducible, con potencial de extenderse a otras arquitecturas y dominios.

MATERIAL Y MÉTODOS

El descriptor $\boldsymbol{X}_3$ se construyó sobre tres resultados matemáticos. El primero garantiza que el reetiquetado de las neuronas no lo altera; el segundo establece que

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

el número de Betti β_0 [Nota: En todo el artículo, β_k denota el k-ésimo número de Betti; se evita la notación habitual H_k para no confundirla con las hipótesis H_0 y H_1 del contraste estadístico.] (la dimensión de la homología de grado cero) coincide con el número de componentes conexas; y el tercero asegura que los diagramas de persistencia apenas varían ante perturbaciones de la nube de activaciones. Los enunciados y las demostraciones correspondientes se recogen en el material suplementario del repositorio.

Sobre esa base, la investigación tuvo un enfoque cuantitativo, ya que la pertenencia de cada red a su familia arquitectónica se estudió mediante variables numéricas. Su alcance fue correlacional, explicativo e inferencial, dado que primero se cuantificó la asociación entre los descriptores y la organización en familias, después esta se explicó mediante propiedades matemáticas demostrables y, por último, se sometió a contrastes de hipótesis.

De este planteamiento se derivó el diseño, que corresponde a un experimento controlado de corte transversal. Es controlado porque la variable independiente (la familia arquitectónica) se manipuló de forma deliberada al construir el repertorio, es decir, el conjunto de redes entrenadas bajo condiciones prefijadas; y es transversal porque todas las mediciones se obtuvieron en un único momento, tras la convergencia del entrenamiento de cada red. El propósito del descriptor es replicar este diseño, es decir, reconocer a qué familia pertenece cada red a partir únicamente de sus activaciones.

En correspondencia con ese diseño, la población estuvo formada por las redes neuronales profundas entrenadas para clasificar imágenes, y la muestra, de tipo no probabilístico por conveniencia, por un repertorio de $n = 40$ redes en cada dominio (en cada banco de imágenes). Las redes se construyeron en cinco familias; las familias A, B y C, con diez redes cada una, fueron perceptrones multicapa (MLP), que difirieron en la composición de sus capas: la A empleó la activación ReLU (la función que conserva solo los valores positivos de cada neurona), la B añadió la normalización por lotes BatchNorm y la activación suavizada GELU, y la C incorporó bloques residuales (capas con atajos que facilitan el entrenamiento profundo). Las familias D y E, con cinco redes cada una, fueron redes convolucionales (CNN), pequeñas en la D y con bloques residuales en la E. Las cuarenta redes cubrieron exactitudes de prueba de 48.12% a 90.95% (una amplitud de 42.83 puntos porcentuales, pp), de modo que la variación entre familias se fijó por construcción y no se estimó a partir de datos observados. El tamaño de la muestra se fijó en cuarenta redes porque, según el análisis de potencia de Cohen (1988), basta para detectar con potencia adecuada efectos grandes, como los que se esperan entre familias de diseño distinto; por la misma razón se adoptó la validación cruzada, que reserva una parte de las redes para evaluar el modelo ajustado con las demás y que es el remuestreo apropiado cuando las muestras son limitadas (Arlot y Celisse, 2010).

Con esa muestra, la variable independiente fue la familia arquitectónica, un factor nominal con cinco niveles (A, B, C, D y E) distribuidos en treinta MLP (diez por familia) y diez CNN (cinco por familia), asignados por diseño y no por observación posterior, por lo que, las diferencias detectadas pueden atribuirse a la arquitectura. Las variables dependientes fueron de dos órdenes: el primero comprende los rasgos numéricos de los descriptores, con diez medidas del grafo funcional en $\boldsymbol{X}_1$, doce estadísticos de homología persistente en $\boldsymbol{X}_2$ y veinticuatro componentes en su fusión multivista $\boldsymbol{X}_3$; el segundo comprende las

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

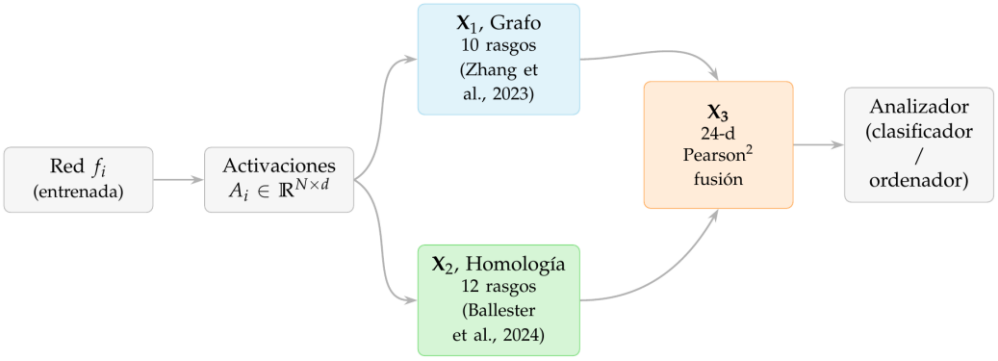
métricas de desempeño con que se evalúa la discriminación entre familias, presentadas en el plan de análisis. Por último, la exactitud de prueba de cada red se tomó como variable interviniente, ya que interviene en la ponderación de la fusión y actúa como criterio externo en las pruebas de asociación.

Para obtener esas variables, las cuarenta redes de cada dominio se entrenaron con tres bancos públicos de imágenes, de acceso libre. CIFAR-10 y CIFAR-100 (Krizhevsky, 2009) contienen 60000 fotografías a color de 32×32 píxeles de animales y vehículos, en 10 y en 100 clases, y se descargan de https://www.cs.toronto.edu/~kriz/cifar.html; SVHN (Netzer et al., 2011) contiene más de 600000 imágenes de dígitos del mismo tamaño, recortadas de números de casas fotografiados desde la calle, y se descarga de http://ufldl.stanford.edu/housenumbers/.

Una vez entrenadas las redes, las variables dependientes se midieron con tres instrumentos matemáticos: el grafo funcional ($\boldsymbol{X}_1$), la homología persistente ($\boldsymbol{X}_2$) y la fusión multivista ponderada, que produce el descriptor multivista ($\boldsymbol{X}_3$). Los tres parten de la misma entrada, la matriz de activaciones $A_i \in \mathbb{R}^{N \times d_i}$ de cada red f_i, que registra la respuesta de cada una de las d_i neuronas de una capa oculta fija ante N imágenes propagadas hacia adelante; sus filas son los ejemplos y sus columnas las neuronas (en las CNN, los canales de salida tras promediar el mapa espacial). A todo descriptor se le exigieron dos condiciones: que la variación entre familias domine a la variación dentro de cada familia, y que no dependa de las etiquetas de las neuronas, que son arbitrarias (propiedad de invariancia demostrada en el material suplementario). La Figura 1 resume el flujo: la matriz de activaciones se procesa por dos ramas, el grafo ($\boldsymbol{X}_1$) y la homología ($\boldsymbol{X}_2$), que se funden en $\boldsymbol{X}_3$ antes de pasar al analizador (el clasificador u ordenador que lo evalúa).

Figura 1

Visión general para la obtención de X_3.



Primer instrumento, el grafo funcional (X_1): este instrumento interpretó las neuronas como nodos de una red funcional. Dos neuronas cuyas activaciones fluctúan sincronizadas se consideraron funcionalmente conectadas, y el grafo resultante se resumió, siguiendo a Zhang et al. (2023), mediante diez medidas de teoría de grafos. Sea $\tilde{A} = A[:, 1: min(d, 80)]$ la submatriz de las primeras $min(d, 80)$ neuronas, cuyo truncamiento determinista

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

reduce el coste computacional. Con la convención de conectividad funcional de Bullmore y Sporns (2009), la correlación entre las neuronas i y j se define como $C_{ij} = Cov(\tilde{a}_i, \tilde{a}_j)/(\sigma_{\tilde{a}_i}\sigma_{\tilde{a}_j})$, donde $\tilde{a}_i$ es la i-ésima columna de $\tilde{A}$. De esta matriz se construyó un grafo no ponderado $G = (V, E)$, con nodos V (las neuronas) y aristas E (sus conexiones), que conserva solo los vínculos suficientemente intensos mediante la umbralización $E_{ij} = \mathbf{1}[|C_{ij}| > \theta]$ con $\theta = 0.4$. Este umbral no se optimizó sobre los datos, sino que se adoptó de Zhang et al. (2023) para preservar la equidad de la comparación, y el material suplementario incluye un barrido de sensibilidad sobre $\theta \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$. El descriptor $X_1 \in \mathbb{R}^{10}$ reúne diez magnitudes que resumen la estructura de G

$$X_1 = [\rho, \overline{C}, \mathcal{T}, \overline{k}, \sigma_k, k_{max}/|V|, n_{\mathrm{cc}}/|V|, |E|/|V|, \overline{c}_B, \sigma_{c_B}]^{\top} \quad (1)$$

en donde ρ es la densidad (fracción de pares de neuronas que quedaron conectados), $\overline{C}$ el coeficiente de agrupamiento medio (proporción de vecinos de una neurona que también están conectados entre sí), $\mathcal{T}$ la transitividad (fracción de tríos conectados que cierran un triángulo), $\overline{k}$ y σ_k la media y la desviación típica del grado (número de conexiones de cada neurona), $k_{max}/|V|$ el grado máximo relativo al número de nodos, $n_{\mathrm{cc}}/|V|$ la fracción de componentes conexas (grupos de neuronas aislados entre sí), $|E|/|V|$ el número de aristas por nodo, y $\overline{c}_B$ y σ_{c_B} la media y la desviación típica de la centralidad de intermediación (frecuencia con que una neurona aparece en los caminos más cortos entre otras dos), denotada así para evitar el choque con los números de Betti β_k. Reetiquetar las neuronas deja invariante a X_1, ya que sus diez medidas son invariantes de grafo (con el truncamiento fijo a las primeras 80 neuronas, la propiedad se conserva en la práctica).

Segundo instrumento, la homología persistente (X_2): a diferencia del anterior, este instrumento descartó las conexiones e interpretó las activaciones como una nube de puntos, un punto por cada ejemplo, cuya forma se analizó escala por escala; al crecer el radio ϵ, las componentes conexas (grado cero, β_0) se fusionan y los ciclos (grado uno, β_1) se cierran. Ese recorrido multiescala constituye la filtración de Vietoris–Rips (Edelsbrunner et al., 2002), y la persistencia de cada rasgo (cuánto perdura al crecer el radio) es la que distingue la señal del ruido de muestreo (Chazal y Michel, 2021; Zhang et al., 2024; Ballester et al., 2026). Como el coste de esta filtración crece rápidamente con el número de puntos y con la dimensión, la nube se comprimió previamente. El análisis de componentes principales (PCA) redujo la submatriz $A[:50] \in \mathbb{R}^{50\times d}$, formada por los primeros 50 ejemplos de evaluación, hasta $r = min(15, d, N - 1)$ componentes, y la nube se denotó como $P = \mathrm{PCA}(A[:50], r)$, reescalada al intervalo $[0, 1]$ mediante su mínimo y su máximo globales. Por lo que la escala absoluta de las activaciones, que es distinta entre familias con y sin normalización por lotes, no interviene en la filtración.

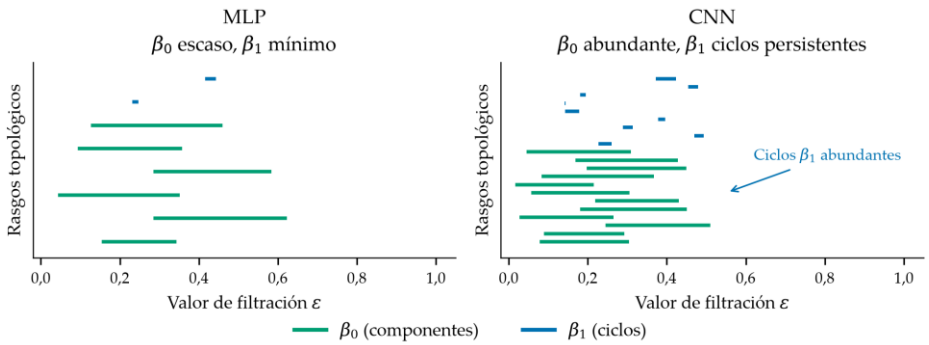
La homología persistente de la nube registra el nacimiento y la muerte de cada clase de grado k a lo largo de la filtración (Edelsbrunner et al., 2002), información que condensa el diagrama de persistencia $\mathrm{Dgm}_k(P) = \{(b_j^k, d_j^k)\}_j$, $k \in \{0,1\}$, donde b_j^k y d_j^k son los radios de nacimiento y muerte del j-ésimo rasgo de grado k, y $p_j^k = d_j^k - b_j^k$ su persistencia. La

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

Figura 2 ilustra esta información como código de barras, con un MLP (panel izquierdo) y una CNN (panel derecho), con los rasgos de grado cero (β_0) en verde y los de grado uno (β_1) en azul; la CNN acumula más componentes y ciclos, y con mayor persistencia.

Figura 2

Códigos de barras de persistencia (esquema ilustrativo).



Como cada diagrama tiene tamaño distinto y el analizador requiere longitud fija, las persistencias $p_k = \{p_j^k\}_j$ se resumieron mediante seis estadísticos descriptivos dados por

$$\sigma_6(\mathrm{Dgm}_k) = [\overline{p}_k, std(p_k), max(p_k), \sum p_k, |\mathrm{Dgm}_k|, P_{75}(p_k)] \quad (2)$$

en donde $\overline{p}_k$ es la persistencia media, $std(p_k)$ su desviación típica, $max(p_k)$ la persistencia del rasgo más duradero, $\sum p_k$ la persistencia total, $|\mathrm{Dgm}_k|$ el número de rasgos del diagrama y $P_{75}(p_k)$ el percentil 75 de las persistencias. Al aplicar σ_6 a los diagramas de grado cero y de grado uno se obtiene $X_2 = [\sigma_6(\mathrm{Dgm}_0), \sigma_6(\mathrm{Dgm}_1)] \in \mathbb{R}^{12}$. Los seis estadísticos son los dados por Ballester et al. (2024), con una única modificación, que la filtración de Vietoris–Rips se ejecutó sobre la nube de puntos de activaciones reducida por PCA, y no sobre un grafo de correlación filtrado (Purvine et al., 2023); ambos diagramas se calcularon con Ripser (Bauer, 2021), la implementación de referencia para códigos de barras de Vietoris–Rips. Los diagramas completos (Figura 2) son estables ante perturbaciones de la nube, según el teorema de estabilidad recogido en el material suplementario; el conteo de rasgos $|\mathrm{Dgm}_k|$ y la persistencia total quedan fuera de esa garantía, por lo que, su estabilidad se comprobó con un repertorio replicado.

Tercer instrumento, el descriptor multivista (X_3): este instrumento integró las dos vistas en un solo descriptor en tres etapas, la estandarización de cada variable, una ponderación según su asociación con la exactitud y una predicción por vista mediante regresión Ridge, una regresión que penaliza los coeficientes grandes para prevenir el sobreajuste. Esta estrategia se alineó con las fusiones multivista con ponderación aprendida de la literatura reciente (Chen et al., 2023; Hao et al., 2025). Con el fin de formalizar

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

estas etapas, sea $y \in \mathbb{R}^n$ el vector de exactitudes de prueba de las n redes. Cada vista se estandarizó, $X_v^* = \text{StandardScaler}(X_v)$, es decir, con cada variable llevada a media cero y varianza uno, y el peso de la variable j de la vista v crece con el cuadrado de su correlación con la exactitud

$$w_j^{(v)} = \frac{\left[r(X_{v,j}^*, y)\right]^2}{\sum_l [r(X_{v,l}^*, y)]^2 + \varepsilon}, v \in \{1,2\}, \varepsilon = 10^{-10} \quad (3)$$

en donde $w_j^{(v)}$ es el peso asignado a la variable j de la vista v ($v = 1$ para el grafo y $v = 2$ para la homología persistente), $X_{v,j}^*$ es la columna estandarizada de esa variable, $r(\cdot,\cdot)$ es el coeficiente de correlación de Pearson entre dos vectores (asociación lineal entre -1 y 1), el índice l recorre todas las variables de la vista, de modo que los pesos de cada vista suman la unidad, y ε es una constante mínima que evita la división por cero. A continuación, una regresión Ridge (Hoerl y Kennard, 1970) con parámetro de penalización $\lambda = 1$ condensó cada vista en una predicción de la exactitud, $\hat{y}^{(v)} = X_v^* \hat{\beta}^{(v)}$, donde $\hat{\beta}^{(v)}$ son los coeficientes estimados. El descriptor final concatenó las dos vistas ponderadas y las dos predicciones así,

$$\boxed{X_3 = [X_1^* \odot w^{(1)}, X_2^* \odot w^{(2)}, \hat{y}^{(1)}, \hat{y}^{(2)}] \in \mathbb{R}^{24},} \quad (4)$$

en donde $X_1^* \odot w^{(1)}$ son las diez medidas del grafo estandarizadas y multiplicadas por sus pesos, $X_2^* \odot w^{(2)}$ los doce estadísticos de persistencia tratados de igual forma, $\odot$ el producto elemento a elemento, e $\hat{y}^{(1)}$ e $\hat{y}^{(2)}$ las predicciones de la exactitud obtenidas con cada vista, lo que suma $10 + 12 + 2 = 24$ componentes. Como esta construcción emplea la exactitud, se planteó un estudio de ablación, que compara el descriptor con y sin uno de sus componentes. La ecuación (4) define la variante adaptativa $X_3^{(a)} \in \mathbb{R}^{24}$, mientras que la de control $X_3^{(c)} \in \mathbb{R}^{22}$ elimina por completo la exactitud, ya que concatena las dos vistas, cada una dividida por su norma euclídea en cada red, sin ponderación ni predicciones, y deja la estandarización por columna al analizador que la ajusta dentro de cada pliegue. Una tercera variante, la de rangos $X_3^{(r)}$, sustituye en la ecuación (3) la correlación de Pearson por la de Spearman (correlación entre rangos). Si las variantes adaptativa y de control alcanzan un AUC de discriminación de familias arquitectónicas (AUC familiar) equivalente, la ventaja de la fusión no depende de la exactitud y queda libre de fuga de información, es decir, del uso indebido de la exactitud de las redes evaluadas.

Definidos los instrumentos, el procedimiento experimental siguió un flujo cronológico de cuatro etapas, resumido en la Figura 1. Primero se generó y entrenó el repertorio de redes, después se extrajeron las activaciones, a continuación se construyeron las dos vistas y, por último, se calculó el descriptor multivista (X_3). El flujo completo se ejecutó primero sobre CIFAR-10, que constituyó el análisis principal, y después se repitió sobre CIFAR-100 y SVHN como validación multidominio; los dominios se analizaron por separado, ya que mezclarlos habría hecho que X_3 reflejara la tarea y no la arquitectura, con lo que, la evidencia total sumó ciento veinte redes entrenadas.

En la primera etapa, el entrenamiento empleó el optimizador AdamW, un método de ajuste por

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

gradiente con decaimiento de pesos desacoplado (Loshchilov y Hutter, 2019), con un decaimiento de 10^{-4}. Con el fin de que los resultados fueran reproducibles, se fijó una semilla (2024), es decir, el valor inicial del generador de números aleatorios establecido en las bibliotecas NumPy y PyTorch, de modo que, dada esa semilla previa, cada ejecución produce la misma secuencia de números aleatorios. Los hiperparámetros (los ajustes que se fijan antes de entrenar) variaron según la configuración: la tasa de aprendizaje (el tamaño del paso de ajuste) fue de 10^{-4} a 5×10^{-3}; el *dropout* (la desactivación aleatoria de una fracción de neuronas durante el entrenamiento), de 0.1 a 0.6; y el número de épocas (réplicas completas del recorrido por los datos de entrenamiento), de 5 a 25 en los perceptrones multicapa y de 15 a 30 en las redes convolucionales. Con esta variación se indujo en cada familia la dispersión de exactitud que el descriptor debe discriminar.

En las etapas siguientes, tras la convergencia se procesaron de ocho a diez minilotes (pequeños grupos de ejemplos) sin calcular gradientes, y se extrajeron las activaciones de la última capa oculta en los MLP o del agrupamiento promedio global del último bloque convolucional en las CNN. Sobre esa matriz se calcularon X_1 y X_2 y, con ambos, el descriptor X_3 de cada red.

Completado ese flujo, el plan de análisis se organizó en torno a una comparación central, la de si X_3 discrimina las familias mejor que cada vista por separado; su hipótesis nula H_0 es que el AUC familiar de X_3 no supera al de la mejor vista individual, mientras que la alternativa H_1 plantea que sí lo supera, con un nivel de significancia de $\alpha = 0.05$. Para resolverla sin suponer previamente ninguna distribución teórica se recurrió al bootstrap, un remuestreo con reemplazo de los propios datos que estima la incertidumbre de una medida y sus intervalos de confianza. La comparación se decidió mediante un bootstrap emparejado sobre la diferencia ΔAUC familiar $(B = 2000)$ aplicado a los mismos remuestreos en los tres dominios, respaldado con pruebas de Wilcoxon de rangos con signo, una prueba no paramétrica para muestras emparejadas, a nivel de pliegue.

Además de esa comparación, el desempeño se resumió con seis indicadores. El principal fue el AUC familiar, ya definido al presentar las variantes del descriptor, al que acompañó el AUC por pares, que evalúa si el descriptor ordena correctamente cada pareja de redes según su exactitud. A ellos se sumaron la d de Cohen, es decir, la diferencia de medias estandarizada entre dos grupos, dos estadísticos de rango que se definen a continuación y un conjunto de diez pruebas con corrección de comparaciones múltiples; cada cifra se acompañó de un intervalo de confianza (IC) al 95%.

Dentro de ese conjunto, cinco pruebas emplearon el contraste de Mann–Whitney (Mann y Whitney, 1947), una prueba no paramétrica que compara dos muestras independientes mediante rangos; su tamaño del efecto asociado $|r|$, acotado entre 0 y 1, cuantifica cuánto separa el descriptor a los dos grupos. En este trabajo se adoptó porque no exige normalidad, supuesto inverificable con cuarenta redes por dominio. Para cada descriptor, la prueba comparó sus puntuaciones en las redes del cuarto superior de exactitud (diez redes, siete en la subpoblación MLP) con las del cuarto inferior.

Las otras cinco pruebas se basaron en la τ de Kendall (Kendall, 1938), un coeficiente de correlación de rangos que mide la concordancia entre dos ordenamientos contando sus pares concordantes y discordantes, sin presuponer una relación lineal. Su

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

aplicación enfrentó el ordenamiento de las cuarenta redes inducido por la predicción Ridge de cada descriptor, con λ elegida por validación cruzada interna entre veinte valores de 10^{-3} a 10^{3}, con el ordenamiento real por exactitud de prueba.

En conjunto, las diez pruebas evaluaron, bajo la hipótesis nula de ausencia de asociación, si cada descriptor contiene información sobre la exactitud, con lo que establecieron que es informativo, pero no que sea superior. Como el conjunto involucró $m = 10$ comparaciones, la tasa de descubrimientos falsos se controló con el procedimiento de Benjamini–Hochberg (Benjamini y Hochberg, 1995), que mantiene las pruebas en $q \leq 0.05$, y se informó también del umbral más severo de Bonferroni, que divide α entre el número de pruebas, $\alpha/m = 0.005$. El mismo control de Benjamini–Hochberg se aplicó a las correlaciones de Spearman entre cada par de rasgos de X_1^* y X_2^*, con las que se midió el solapamiento entre ambas vistas.

En cuanto a la incertidumbre, junto al bootstrap emparejado se empleó la variante *out-of-bag*, traducible como fuera de bolsa ($B = 300$), en la que cada réplica ajusta el modelo sobre la muestra remuestreada con reemplazo y lo evalúa únicamente sobre las redes excluidas de esa muestra. Como los pesos de la ecuación (3) y las predicciones Ridge dependen de la exactitud y, ambos se reajustaron solo sobre el subconjunto de entrenamiento de cada pliegue o remuestreo, nunca con las redes reservadas, lo que evitó la circularidad en la validación. El AUC familiar se calculó mediante un clasificador logístico de uno contra el resto con regularización L_2 de intensidad $C = 1$, con validación cruzada de cinco pliegues estratificada por familia, de modo que cada pliegue conserva la proporción de las cinco familias, y sus IC recogieron la variación entre particiones bajo validación cruzada repetida. Los IC del AUC por pares procedieron, en cambio, del bootstrap fuera de bolsa sobre las $n = 40$ redes, con $\lambda = 1$, mientras que los de la τ de Kendall provinieron de la validación cruzada repetida y los de la d de Cohen, de las predicciones obtenidas dejando una red fuera; en los tres indicadores, el ordenamiento de las redes lo produjo la regresión Ridge, con los mismos hiperparámetros para cada descriptor en los tres conjuntos de datos.

Finalmente, como comprobación de reproducibilidad, las cuarenta arquitecturas de CIFAR-10 se reentrenaron con la misma configuración y la semilla 2024, y se extrajeron de nuevo sus activaciones para repetir el análisis sobre ese repertorio replicado. El código, los registros de entrenamiento, los datos de los descriptores y el material suplementario (demostraciones, solapamiento entre vistas, barrido de θ, robustez y tablas completas por dominio) están disponibles en https://github.com/Dcrr1717/multiview-nn-descriptor.

RESULTADOS Y DISCUSIÓN

Antes de comparar los descriptores se verificó el solapamiento entre las dos vistas. En los $n = 40$ modelos de CIFAR-10, la correlación de Spearman entre pares de rasgos de X_1^* y X_2^* presentó una mediana de $|\rho| = 0.235$, un rango intercuartílico (el tramo que concentra la mitad central de los valores) de $[0.11, 0.47]$ y un máximo de 0.829 ($p = 4.0 \times 10^{-11}$). Tras la corrección de Benjamini–Hochberg 41 de los 110 pares evaluables mostraron una asociación significativa. El patrón se repitió en CIFAR-100 y SVHN con medianas de 0.271 y 0.239, y un 18% y 30% de pares significativos (material suplementario). Ambas vistas compartieron, por tanto, información

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

parcial, algo esperable porque derivan de las mismas activaciones, aunque la mayoría de los pares presentó una asociación débil; si fuesen redundantes, la fusión no elevaría el AUC familiar respecto de cada vista aislada (Zhao et al., 2017).

Con esa complementariedad establecida se determinó, en CIFAR-10, el desempeño discriminativo de los cuatro descriptores (las dos vistas y las dos variantes de X_3) sobre las $n = 40$ redes entrenadas, mediante el AUC familiar, el AUC por pares y la τ de Kendall, como se muestra en la Tabla 1, donde se resalta en negrita el mejor valor de cada columna.

Tabla 1

Desempeño discriminativo ($n = 40$), media [IC 95%].

Descriptor	AUC familiar ↑	AUC por pares ↑	τ ↑
X_1 (Grafo funcional)	0,854 [0,82, 0,89]	0,734 [0,55, 0,87]	0,442 [0,37, 0,51]
X_2 (Persistencia)	0,869 [0,83, 0,92]	0,656 [0,47, 0,84]	0,316 [0,27, 0,37]
$X_3^{(c)}$ (pesos iguales)	**0,949** [0,92, 0,97]	**0,755** [0,55, 0,91]	**0,522** [0,46, 0,58]
$X_3^{(a)}$ (adaptativa)	0,948 [0,91, 0,98]	0,731 [0,58, 0,89]	0,444 [0,36, 0,50]

Con base en la Tabla 1, la fusión de ambas vistas elevó de forma apreciable el desempeño discriminativo. En concreto, la variante $\boldsymbol{X}_3^{(c)}$ alcanzó un AUC familiar de 0.949 (IC 95% $[0.92, 0.97]$) y superó a las dos vistas individuales, con 0.854 para $\boldsymbol{X}_1$ y 0.869 para $\boldsymbol{X}_2$, es decir, unos 8 pp más que la mejor vista individual y 9.5 pp más que $\boldsymbol{X}_1$. Los otros dos indicadores de la tabla coincidieron en el ordenamiento, por lo que, la ventaja de la fusión no dependió de la métrica elegida.

Esta comparación quedó libre de fuga de información, ya que $\boldsymbol{X}_3^{(c)}$ no emplea la exactitud en su construcción (Materiales y métodos) y alcanzó un AUC familiar equivalente al de la variante adaptativa, de modo que la ventaja no dependió de la exactitud. A su vez, las diez pruebas de asociación del conjunto, cinco basadas en la τ de Kendall y cinco en el $|r|$ de Mann–Whitney, superaron tanto la corrección de Benjamini–Hochberg ($q \leq 0.05$) como el criterio más estricto de Bonferroni ($\alpha/10 = 0.005$).

Con el fin de descartar que el tamaño de la red (su número de parámetros y su profundidad) explique el resultado, se contrastó además una línea base trivial que solo emplea el número de parámetros, su logaritmo y profundidad. Esta alcanzó apenas 0.754 de AUC familiar (IC 95% $[0.71, 0.80]$), es decir, 0.195 por debajo de $\boldsymbol{X}_3$, y la brecha tampoco se cerró dentro de los perceptrones multicapa (0.758 y 0.894, respectivamente), por lo que, el desempeño observado no se redujo a un efecto de escala.

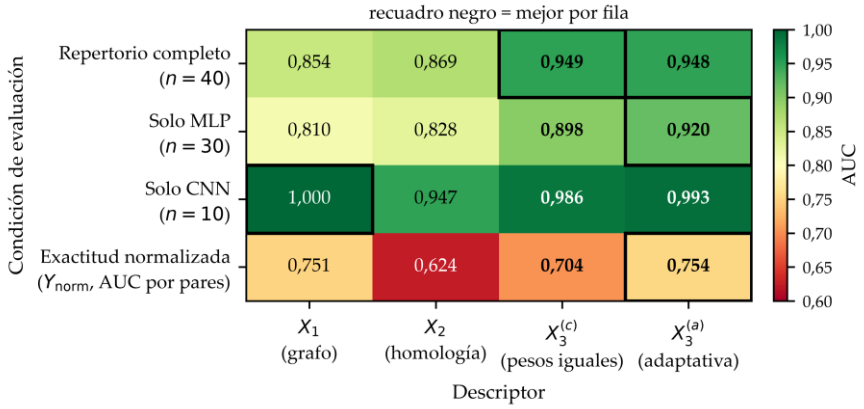
Con el bootstrap emparejado ($B = 2000$) sobre las probabilidades fuera de pliegue de una única partición (las que el clasificador asignó a cada red cuando esta quedó excluida del ajuste), $\boldsymbol{X}_3$ superó al grafo funcional ($\Delta = 0.116$, IC 95% $[0.036, 0.208]$, $p <$

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

0.001) y a la línea base trivial ($p < 0.001$). El competidor más exigente fue la vista topológica $\boldsymbol{X}_2$, respecto de la cual el margen se estrechó a $\Delta = 0.076$ en CIFAR-10, es decir, no cruzó el umbral de significancia ($n = 40$, $p = 0.075$); por ello, esa ventaja sobre la topología pura requiere confirmarse con una muestra mayor.

Figura 3

Desempeño de los cuatro descriptores en cuatro condiciones de evaluación.



Por otro lado, buena parte de los 42.83 pp de dispersión en la exactitud correspondió a un escalón estructural entre los dos tipos de red (MLP o CNN), ya que los MLP se mantuvieron en $\lesssim 60\%$ de exactitud y las CNN ascendieron a $\gtrsim 70\%$. Esta situación implicaba que un descriptor podría destacar solo por separar ambos tipos, lo que constituía un riesgo que atendieron los dos controles resumidos en la Figura 3. El primero restringió el análisis a un mismo tipo y el segundo normalizó la exactitud dentro de cada tipo (Y_{norm}), y en esta figura las columnas son los cuatro descriptores; en las tres primeras filas se muestra el AUC familiar en el repertorio completo, en los MLP y las CNN; y en la última el AUC por pares con la exactitud normalizada. Los recuadros negros señalan al mejor descriptor de cada fila y las tablas por métrica figuran en el material suplementario.

El primer control aisló los MLP ($n = 30$, con apenas 11.8 pp de dispersión en la exactitud), condición en la cual la distinción entre CNN y MLP se anuló por construcción. Aun así, X_3 mantuvo su ventaja, con un AUC familiar de 0.898 ($X_3^{(c)}$, sin exactitud) y 0.920 ($X_3^{(a)}$), superior al de ambas vistas (Figura 3); la τ de Kendall repitió el patrón (0.491, respecto de 0.430 y 0.184) y el AUC por pares resultó equiparable (0.728 y 0.727).

El segundo control estandarizó la exactitud con puntuaciones z, es decir, restando la media y dividiendo por la desviación típica, dentro de cada tipo (Y_{norm}), con lo que, el escalón entre tipos se desvaneció; bajo esta normalización, $X_3^{(a)}$ volvió a encabezar, con un AUC por pares de 0.754 (respecto de 0.751 y 0.624) y un AUC familiar de 0.947, aunque la variante de control quedó por debajo de X_1 en el AUC por pares (0.704 y 0.751, respectivamente). Asimismo, la prueba de Wilcoxon que examina si X_3 predice la exactitud de un modelo individual mejor que X_1 no alcanzó la significancia

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

($p = 0.59$ para la variante de control y $p = 0.78$ para la adaptativa).

Como los resultados anteriores se circunscribían a CIFAR-10, el mismo repertorio de 40 redes se reentrenó en CIFAR-100, donde la tarea es más exigente al pasar de 10 a 100 clases, y en SVHN que es un conjunto de dígitos más sencillo; es decir, solo varió la tarea como reflejaron las exactitudes (6.84% a 66.79% en CIFAR-100 y 70.03% a 96.04% en SVHN), de modo que toda regularidad compartida entre dominios debía atribuirse a la arquitectura y no a la tarea.

Con ese propósito, la Tabla 2 consolida los tres dominios, con el ΔAUC respecto de la línea base trivial y de la más fuerte de X_1 y X_2 (* significativo a $\alpha = 0.05$, ** a $\alpha = 0.01$, *** a $\alpha = 0.001$).

Tabla 2

AUC familiar en los tres dominios (sin fuga de información).

	AUC familiar			**Δ AUC familiar de $X_3^{(c)}$**		
Conjunto	X_1	X_2	$X_3^{(c)}$	**vs. trivial**	**vs. mejor vista**	**Conclusión**
CIFAR-10	0,854	0,869	**0,949**	0,195 ***	0,076 (p=0,075)	No se rechazó H₀
CIFAR-100	0,778	0,870	**0,957**	0,210 ***	0,081* (p=0,011)	Se rechazó H₀
SVHN	0,928	0,913	**0,972**	0,193 **	0,037 (p=0,071)	No se rechazó H₀

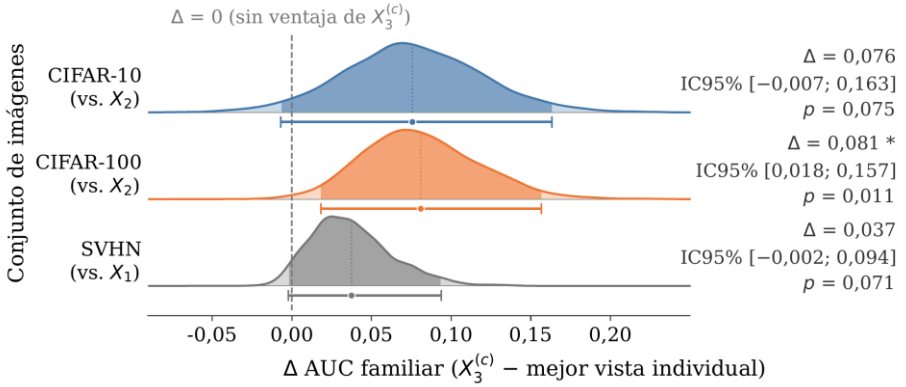
Como muestra la Tabla 2, la variante sin exactitud $\boldsymbol{X}_3^{(c)}$ exhibió el AUC familiar más alto en los tres dominios, superando con holgura a la línea base trivial, y solo en CIFAR-100 superó de forma significativa a la mejor vista individual. A su vez, 29 de las 30 pruebas no paramétricas de los tres dominios superaron la corrección de Benjamini–Hochberg y 28 resistieron incluso Bonferroni. La única caída correspondió al $|r|$ de Mann–Whitney de $\boldsymbol{X}_1$ en CIFAR-100 ($p = 0.273$), cuya τ asociada ($p = 0.037$) pasó Benjamini–Hochberg pero no Bonferroni.

Respecto de la mejor vista individual de cada dominio ($\boldsymbol{X}_2$ en CIFAR-10 y CIFAR-100, y $\boldsymbol{X}_1$ en SVHN), la Figura 4 despliega la distribución del ΔAUC de $\boldsymbol{X}_3^{(c)}$, obtenida al remuestrear con reemplazo las mismas $n = 40$ redes en $B = 2000$ réplicas. Las tres densidades se concentraron a la derecha del cero, con fracciones de réplicas positivas de 0.96, 0.99 y 0.96; solo el intervalo de CIFAR-100 excluyó el cero ($[0.018, 0.157]$), mientras que los de CIFAR-10 ($[-0.007, 0.163]$) y SVHN ($[-0.002, 0.094]$) lo cruzaron apenas por su límite inferior, y en SVHN ambas vistas quedaron muy próximas al máximo (0.928 y 0.913).

CIENCIA INTERDISCIPLINARIA INTERNACIONAL
ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

Figura 4

Distribución bootstrap ($B = 2000$) del ΔAUC de $X_3^{(c)}$ respecto de la mejor vista individual de cada dominio.



La significancia anterior acreditó la existencia de la asociación, pero no informó de su magnitud, para lo cual, se recurrió a la d de Cohen y al $|r|$ de Mann–Whitney entre los cuartos extremos de exactitud. En los cinco descriptores y los tres dominios, 13 de los 15 intervalos de la d excluyeron el cero y 12 mantuvieron su límite inferior por encima de 0.8, con una mediana de $d = 2.293$. Los dos intervalos que cruzaron el cero fueron el de X_1 en CIFAR-100 ($d = 0.563$) y el de la variante de control en SVHN ($d = 0.708$, con $|r| = 0.800$ significativo).

Finalmente, cuatro comprobaciones adicionales detalladas en el material suplementario (estabilidad de los pesos por pliegue, aporte de las predicciones Ridge, AUC por familia y una línea base de metadatos más rica) apuntaron en la misma dirección. A ellas se sumó la réplica de reproducibilidad de CIFAR-10, en la que $X_3^{(c)}$ alcanzó un AUC familiar de 0.929 (IC 95% $[0.89, 0.95]$), mientras que X_1 obtuvo 0.859, X_2, 0.880 y la línea base trivial, 0.755; el orden se conservó, aunque los márgenes se redujeron a 0.056 respecto de la mejor vista ($p = 0.167$) y a 0.178 respecto de la trivial ($p = 0.010$).

Los resultados obtenidos responden a la necesidad, planteada en la Introducción, de contar con una caracterización unificada de las activaciones validada bajo control estadístico. En respuesta a esa necesidad, el descriptor multivista X_3 supera a cada vista aislada en el AUC familiar, ventaja que se sostiene bajo el conjunto de pruebas no paramétricas con corrección de comparaciones múltiples y bajo los dos controles del factor de confusión. La interpretación de estos resultados indica que X_3 captura la organización compartida por cada familia arquitectónica y no un efecto del tamaño, ya que la línea base trivial queda claramente por debajo y la ventaja persiste dentro de la subpoblación MLP. En ella, la frontera entre arquitecturas densas y convolucionales se anula por construcción.

En comparación con los estudios previos, Zhang et al. (2023) emplearon únicamente las medidas del grafo funcional y Ballester et al. (2024) solo los resúmenes de persistencia, sin controlar el factor de confusión arquitectónico, y los antecedentes más próximos,

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

citados en la Introducción, tampoco fusionaron ambas vistas (Corneanu et al., 2020; Zhang y Lin, 2024; Songdechakraiwut y Wu, 2025). A diferencia de esas aproximaciones parciales, la fusión aquí validada supera a la mejor vista aislada y reproduce el patrón en los tres dominios, lo que confirma la utilidad de integrar ambas descripciones.

Sin embargo, bajo la normalización Y_{norm}, la prueba de Wilcoxon no mostró ventaja de X_3 sobre X_1 al predecir la exactitud individual, lo que delimita su alcance a caracterizar familias arquitectónicas. A su vez, el margen sobre las vistas individuales fue menor en SVHN, donde ambas ya discriminaban casi al máximo, lo que indica que fundir fuentes complementarias rinde más cuanto más difícil es la discriminación.

Otro resultado no anticipado es que ponderar con la exactitud no aportó ventaja. La variante de pesos iguales $X_3^{(c)}$ iguala a la adaptativa en el AUC familiar y la supera en el AUC por pares y en la τ de Kendall (Tabla 1), por lo que, se recomienda emplear $X_3^{(c)}$ y reservar $X_3^{(a)}$ para la ablación. Estos hallazgos tienen implicaciones directas, ya que la caracterización multivista permite comparar y auditar redes entrenadas más allá de la exactitud, y la variante de pesos iguales simplifica su aplicación sin sacrificar el desempeño.

Finalmente, las líneas futuras contemplan el incremento de la muestra hasta $n \geq 100$, la incorporación de arquitecturas más diversas, el uso del banco de imágenes Tiny ImageNet de mayor resolución y con 200 clases, la extensión a dominios no visuales. A más largo plazo, se propone X_3 con dos usos: como firma, es decir, descriptor identificador del modelo en su despliegue, y como aproximación (proxy) en la búsqueda de arquitecturas neuronales (NAS) mediante el procedimiento automático que entrena y compara muchas arquitecturas candidatas en las fases tempranas del entrenamiento (Mellor et al., 2021).

CONCLUSIONES

Esta propuesta integra ambos descriptores bajo un protocolo estadístico capaz de controlar el factor de confusión arquitectónico, ya que reúne el grafo funcional y la homología persistente en una sola caracterización de las activaciones. Esta integración ha sido poco explorada bajo control estadístico, y su validez se confirmó al discriminar las familias arquitectónicas.

En efecto, la variante recomendada de pesos iguales que no emplea la exactitud alcanza un AUC familiar de 0.949 en CIFAR-10, unos 8 puntos porcentuales por encima de la mejor vista aislada. Este patrón se reproduce en CIFAR-100 y SVHN, supera a la línea base trivial en los tres dominios y persiste dentro de la subpoblación MLP; de modo que, no se deriva de separar CNN de MLP. Pese a esa ventaja, con $n = 40$ el margen respecto de la vista individual más fuerte solo resulta significativo en CIFAR-100, y X_3 reconoce familias arquitectónicas sin evaluar la calidad de un modelo concreto, ya que, con la exactitud normalizada mediante puntuaciones z, no supera al grafo funcional en la predicción de la exactitud individual.

En consecuencia, el trabajo futuro apunta a superar la dicotomía entre MLP y CNN, y a evaluar los dos usos de X_3 propuestos en la discusión.

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

REFERENCIAS BIBLIOGRÁFICAS

Alain, G., y Bengio, Y. (2016). Understanding intermediate layers using linear classifier probes. *arXiv preprint arXiv:1610.01644*. https://doi.org/10.48550/arXiv.1610.01644

Arlot, S., y Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

Ballester, R., Arnal Clemente, X., Casacuberta, C., Madadi, M., Corneanu, C. A., y Escalera, S. (2024). Predicting the generalization gap in neural networks using topological data analysis. *Neurocomputing*, *596*, 127787. https://doi.org/10.1016/j.neucom.2024.127787

Ballester, R., Casacuberta, C., y Escalera, S. (2026). *Topological data analysis for neural networks*. Springer. https://doi.org/10.1007/978-3-032-08283-1

Baltrusaitis, T., Ahuja, C., y Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Bauer, U. (2021). Ripser: efficient computation of Vietoris–Rips persistence barcodes. *Journal of Applied and Computational Topology*, *5*, 391–423. https://doi.org/10.1007/s41468-021-00071-5

Benjamini, Y., y Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bullmore, E., y Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Chazal, F., y Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, *4*, 667963. https://doi.org/10.3389/frai.2021.667963

Chen, Z., Fu, L., Yao, J., Guo, W., Plant, C., y Wang, S. (2023). Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, *95*, 109–119. https://doi.org/10.1016/j.inffus.2023.02.013

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2.ª ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Corneanu, C. A., Escalera, S., y Martinez, A. M. (2020). Computing the testing error without a testing set. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2674–2682. https://doi.org/10.1109/CVPR42600.2020.00275

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*, 1–30. https://jmlr.org/papers/v7/demsar06a.html

Edelsbrunner, H., Letscher, D., y Zomorodian, A. (2002). Topological persistence and simplification. *Discrete & Computational Geometry*, *28*(4), 511–533. https://doi.org/10.1007/s00454-002-2885-2

García, S., Fernández, A., Luengo, J., y Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, *180*(10), 2044–2064.

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

https://doi.org/10.1016/j.ins.2009.12.010

Hao, P., Gao, W., y Hu, L. (2025). Embedded feature fusion for multi-view multi-label feature selection. *Pattern Recognition*, *157*, 110888. https://doi.org/10.1016/j.patcog.2024.110888

Hoerl, A. E., y Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Horta, V. A. C., Tiddi, I., Little, S., y Mileo, A. (2021). Extracting knowledge from deep neural networks through graph analysis. *Future Generation Computer Systems*, *120*, 109–118. https://doi.org/10.1016/j.future.2021.02.009

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93. https://doi.org/10.2307/2332226

Klabunde, M., Schumacher, T., Strohmaier, M., y Lemmerich, F. (2025). Similarity of neural network models: A survey of functional and representational measures. *ACM Computing Surveys*, *57*(9), 1–52. https://doi.org/10.1145/3728458

Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

Liang, P. P., Zadeh, A., y Morency, L. P. (2024). Foundations & trends in multimodal machine learning: Principles, challenges, and open questions. *ACM Computing Surveys*, *56*(10), 1–42. https://doi.org/10.1145/3656580

Loshchilov, I., y Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*. https://openreview.net/forum?id=Bkg6RiCqY7

Mann, H. B., y Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/aoms/1177730491

Mellor, J., Turner, J., Storkey, A., y Crowley, E. J. (2021). Neural architecture search without training. *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 7588–7598. https://proceedings.mlr.press/v139/mellor21a.html

Moor, M., Horn, M., Rieck, B., y Borgwardt, K. (2020). Topological autoencoders. *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 7045–7054. https://proceedings.mlr.press/v119/moor20a.html

Naitzat, G., Zhitnikov, A., y Lim, L. H. (2020). Topology of deep neural networks. *Journal of Machine Learning Research*, *21*(184), 1–40. https://jmlr.org/papers/v21/20-345.html

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., y Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*. http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf

Purvine, E., Brown, D., Jefferson, B., Joslyn, C., Praggastis, B., Rathore, A., Shapiro, M., Wang, B., y Zhou, Y. (2023). Experimental observations of the topology of convolutional neural network activations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(8), 9470–9479. https://doi.org/10.1609/aaai.v37i8.26134

Raghu, M., Gilmer, J., Yosinski, J., y Sohl-Dickstein, J. (2017). SVCCA: Singular vector canonical correlation analysis for deep learning

ISSN: 3121-2948
Periodicidad: Trimestral, Volumen: 4, Número: 4, Año: 2026 páginas 234-252

dynamics and interpretability. *Advances in Neural Information Processing Systems (NeurIPS)*, 6076–6085. https://proceedings.neurips.cc/paper/2017/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html

Recht, B., Roelofs, R., Schmidt, L., y Shankar, V. (2019). Do ImageNet classifiers generalize to ImageNet?. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 5389–5400. https://proceedings.mlr.press/v97/recht19a.html

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., y Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://doi.org/10.1109/ICCV.2017.74

Simonyan, K., Vedaldi, A., y Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*. https://doi.org/10.48550/arXiv.1312.6034

Songdechakraiwut, T., y Wu, Y. (2025). Functional Connectomes of Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-25)*, 20558–20566. https://doi.org/10.1609/aaai.v39i19.34265

Sun, S. (2013). A survey of multi-view machine learning. *Neural Computing and Applications*, *23*(7–8), 2031–2038. https://doi.org/10.1007/s00521-013-1362-6

Zhang, B., Dong, Z., Zhang, J., y Lin, H. (2023). Functional network: A novel framework for interpretability of deep neural networks. *Neurocomputing*, *519*, 94–103. https://doi.org/10.1016/j.neucom.2022.11.035

Zhang, B., He, Z., y Lin, H. (2024). A comprehensive review of deep neural network interpretation using topological data analysis. *Neurocomputing*, *609*, 128513. https://doi.org/10.1016/j.neucom.2024.128513

Zhang, B., y Lin, H. (2024). Functional loops: Monitoring functional organization of deep neural networks using algebraic topology. *Neural Networks*, *174*, 106239. https://doi.org/10.1016/j.neunet.2024.106239

Zhao, J., Xie, X., Xu, X., y Sun, S. (2017). Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, *38*, 43–54. https://doi.org/10.1016/j.inffus.2017.02.007