



Estructuración bibliográfica acerca de Multiview learning para clasificación de imágenes

Bibliographic structure about Multiview learning for image classification

Nancy Yaneth Gélvez-García  ¹, Kevin C. Díaz-M.  ²,
Carlos Enrique Montenegro-Marín  ³, Paulo Alonso Gaona-García  ⁴

Para citar este artículo: N. Y. Gélvez-García, K. C. Díaz-M., C. E. Montenegro-Marín, P. A. Gaona-García, "Estructuración bibliográfica acerca de Multiview learning para clasificación de imágenes", Revista Vínculos, vol 19, no. 2, pp 156-190, 2022. <https://doi.org/10.14483/2322939X.19926>

Recibido: 10/03/2022 / Aprobado: 19/05/2022

Resumen: En el presente artículo se muestra una revisión bibliográfica de la literatura académica relacionada con "Clasificación de imágenes con Multiview Learning", junto con un análisis de la información presente en cada una de las fuentes bibliográficas revisadas, con el fin de proponer una base conceptual, teórica y estadística para trabajos de investigación que desarrolleen o contengan esta temática. De igual manera se presenta brevemente la

forma en la que se aborda el MVL en los diferentes escenarios de aplicación tanto académicos como prácticos.

Palabras clave: Análisis bibliométrico, Clasificación de imágenes, Estado del arte, Multiview Learning.

Abstract: This article shows a bibliographic review of the academic literature related to "Image classification with Multiview learning" together with an analysis of the

¹ MSc. en Ciencias en Sistemas de Información y Comunicación, Universidad Distrital Francisco José de Caldas, Colombia, Bogotá D.C., Colombia. nygelvezg@udistrital.edu.co

² Ingeniero de Sistemas, Universidad Distrital Francisco José de Caldas, Colombia. Koncilia S.A.S, Colombia. kcdiazm@correo.udistrital.edu.co

³ PhD. en Informática, Universidad Pontificia de Salamanca, España, Salamanca. cemontenegrrom@udistrital.edu.co

⁴ PhD. en Ingeniería de la Información y del Conocimiento, Universidad de Alcalá, España, Alcalá., Colombia. pagaonag@udistrital.edu.co

information present in each of the reviewed bibliographic sources, to propose a conceptual basis, theoretical and statistical for research works that develop or contain this theme. In the same way, the way in which the MVL is approached in the different application scenarios, both academic and practical, is briefly presented.

Keywords: Bibliometric analysis, Image classification, State of the art, Multiview Learning.

1. Introducción

Al realizar una investigación académica en cualquier rama del conocimiento, es común iniciar con la revisión bibliográfica, más conocida como la investigación del estado del arte, con el fin de establecer una noción básica en esa área del conocimiento, y de esta manera comprender las diversas vertientes y tendencias que se están desarrollando, tener una concepción de cuáles son los retos en el campo de estudio y sobre todo, inspirar nuevas ideas que se enfoquen en mejorar o llevar más allá el conocimiento que se está trabajando.

Este documento condensa una investigación enfocada en Multiview

Learning (Aprendizaje de múltiples vistas) en primer lugar de forma general y en segundo lugar aplicado al análisis de imágenes. Dentro de este campo se encuentran múltiples proyectos, problemáticas, y mejoras que ya se han trabajado, por tal motivo es importante realizar esta revisión previa para contextualizar y establecer los límites que se tendrán dentro de la investigación.

Por lo tanto, se establece una metodología, una definición y delimitación del problema (Multiview Learning para la clasificación de imágenes), la cual describe de manera clara y precisa la manera en que se trabajará con las diversas fuentes bibliográficas y su contenido, con el fin de guiar al investigador en la tarea de establecer los fundamentos teóricos básicos y necesarios para su proyecto.

2. Procedimiento para el análisis bibliográfico de clasificación de imágenes con Multiview Learning.

2.1 Definición de la metodología de revisión bibliográfica enfoca al Multiview Learning

En la construcción de la metodología a trabajar se establecen unos pasos a seguir los cuales serán evaluados, esto se hace teniendo en cuenta que el Multiview Learning (de ahora en adelante MVL) es una rama del Machine Learning cuyo enfoque

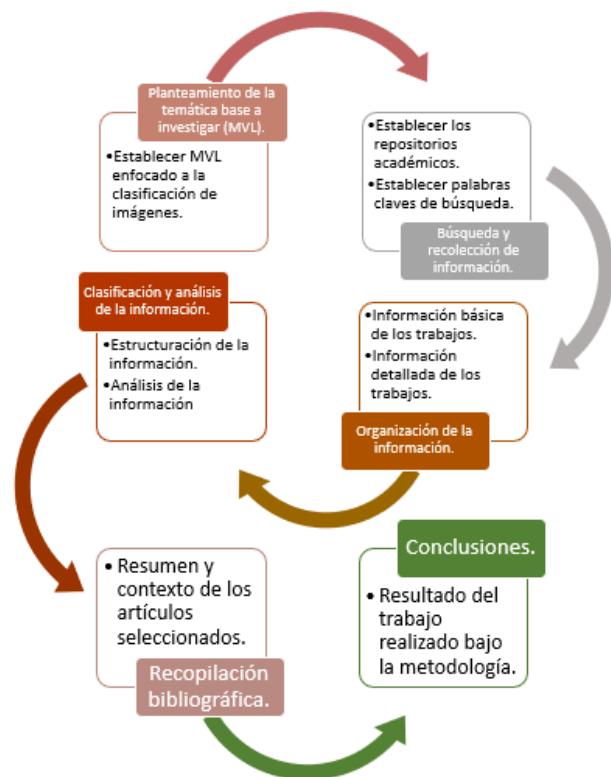
es mejorar el rendimiento y sus resultados al trabajar múltiples vistas. "El aprendizaje de múltiples vistas también se conoce como fusión de datos o integración de datos de múltiples conjuntos de características"[1], es un concepto relativamente nuevo que se ha ido explorando durante la última década con diferentes enfoques, algoritmos y estrategias para el entrenamiento que dependen de los problemas a resolver.

- Planteamiento de la temática base a investigar (MVL).
- Búsqueda general del tema y recolección de múltiples elementos a revisar.
- Organización de información relevante como:
 - Título.
 - Abstract.
 - Palabras Clave.
 - Revista.
 - Fecha.
 - Enlace de la web.
 - Técnicas o metodologías usadas.
 - Temática que aborda.
- Clasificación y análisis del contenido en variables como:
 - Fechas de publicación.

- Frecuencia de técnicas o metodologías que se usan para el Multiview Learning.
- Proporción de las temáticas que se ven trabajadas con Multiview Learning.
- Recopilación bibliográfica
- Conclusiones

En la Figura 1 se presenta de forma gráfica el planteamiento de la metodología.

Figura 1. Diagrama de la metodología de revisión bibliográfica planteada.



Fuente: Elaboración propia.

Estos pasos se establecen basados en una previa revisión de información bibliográfica del tema, en donde se destacan aspectos como la técnica y la temática que aborda cada artículo. Un aspecto importante es el año de publicación del trabajo investigativo, lo recomendable es tener información que haya sido publicada recientemente, pero a su vez no se deben descartar posibles artículos que a pesar de la fecha de publicación pueden ser relevantes por su contenido único.

2.2 Planteamiento de la temática

Para el proceso de investigación se estableció la temática de MVL enfocada al manejo de imágenes, con la finalidad de dar soporte a desarrollos en donde se necesite la clasificación de imágenes.

La temática delimita el espectro del trabajo debido a que MVL puede ser utilizado para trabajar con conjuntos de datos planos, bases de datos que derivan en una búsqueda más extensa y menos enfocada al objetivo.

Al ser un espectro de trabajo enfocado a las imágenes se ha de tener en cuenta además, características, partículas, temáticas y aplicaciones en las que se usan, entre otras, siendo así un amplio espectro sobre el cual MVL se usa para afrontar los escenarios particulares que se presentan.

2.3 Búsqueda y recolección de información

Para la búsqueda de información se utilizaron varios repositorios de información académica, entre los que se destacan:

- Science Direct.
- Springer.
- Oxford Academic.
- IEEE Explore Digital Library.
- Google Academic.

Esto con la finalidad de obtener la información más relevante, se manejaron palabras claves como:

- Multiview Learning.
- Image Classification.
- Multiview Classification.
- Deep Learning Classification Image.

De estas bases de datos se sacaron una cantidad de 50 artículos los cuales a primera vista cumplen con la idea básica de que pertenezcan a la temática a trabajar.

2.4 Organización de la información

Para el control de la información se establecieron los parámetros iniciales como lo son:

- Índice del artículo.
- Nombre del artículo.

- Abstract (Resumen).
- Enlace de consulta.

Esta primera revisión se hace con el fin de filtrar los artículos con una lectura y contextualización de la temática trabajada en cada proyecto. Una vez se tengan seleccionados los trabajos que entraran en la clasificación y análisis, es importante tener en cuenta:

- Palabras clave: da un contexto de los temas que se están trabajando en el artículo, en este caso se establecieron las siguientes palabras clave: Multiview, Machine Learning, classification, image, Deep learning, etc.
- Revista: es el medio en el cual se publica el artículo o un trabajo académico, cada una de ellas tiene una clasificación por cuartiles los cuales escalafonan la revista dando una mejor percepción de la calidad del artículo [2]. En este caso se buscaron en fuentes de información con enfoque a Revistas académicas con cuartiles tipo Q1 y Q2, pero también se tienen en cuenta otras posibles fuentes que puedan aportar a la investigación. La información de cada revista se puede consultar en <https://www.scimagojr.com/> [3].

- Tipo de artículo: es una clasificación dada por el contenido, el cual establece unas etiquetas para el artículo dependiendo del enfoque que tiene, por ejemplo, cuando su foco principal es una revisión bibliográfica, este tendrá dicha etiqueta "Revisión Bibliográfica", si el artículo maneja una sola técnica como SVM este tendrá la etiqueta "Algoritmos", de esta forma se clasifican dependiendo del contenido interno.
- Temática del artículo: es una clasificación basada en qué áreas del conocimiento se usa el multiview (En el caso de que sea un artículo enfocado a la aplicación).
- Técnicas, metodologías o algoritmos: clasifica los temas específicos que se están usando en cada uno de los artículos, con la finalidad de tener un detalle de los planteamientos teóricos que se estén trabajando para resolver el caso de investigación particular de cada artículo.

2.5 Clasificación y análisis de la información

Para la clasificación y posterior análisis se estableció una tabla de información en la cual se condensan aspectos útiles para la consulta rápida, como por ejemplo establecer el nivel de importancia que se le

puede asignar a un artículo dentro del conjunto investigado. Esta cuenta con los siguientes campos:

Tabla 1. Descripción de los campos de la tabla principal de la información.

Campo	Descripción del campo
Id artículo	Identificador del artículo.
Título original	Título en el idioma original.
Título español	Título en español.
Abstract original	Abstract en el idioma original.
Abstract español	Abstract en español.
Revista	Nombre de la revista del artículo.
Calificación	Calificación científica que tiene la revista al momento de realizar el artículo (Q1, Q2, entre otras.)
Categoría revista	Enfoque en conocimiento que tiene la revista.
Fecha artículo	Fecha de publicación del artículo.
URL	Dirección URL donde es tomado el artículo.
Valoración investigación	Calificación de 1 a 3 (1 más importante, 3 menos importante) conforme al aporte del artículo para la investigación.
Tipo de artículo	Categorización del tipo de artículo.

Fuente: Elaboración propia.

Así mismo se crearon 3 tablas auxiliares (Tabla 2, Tabla 3 y Tabla 4) en las cuales se relaciona información importante que se puede extraer de cada artículo y que brinda valor para el análisis, están estructuradas de la siguiente forma:

Tabla 2. Descripción de los campos de la tabla de palabras clave

Campo	Descripción del campo
Id artículo	Identificador del artículo.
Palabras clave	Palabra clave que está contenida en el artículo.

Fuente: Elaboración propia.

Tabla 3. Descripción de los campos de la tabla de temática.

Campo	Descripción del campo
Id artículo	Identificador del artículo.
Temática	Temática que es abordada en el artículo.

Fuente: Elaboración propia.

Tabla 4. Descripción de los campos de la tabla de técnicas, metodologías o algoritmos.

Campo	Descripción del campo
Id artículo	Identificador del artículo.
Palabras clave	Es la técnica, la metodología o el algoritmo que aparece en uso en el respectivo artículo.

Fuente: Elaboración propia.

Con esta estructura se procede a diligenciar la información conforme cada uno de los artículos que se tuvieron en cuenta, tanto por su título, por el abstract, por las palabras y por el posible aporte que pueda tener al tema de MVL.

A continuación, se realiza el análisis de la información teniendo en cuenta los siguientes aspectos:

2.5.1. Por fecha de publicación

Este aspecto se tiene en cuenta para determinar la fecha de publicación de los artículos encontrados y así tomar en consideración los más recientes, de igual manera, establecer desde qué año se ha venido trabajando con MVL en la clasificación de imágenes. A manera de resultado se obtuvo (Figura 2):

Figura 2. Artículos por fecha de publicación.

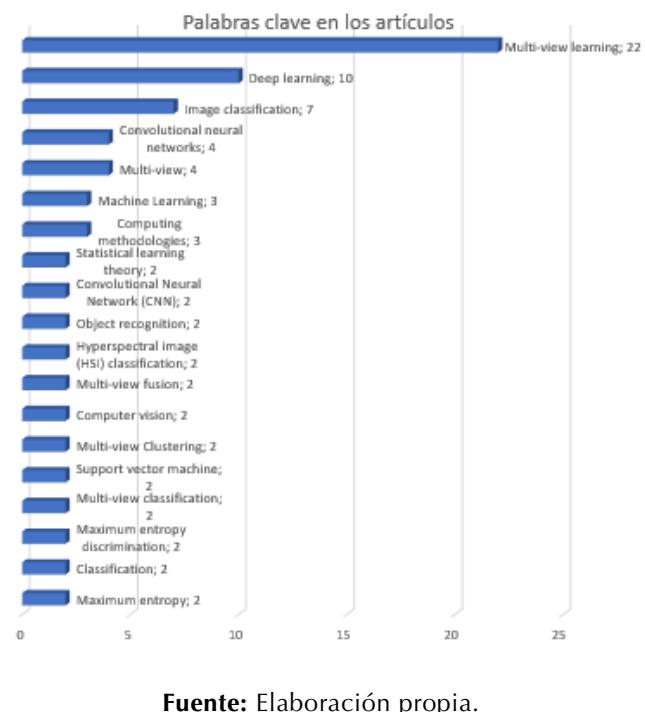


En la Figura 2 se puede observar que ha existido un crecimiento en la investigación de MVL en cuanto al análisis de imágenes se refiere durante los últimos 5 años, y que la temática de MVL se viene trabajando desde el año 2009.

2.5.2 Por palabras clave

En este análisis se buscó establecer qué palabras son utilizadas de forma recurrente por los autores que trabajan MVL y sus temáticas asociadas. Con la información recopilada en la estructura de la Tabla 2 se obtuvieron los siguientes resultados:

Figura 3. Representación de las palabras clave (Keywords) por concurrencia en los artículos.



En la Figura 3 se observa como en la mayoría de los artículos consultados se maneja el concepto de Multiview Learning, seguida de conceptos relacionados como Deep learning, image classification y Convolucionar neural network, los cuales son los temas más relacionados a la hora de

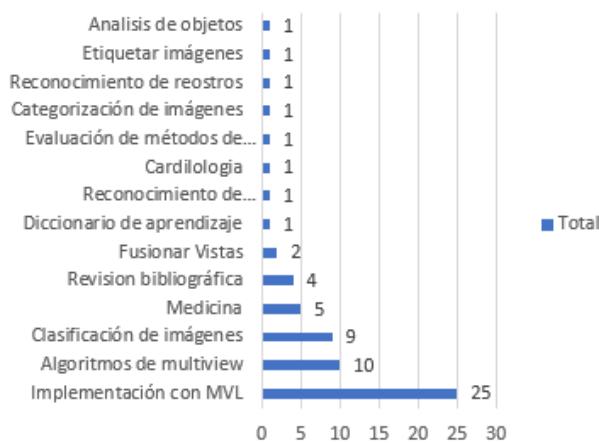
trabajar el MVL enfocado al tratamiento de imágenes. NOTA: Solo se presentan en la figura los conceptos que aparecen más de una vez.

2.5.3. Por técnicas y/o metodologías que usan el multiview

En este análisis se buscó establecer qué temáticas son trabajadas en los artículos, teniendo en cuenta que pueden tener más de una. Un ejemplo de ello es utilizar el MVL enfocado a reconocer violencia en contenido multimedia. Con la información recopilada en la estructura de la Tabla 3 se logró llegar a los siguientes resultados:

Figura 4. Relación de temáticas por concurrencia en los artículos

Relación de temáticas que se trabajan en los artículos



Fuente: Elaboración propia.

Como se evidencia en la Figura 4, la mayoría de los artículos se enfocan en la implementación del MVL para la mejora de un aspecto en la clasificación o en la optimización de algún algoritmo, seguido de otros que se enfocan específicamente en el manejo de MVL y posteriormente en la clasificación de imágenes, cabe resaltar que aparecen temas como la medicina, el análisis de objetos, reconocimiento de violencia o fusión de vistas.

2.5.4. Por técnicas, metodologías o algoritmos que trabajan con el multiview

Por último, un tema importante es la utilización de una técnica, metodología o algoritmo en conjunto con otros como parte del MVL, esto es algo usual en varios artículos o en su defecto se evidencian múltiples técnicas que se usan para abordar diversos problemas que, a pesar de que tengan una misma temática como lo son las imágenes, su propósito es diferente, entre las que se destacan el MVL como base fundamental de la mayoría de los artículos en conjunto con las CNN y las SVMs, también es importante recalcar que entre todas las técnicas se encuentran algoritmos de bajo nivel como DSIFT, SIFT, LBP y HOG para extraer características de las imágenes, y también aparecen variaciones o extensiones de MVL como Deep MVL, MVL semi supervisado, MVL CNN, etc.

En conclusión, este análisis evidencia la gran variedad de temas que se ven relacionados con el MVL lo cual muestra una aceptación de la idea de correlacionar, fusionar y establecer coincidencias entre múltiples vistas, ya sea de imágenes o etiquetas para lograr entrenar sistemas óptimos mediante la utilización de algoritmos y funciones matemáticas, que cada uno de los artículos aborda de la forma más minuciosa y rigurosa posible. Nota: Solo se presentan en la Figura 5 los conceptos que aparecen más de una vez.

Con la información de la Figura 5 se denota que hay temas que son recurrentes en el trabajo con MVL, un ejemplo de ello son los algoritmos de bajo nivel que se usan como base para la creación de las vistas y por ende forman parte fundamental de

algunos procesos de MVL para el análisis de datos.

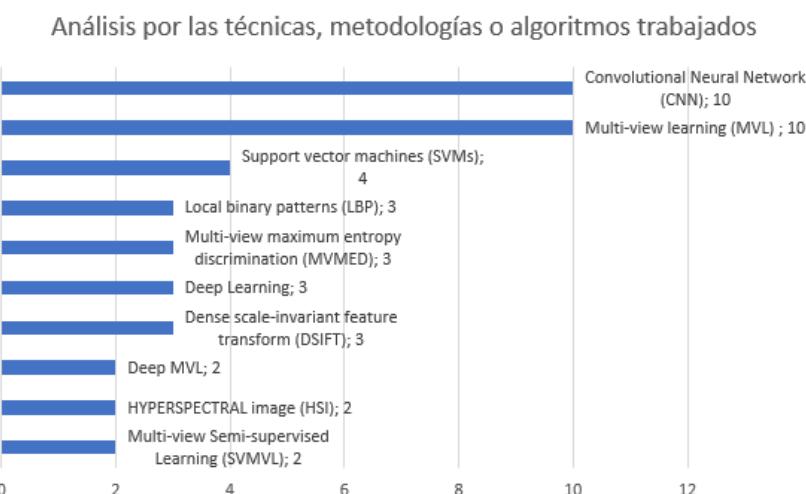
3. Artículos relevantes sobre MVL enfocado a imágenes

Al realizar la lectura de cada uno de los artículos, se buscó rescatar la información más relevante de cada uno de ellos con el fin de hacer un breve resumen de su contenido y de esta manera referencialos en el cuerpo del artículo.

El primero de ellos realiza una clasificación por forma de entrenamiento de los algoritmos que se usan para el MVL en 3 categorías:

- Algoritmos de estilo de coentrenamiento.

Figura 5. Análisis de técnicas, metodologías o algoritmos por concurrencia en los artículos.



Fuente: Elaboración propia.

- Algoritmos de estilo de coregularización.
- Algoritmos de estilo de consistencia de márgenes.

El trabajo titulado “Multiview learning overview: Recent progress and new challenges and new challenges.” [1], presenta algunos algoritmos para cada una de las categorías y su respectivo funcionamiento siendo un punto de referencia clave al ser un artículo de revisión bibliográfica enfocado al MVL.

Un artículo interesante el cual sirve como referencia para el manejo de imágenes es el artículo “Multi-view discriminative marginal metric learning for makeup face verification” [4] ya que toma como referencia algunos algoritmos de bajo nivel como los son HOG, SIFT, TPLBP, LBP Y FBD, con el fin de determinar diferencias entre dos imágenes (una imagen con maquille y otra sin el maquillaje, esto en la misma persona), posteriormente aplican el principio básico de multiview discriminative marginal metric learning (MDMML) para aplicar el modelo de clasificación.

Al realizar este estudio se encontraron algunos algoritmos interesantes y relevantes como los algoritmos de bajo nivel, sobre los cuales también se soporta

otro algoritmo de MVL llamado multi-view maximum entropy discriminant (MVMED+) basado en: model for learning with different numbers of views. o en español (modelo discriminante de máxima entropía de múltiples vistas (MVMED+)); este fue trabajado en el artículo “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features” [5], cuyo objetivo es lograr reconocer imágenes que representan violencia física, para ello se soportan en los algoritmos de bajo nivel que luego se unirán con MVL, estos son:

- DSIFT Feature: una variante del algoritmo SIFT que busca manejar descriptores más densos de los que se manejan en el algoritmo de SIFT, para ello realizan más iteraciones en un tamaño específico de la grilla que se usa en SIFT, para así lograr extraer características invariantes de la imagen.
- HOG Features: el Histograma orientado por gradiente es un descriptor de características usado para la detección de objetos, este se usó en conjunto con una máquina de soporte vectorial o support vector machine (SVM) para obtener el histograma de cada píxel y así normalizar las imágenes.

- LBP Features (Características de Patrones Binarios Locales): es un descriptor basado en la textura en donde se compara un píxel en conjunto con sus vecinos circundantes para así obtener un ponderado de esa textura local y establecer un valor en una escala de grises.

Con la contextualización anterior en el artículo se introduce de forma matemática y algorítmica el concepto de la discriminación por máxima entropía, en donde a grandes rasgos se trata de optimizar una función objetivo en la cual se incluyen los resultados de cada uno de los algoritmos de bajo nivel trabajados previamente. Adicionalmente establecer los valores requeridos para un SMO (Sequential Minimal Optimization). Todo esto en conjunción logra establecer unos resultados para un conjunto de imágenes en comparación con las tratadas de forma individual o en combinaciones pares de las vistas como se muestra en la Figura 6 del artículo.

Figura 6. Precisión de MVMED+ con distinto número de vistas [5].

The accuracy of MVMED+ with different number of views.			
Views	Life	Scene	Physical
view1	77.82 ± 0.47	89.94 ± 0.62	88.35 ± 0.70
view2	78.81 ± 0.45	92.38 ± 0.52	91.70 ± 0.47
view3	85.99 ± 0.57	92.26 ± 0.47	88.74 ± 0.95
view1 + view2	80.29 ± 0.63	94.06 ± 0.54	92.58 ± 0.63
view1 + view3	85.56 ± 0.71	93.07 ± 0.36	90.60 ± 0.60
view2 + view3	86.07 ± 0.46	94.29 ± 0.63	92.52 ± 0.70
view1 + view2 + view3	86.12 ± 0.42	95.10 ± 0.52	92.60 ± 0.41

Otro trabajo importante es el artículo denominado “A Survey on Multi-view Learning” [6], en el cual se presentan los dos principios de relación entre las múltiples vistas:

- El principio de consenso: el cual tiene como objetivo manejar la información de las diferentes vistas y maximizar el “consenso” o la igualdad de la información entre las vistas. La conexión entre estas vistas muestra, bajo algunas superposiciones y leves desigualdades, que pueden existir al maximizar la correlación de la información, dependiendo del caso en concreto, también se puede aplicar utilizando una minimización.
- El principio de complemento: este principio se basa en que cada vista puede contener información que otra vista no posee, mostrando diferente información sobre un mismo objeto de estudio en capas que de por sí, manejan información que en esencia no es identica. Una forma de manejar este concepto es utilizar múltiples clasificadores para extraer información complementaria y que, de forma paralela, se puedan entrenar para hallar comportamientos que unan la información pero que, a su vez, den el complemento esperado.

Continuando con el artículo se plantea el proceso de creación de uso básico de las técnicas de MVL, partiendo de la manera en la que se deben crear las vistas, cómo se deberían combinar, manejando técnicas de coentrenamiento y estableciendo algunos algoritmos de múltiple aprendizaje que se han trabajado a través de la evolución de MVL, como lo son el Múltiple Kernel Learning (MKL) y el Subspace Learning-based Approaches, los cuales exploran todos los algoritmos y técnicas que cada uno de estos conllevan.

En el artículo titulado “Multi-View Image Classification With Visual, Semantic and View Consistency” [7], desarrolla la idea de consenso entre las múltiples vistas, lo que implica que entre las vistas que se están manejado existen características que sean iguales o semejantes. Para ello, aplicaron cada una de las coherencias en vistas separadas y posteriormente utilizaron clasificadores preentrenados para combinar las coherencias, utilizando un método denominado “Linear Multi-View Combination”.

Posteriormente, comenzaron a experimentar con varios datasets para observar comportamientos y objetos que clasifica; también para realizar una comparación en desempeño conforme a otros métodos que se utilizaron como línea

base de comparación, lo cual se puede observar en la Figura 7.

Figura 7. Tabla de “Performance Comparison os VSVC another Baseline Methods” [7].

TABLE V PERFORMANCE COMPARISONS OF VSVC AND OTHER BASELINE METHODS ON THE CUB-200-2011 DATASET. BB: BOUNDING BOX		
Methods	BB	Performance (%)
FC-VGG [45]	no	70.4
bilinear CNN [61]	no	84.1
LRBP[62]	no	84.2
PR-CNN [16]	yes	73.5
Triplet-A [63]	yes	80.7
STN [64]	no	84.1
BoostCNN [65]	no	86.2
VGG-D [66]	yes	82.0
RBF [73]	no	79.0
AGAL [74]	yes	85.5
VSVC(average)	no	87.5
VSVC	no	87.9

Un artículo que es importante tener en cuenta por establecer un marco de aprendizaje semi supervisado es el artículo “Multiview semi-supervised learning for image classification” [8], en el cual se toma la información de imágenes que está preetiquetada. Se establece un sistema de etiquetado de imágenes para entrenar y clasificar las imágenes de forma independiente. Posteriormente, con estas vistas, se usa un principio de mínima entropía para establecer la categoría de la etiqueta que debería llevar las imágenes y así poder clasificar la imagen de forma certera. Algo similar se observa en el artículo de la conferencia de la IEEE ICME del 2009 llamado “Multi-view multi-label active learning for image classification”

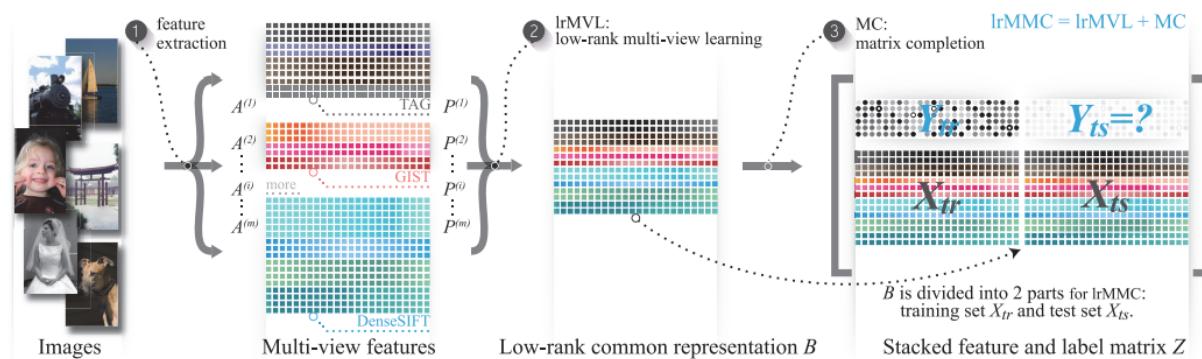
[9], el cual lleva a cabo un aprendizaje activo de múltiples vistas estableciendo una serie de incertidumbres en ellas para fusionarse posteriormente a partir de una muestra-etiquetada que se toma como base.

En el tema de multiview, es algo recurrente buscar la mejora de los algoritmos que ya se están utilizando, esto conforme a la complejidad en los tiempos de clasificación y procesamiento. Un artículo que trata sobre ello es “Low-Rank Multi-View Learning in Matrix Completion for Multi- Label Image Classification” [10], en este artículo se basan en la complejión de matrices para realizar la clasificación de múltiples etiquetas, ya que es un método que se ha trabajado, aquí también se utilizan algoritmos de bajo nivel como DSIFT, TAG, GIST para obtener las múltiples vistas de características de las imágenes. Posteriormente se aplica el

modelo planteado “low-rank multi-view matrix completion (IrMMC) model” y finalmente se utiliza la matriz de complejión para clasificar la imagen, tal como se muestra en la Figura 8.

Un desafío que se presenta en la literatura consiste en el manejo de imágenes en 2D que representan objetos en tercera dimensión (3D). Para abordar esto, se ha trabajado el MVL apoyado en técnicas de redes neuronales convolucionales, como se evidencia en el trabajo titulado “Multi-view Convolutional Neural Networks for 3D Shape Recognition” [11]. En este artículo, abordan las imágenes de perspectivas de un objeto de manera separada; cada una de estas imágenes ingresa a una red neuronal, el resultado de cada una de ellas se agrupa y, posteriormente, se transfiere la información a una red neuronal convolucional con el objetivo de identificar que estas vistas coinciden con

Figura 8. Muestra del proceso de multiview con IrMVL y MC. [10].



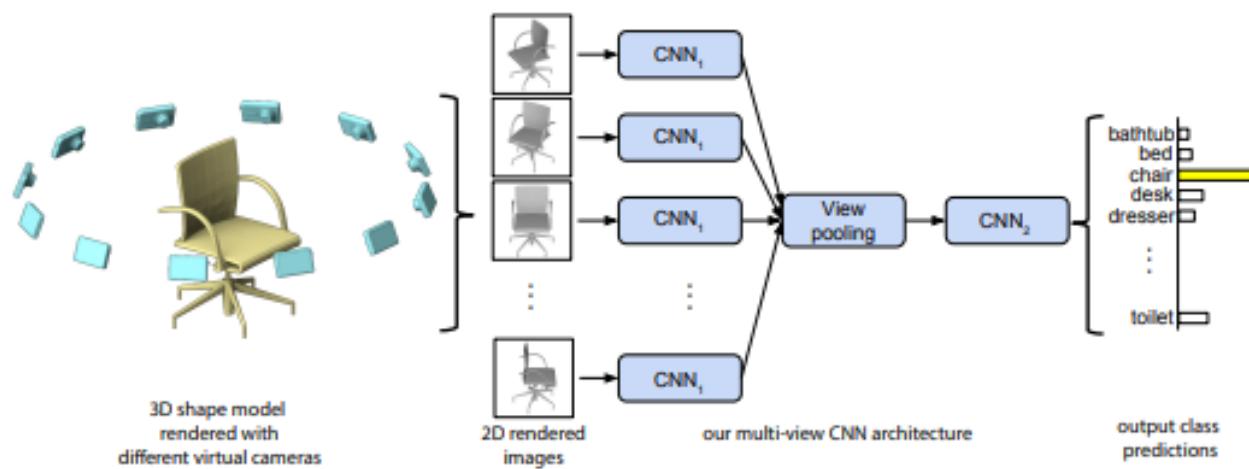
un objeto específico (una imagen de la figura 3D relacionada con múltiples imágenes 2D). Este proceso se puede ver representado en la Figura 9.

Otro gran desafío que se trabaja en el MVL consiste en buscar formas de mejorar el rendimiento de las clasificaciones. Esto ha llevado a la integración de procesos con ramas de la inteligencia artificial, como las redes neuronales y, más específicamente, las redes neuronales convolucionales. Este enfoque se refleja en el artículo “Multi-view classification with convolutional neural networks” [12]. En este trabajo, se utilizan las redes neuronales convolucionales para extraer información de objetos (Plantas insectos y autos) y así implementar tres formas diferentes de fusionar las múltiples vistas: (1) fusionar mapas de características convolucionales a

diferentes profundidades de red (Early fusion); (2) fusión de representaciones latentes de cuellos de botella antes de la clasificación (Late fusion); y (3) fusión de puntuación (Score fusion). De esta manera, estiman cuál es la mejor forma de combinar estas múltiples vistas para los entornos de dominio planteados.

Como se ha observado a lo largo de los artículos revisados, se manejan una gran variedad de algoritmos enfocados en diferentes partes del proceso de MVL. Una parte crucial del multiview es la fusión, concatenación o correlación de las múltiples vistas generadas desde los métodos establecidos en cada investigación. La forma de combinar todas estas vistas se lleva a cabo con redes neuronales. En un nuevo artículo titulado “Deep Multi-View Learning using Neuron-

Figura 9. Multi-view CNN para el reconocimiento de formas en 3D [11].



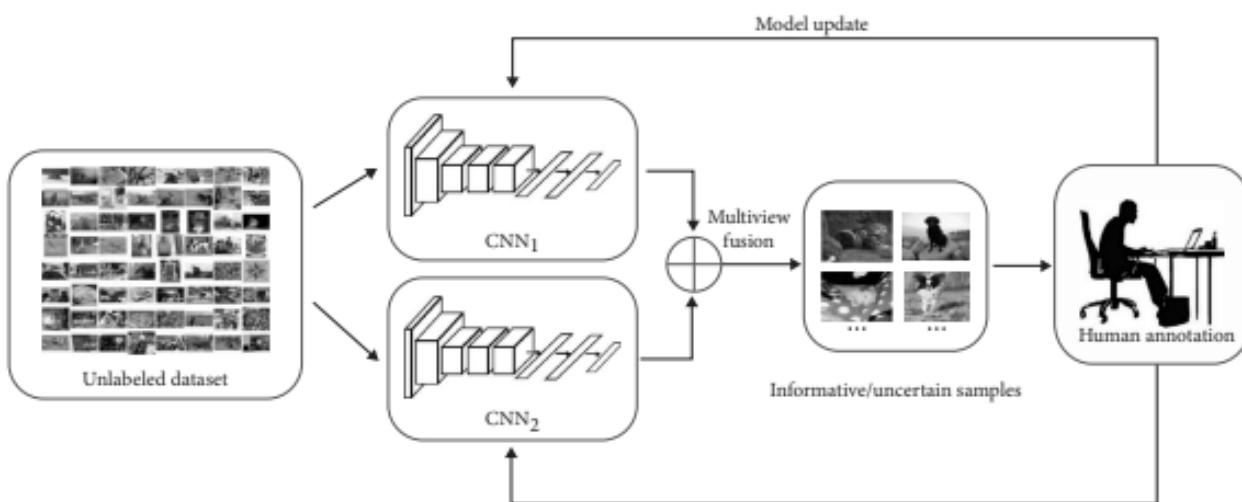
Wise Correlation-MaximizingRegularizers" [13], se toma como base para la fusión de las vistas las redes neuronales profundas (DNN) en su forma básica, seguidamente se introduce el concepto a trabajar: Regularizadores por correlación (CorrReg) los cuales se aplican en las capas de fusión para regularizar los parámetros y optimizar la validación del set de datos.

Continuando con la línea de la clasificación de imágenes utilizando redes neuronales, en el siguiente artículo titulado "A Deep Multiview Active Learning for Large-Scale Image Classification" [14], se presenta el "Deep multiview active learning (DMAL) framework", el cual busca mejorar el aprendizaje activo de múltiples vistas (Multiview active learning (MAL)). Esto se logra realizando pruebas con

variaciones en la cantidad y tamaño de las imágenes, las cuales acceden a dos redes neuronales convolucionales generando dos capas que se unen usando Multiview Learning. Posteriormente, se toman anotaciones de una persona y se retroalimenta activamente a las redes neuronales para complementar la información y mejorar la clasificación.

En la Figura 10, se muestra el proceso en el cual las imágenes son analizadas en dos redes neuronales convolucionales. El resultado ingresa al modelo de multiview fusión, el cual analiza las imágenes y define cuáles puede entender y cuáles no. Posteriormente, esta información se presenta a una persona que retroalimenta a las redes neuronales convolucionales, con

Figura 10. Muestra del proceso de Multiview Learning active [14].



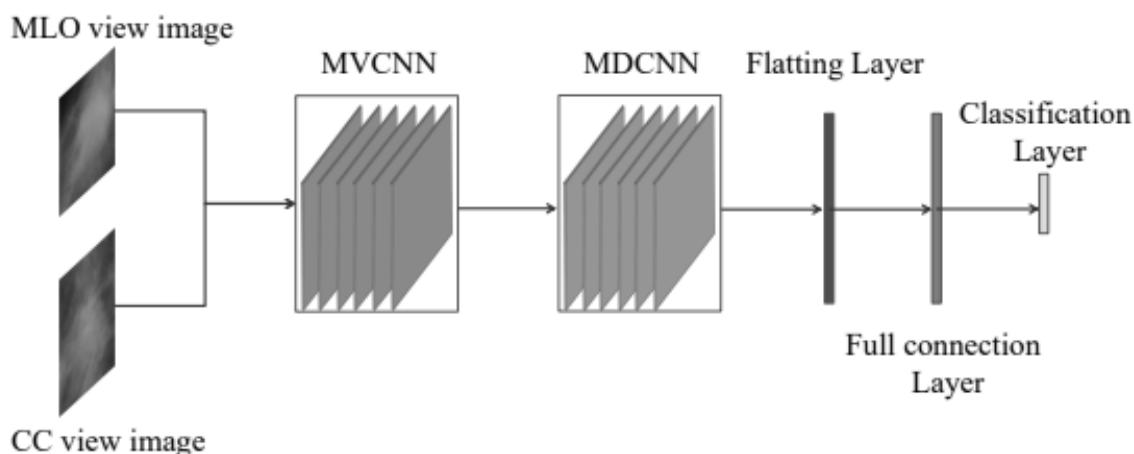
el fin de mejorar los resultados obtenidos en cada una de ellas.

En sí, el concepto de redes convolucionales utilizadas en conjunto con el MVL es algo que se ha ido trabajando, pero a su vez ha tenido implementaciones interesantes en ámbitos como la medicina. Uno de ellos es la clasificación de imágenes mamográficas, como se presenta en el artículo “Multi-View Convolutional Neural Networks for Mammographic Image Classification” [15], aquí se explica cómo utilizar dos técnicas para obtener las vistas de las imágenes mamográficas (mediolateral oblique (MLO) y Craniocaudal (CC)). Una vez obtenidas estos dos tipos de imágenes, proceden a utilizar una arquitectura MVMDCCNN en la cual aparecen dos subredes una convolucional y una vista oblicuo medio lateral, para poder extraer las

características necesarias de las imágenes y así, al final, utilizar funciones de pérdida de entropía cruzada para el proceso de entrenamiento profundo del modelo. Esto se muestra en la Figura 11.

Es importante recalcar que detrás de estas estructuras de multiview hay una parte teórica y algorítmica que se basa en diversos modelos. Por ello, se puede dar lugar a adentrarnos en diversos algoritmos frameworks que se trabajan en el MVL, un artículo que trabaja el Multiview con un Framework de Discriminación por Máxima Entropía es el trabajo denominado: “Multi-View Maximum Entropy Discrimination” [16], en dicho trabajo se manejan conceptos como Maximum Entropy Discrimination (MED) implementados con support vector machines (SVMs) para que al usar las múltiples vistas se pueda

Figura 11. Muestra de la arquitectura MVMDCCNN.



Fuente: [15].

establecer un margen de clasificación idéntico entre las múltiples vistas. Es por ello que realizan tres pruebas con sets de pruebas y al final se observa cómo el modelo denominado MVMED logra una mejoría con respecto a otros métodos.

Otro artículo que se enfoca en un teorema para ser implementado en conjunto con el MVL es "PAC-Bayes analysis of multi-view learning" [17], en el cual se toma como base los límites PAC-Bayes. Seguidamente, se realizan diversas variaciones del algoritmo con máquinas de soporte vectorial SVM semi supervisadas para luego evaluar estas variaciones en unos sets de datos y encontrar los rangos en los cuales se comporta mejor una variación que otra.

MVL es una rama de investigación que, al integrarse en el aprendizaje automático, genera nuevas mejoras en diversas temáticas. Una de ellas es la temática del aprendizaje de múltiples fuentes, tal como se denota en el artículo "Multi-source Transfer Learning with Multi-view Adaboost" [18], en el cual MVL es una parte fundamental para complementar el algoritmo Adaboost, y de esta manera, lograr que las diversas fuentes de información que recibe el algoritmo se puedan manejar generando el nuevo algoritmo MsTL-MvAdaboost.

Es amplio el trabajo que se ha realizado en los últimos años donde interviene MVL. Uno de los trabajos académicos en los cuales se busca adaptar los dominios que no se encuentran etiquetados a dominios que ya se encuentran adaptados, tomando como base el aprendizaje de multiview que ya se tenga, es el trabajado en el artículo "Robust Multi-view Representation: A Unified Perspective from Multi-view Learning to Domain Adaption" [19]. En dicho artículo se presenta una revisión de algoritmos enfocados a los clústeres de múltiples vistas, el aprendizaje de múltiples vistas (Multi-view), y como se pueden adaptar los dominios mediante algoritmos de transferencia del aprendizaje.

Una idea que se trabajó con MVL es el manejo de diccionarios de aprendizaje, en donde se utilizan estos diccionarios para definir y encontrar características y así poder resolver problemas discriminando las imágenes como tal. Esta técnica, al ser usada, tiene algunos aspectos a mejorar como el rendimiento el manejo del ruido en múltiples escenarios etc. Es aquí donde se centra el artículo titulado: "Multi-view low-rank dictionary learning for image classification" [20], ya que realizan un trabajo con MVL supervisado que aprende subdiccionarios y aplica una fórmula de optimización para denotar la data presente en el mapa de la imagen y posteriormente

correlacionarla, discriminando posibles errores pero a un nivel bajo ya que regulariza algunos términos y hace más precisa la correlación.

Como se ha ido viendo en esta revisión, es usual que el MVL se relacione con otras técnicas de aprendizaje automático. En el caso del artículo “Weighted Multi-view Clustering with Feature Selection” [21], se utilizó la técnica de Clusterizado junto con una ponderación dada a las instancias de cada una de las vistas para poder utilizar Multi-view k-means. Este proceso permite realizar una optimización sobre las características que se van a agrupar y obtener así una vista representativa del conjunto de vistas y lo que buscan representar.

MVL no solo se usa para la clasificación, sino también para reconstrucción de imágenes digitales, como se presenta en el artículo: “Automatic 3D building reconstruction from multi-view aerial images with deep learning” [22]. En este trabajo, mediante el uso de Deep learning y un método de Matching de Multiview, se logra identificar las áreas de las construcciones. A través de múltiples vistas de la imagen, se establecen variaciones en su forma y en su diseño para instaurar posibles relieves que, en conjunto con el Deep learning, logran establecer las

características de la construcción para después construir el modelo en 3D.

Continuando con las aplicaciones del MVL en imágenes geográficas, se encontró el trabajo denominado “Forest Fire Recognition Based on Feature Extraction from Multi-View Images” [23], en donde se reconocen incendios extrayendo características como los componentes presentes en la imagen (humo y árboles). Se extraen características de color como el rojo y amarillo del fuego, todo a partir de imágenes en las cuales se realiza un proceso dinámico de extracción. Posteriormente, con redes neuronales gráficas y convolucionales, se depuran las imágenes que contenían incendios y se establecen las dos capas que se manejan en el proceso de segmentación y clasificación.

A medida que MVL avanza en su utilización, se van encontrando temáticas a mejorar. Una de ellas es poder ser optimizada y utilizada en la reconstrucción de escenas que son más complejas de lo usual para los casos de aprendizaje y entrenamiento de un sistema clasificador de imágenes. En este caso, el artículo “Differentiable Diffusion for Dense Depth Estimation from Multi-view Images” [24] trabajado el concepto de Depth Diffusion para posteriormente diferenciar los cambios de los pixeles en la imagen y así

ser optimizados mediante Multi-view Stereo a varios niveles de las imágenes.

Un artículo que es interesante revisar a nivel conceptual es el denominado “Deep multi-view learning methods: A review” [25], ya que en este se hace una revisión de cómo el MVL ha tenido un gran impacto sobre todo utilizando técnicas de aprendizaje profundo, como análisis profundo con correlación, factorización de matrices, análisis de cuello de botella entre otros. Este trabajo incluye algunas comparaciones y muestran algunas aplicaciones de Deep MVL.

También se revisó un trabajo de grado doctoral titulado “Multi-View Imaging: Depth Estimation And Enhancement” [26], en el cual se aborda el manejo de múltiples vistas para soportar el “Stereo Problem”, el cual es un problema de correspondencia entre dos imágenes separadas por una distancia horizontal mínima. Para ello, se utilizan las redes neuronales junto con Multi-View Super-Resolution con el fin de lograr una mejora de las imágenes al hacer las correspondencias entre ellas.

Como se ha ido mencionando, MVL no solo se unen para la clasificación de imágenes, sino que también se ha adentrado en el reconocimiento de múltiples imágenes para poder determinar, a partir de un conjunto de características similares, si un elemento

de una imagen va variando con respecto a otra, pero sigue siendo el mismo. En este trabajo titulado “Multiple View Oriented Matching Algorithm for People Reidentification” [27], se evidencia cómo, a partir de múltiples vistas en diferentes imágenes, se busca obtener la identificación de las personas, su cambio de dirección y una actualización de la perspectiva actual. Un ejemplo se encuentra en la Figura 12, la cual ejemplifica la idea base del modelo para su caracterización.

MVL en su versión de Deep MVL se ha enfocado en mejorar tiempos y optimizar procesos, En el artículo “Deep Multiview Learning for Hyperspectral Image Classification” [28], se trabaja un método de aprendizaje profundo de múltiples vistas para abordar la clasificación de imágenes hiperespectrales (HSI), optimizando la cantidad de etiquetas que se usan para su respectiva clasificación. Esto se logra extrayendo características invariantes en cada vista para posteriormente entrar a una red no supervisada para su entrenamiento y posterior clasificación con Deep MVL.

El aspecto hiperespectral ha tomado una relevancia en MVL, ya que al manejar múltiples dimensionalidades se busca poder realizar una mejora en la clasificación y, por ende, en su rendimiento. Es por ello que en el artículo

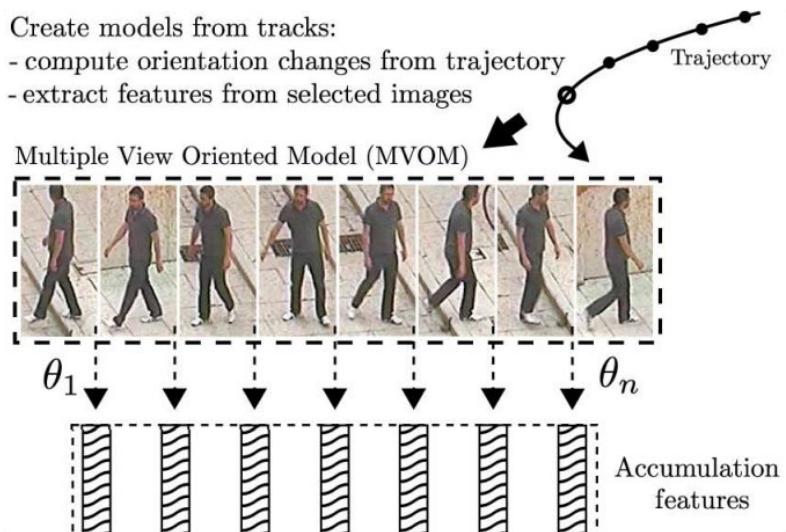
denominado “Hyperspectral Image Classification via Spatial Window-Based Multiview Intact Feature Learning” [29], se propone el modelo spatial window-based multiview intact feature learning method (SWMIFL) para la clasificación de estas imágenes. Este modelo extrae características de las imágenes y aplicando MMVL en vistas intactas para llevar a cabo el proceso de aprendizaje, junto con Multiple morphological component analysis (MMCA) para lograr extraer las características necesarias en el sistema de entrenamiento con SVM.

Otra de las aplicaciones de Deep MVL ha sido la segmentación de múltiples clases para el manejo de ecografías, como se presenta en “Multiview and Multiclass Image Segmentation using Deep Learning

in Fetal Echocardiography” [30], donde se lleva a cabo un proceso de entrenamiento para la detección de cardiopatía congénita (CHD). Esto se realiza a partir de una mejora en la V-Net con abandono espacial, normalización de grupos y supervisión profunda en la segmentación de la información, la cual se integra con las múltiples vistas para completar cualquier información sobre anomalías que no se esté teniendo en cuenta.

En el artículo “Multiview Similarity Learning for Robust Visual Clustering” [31], se plantea un método robusto de aprendizaje de similitud de múltiples vistas (RMvSL). Este método busca implementar rangos bajos en las muestras de las imágenes para eliminar ruido y valores atípicos, de tal forma que se pueda

Figura 12. Muestra de Multiple View Oriented Matching Algorithm [27].



establecer un esquema de múltiples vistas. Este esquema se puede trabajar bajo una función de optimización que realiza la reconstrucción de los datos y sirve para el aprendizaje a la hora de fusionar diferentes gráficos con diversas vistas, logrando así un buen desempeño.

Como se ha visto a lo largo de esta revisión, siempre se está en busca de una mejora con respecto al MVL. En el artículo "Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours" [32], se presenta un modelo supervisado adaptativo denominado: Adaptive MULTiview SEMi-supervised model (AMUSE), el cual busca adaptarse a las ponderaciones en la regularización para así optimizar el algoritmo frente a un nuevo modelo. Esto se logra mediante la toma de los vecinos más cercanos a la matriz que se está trabajando, considerándolos en la función de minimización para adaptarse a los vecinos cercanos. Posteriormente, extienden el multiview con un modelo de clasificación semisupervisada.

Otro ámbito al cual se ha aplicado MVL ha sido las tecnologías basadas en el internet, tal como se muestra en el artículo "Multiview Hessian Regularization for Image Annotation" [33], en donde se basan en el trabajo de la regularización laplaciana que se ha aplicado al aprendizaje automático semisupervisado.

A partir de ahí, se desarrolla la idea de regularización Hessiana Multivista (mHR), con la finalidad de realizar la combinación óptima de múltiples regularizaciones Hessianas, cada una siendo una vista particular de las instancias. Al optimizar esa combinación, se puede mejorar la clasificación de los datos.

En el trabajo titulado "Multi-view metric learning for multi-instance image classification" [34], se aborda la clasificación de imágenes utilizando la extracción de características de las imágenes. Se generan tres vistas de la información las cuales son complementarias entre sí, a su vez, se crea una función para hacer el cálculo entre las vistas y su respectivo aprendizaje, es aquí donde se propone el MVML para optimizar las métricas de dichas distancias y así mejorar su cálculo en las imágenes, teniendo este aspecto en cuenta a la hora de la clasificación de las imágenes.

Como parte de la optimización de la clasificación en el artículo "Auto-Weighted Multi-View Learning for Image Clustering and Semi-Supervised Classification" [35], se plantea el modelo de aprendizaje de vista múltiple que realiza la clasificación de agrupamiento/semisupervisada y el aprendizaje de estructura local simultáneamente. Logra asignar a cada vista un peso óptimo sin la necesidad

explicita de parámetros, optimizando así el modelo de clasificación de imágenes.

Con MVL se han abierto la posibilidad de realizar la concatenación de información que, a primera vista, no parece relacionarse. Un ejemplo de ello es el trabajo “Reconstructing Perceived Images From Human Brain Activities With Bayesian Deep Multiview Learning” [36], en el cual se crea un módulo multivista enfocado a la reconstrucción de imágenes visuales. Esto se realiza a partir de la actividad cerebral registrada en una resonancia magnética funcional, donde en el entrenamiento se ha logrado establecer una relación estadística de las dos vistas: la resonancia y la imagen que actúa como estímulo visual. De esta manera, al obtener una resonancia es posible obtener el estímulo visual aplicado mediante una red neuronal profunda. En la Figura 13, se expone la idea general del desarrollo de este artículo, destacando la primera parte de entrenamiento, donde se relacionan las dos vistas, y cómo, en la prueba de predicción deben comportarse las actividades cerebrales para posteriormente reconstruir el estímulo esperado.

Un artículo que sirve como base para la clasificación supervisada de imágenes es “Experiments with Multi-view Multi-instance Learning for Supervised Image Classification” [37]. En este trabajo utilizan

un clasificador para cada vista, que les permite aplicar el aprendizaje de las múltiples vistas al clasificar las imágenes, siendo este aplicado en varios conjuntos de las mismas.

Una ventaja encontrada al utilizar MVL para el manejo de imágenes es la capacidad de preservar las características de los datos al extraerlos de la fuente original. Esto se explora en el trabajo “Deep Multi-View Enhancement Hashing for Image Retrieval” [38], donde se desarrolla una red neuronal profunda multivista centrada en el aprendizaje del hash (del vecino más cercano en un espacio de datos), lo cual permite recuperar la información de manera eficiente y así evaluar activamente la relación entre las vistas, teniendo en cuenta el ruido como parte de la entrada. En la Figura 14 se evidencia el proceso de tratamiento de la información en este trabajo, donde se presenta el marco propuesto de mejora de hash de múltiples vistas profundas (D-MVE-Hash).

Un tema recurrente en el manejo de imágenes es cuando estas imágenes contienen alguna etiqueta. Esto a llevado al desarrollo de trabajos enfocados en la comprensión de dichas etiquetas, pero a menudo se descuidan las interacciones intramodales que pueden existir entre estas etiquetas. Por esta razón en el artículo

“Multimodal Transformer with Multi-View Visual Representation for Image Captioning” [39], se propone un modelo Transformador Multimodal (MT) que permite la comparación de etiquetas asignadas a imágenes preexistentes. Este modelo combina interacciones intramodales e intermodales con un razonamiento multimodal complejo para

generar etiquetas precisas y acordes al contexto presentando. La Figura 15 ilustra cómo, para una imagen dada, se usa un detector que asigna las etiquetas, generando más vistas que posteriormente entran a un codificador para lograr asignar correctamente la etiqueta.

Figura 13. Muestra de Entrenamiento y predicción del modelo de decodificación Neuronal [36].

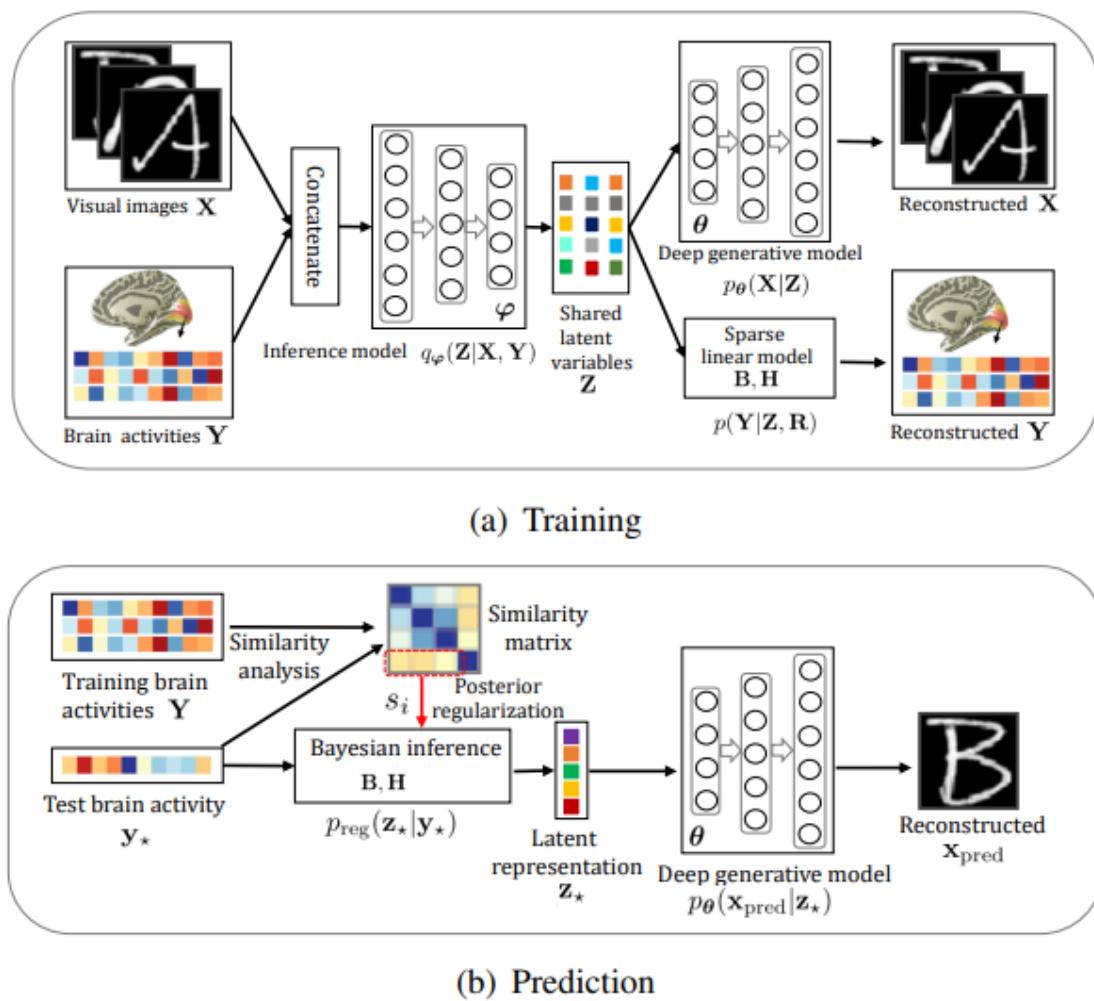
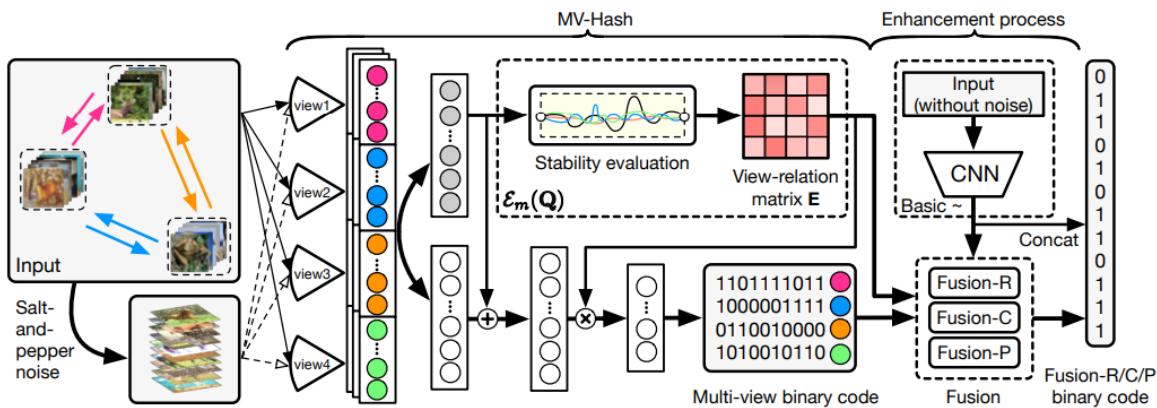
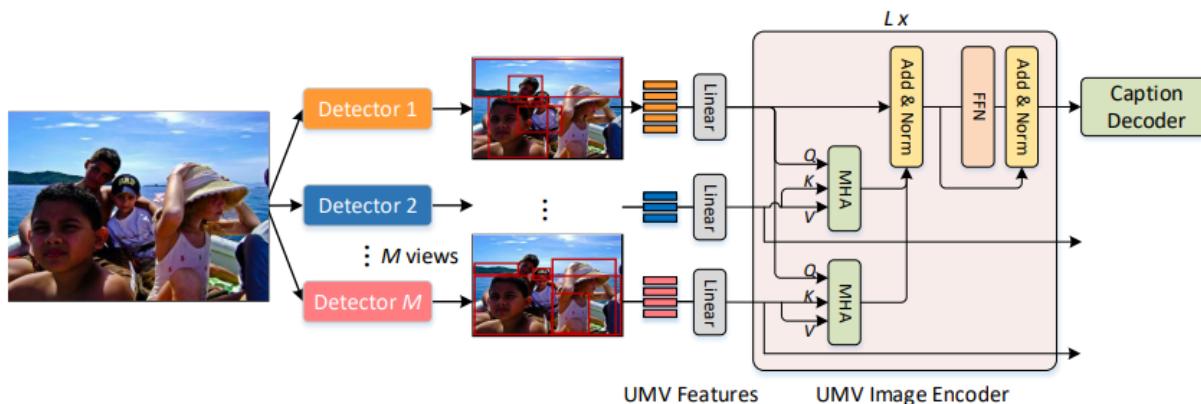


Figura 14. Muestra de (D-MVE-Hash) framework [38].**Figura 15.** Muestra de la estructura del manejo de etiquetas Multimodal [39].

Continuando con los procesos de etiquetado en imágenes con MVL, se ha planteado un modelo en el trabajo de conferencia titulado: "Joint Multi-View Representation Learning and Image Tagging" [40]. En este trabajo se aborda el reto del manejo semántico de las etiquetas de las imágenes, para ello se trabajan dos subespacios donde se definen localmente imágenes ya categorizadas previamente,

así como imágenes con diferentes etiquetas basadas en el mismo ejemplo. Con estas geometrías de múltiples vistas, se establecen puntos vecinos para modelar la estructura semántica de las imágenes, entendiendo que una imagen con x características geométricas puede corresponder a una etiqueta basada en la geometría compartida entre las múltiples vistas, posteriormente se lleva a cabo un

proceso de aprendizaje para entrenar el predictor, que es una función de optimización, y a su vez, esta función se entrena mediante una máquina de soporte vectorial.

Un tema relevante en la investigación en MVL enfocada a las imágenes es el manejo de las múltiples características que cada una de las vistas pueda tener, este tema es abordado en el artículo “Generative multi-view and multi-feature learning for classification” [41]. Al darse cuenta de esta particularidad de las múltiples vistas, se propone un modelo bayesiano generativo, que además de relacionar la característica que se están trabajando con la vista que se está examinando, pueda aprender una representación discriminante de las características en diversas categorías. Esto se logra a partir de un modelo condicional de clases que alterna los parámetros y variables para que el sistema aprenda con base en resultados del mundo real.

MVL ha sido incorporado en diversas áreas, incluida la medicina, tal es el caso del trabajo titulado: “Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis” [42]. En este estudio se busca diagnosticar una patología a partir de imágenes, específicamente para la detección de varios tipos de cáncer, centrándose en el neuroblastoma, que es

el tumor sólido extracranal más común en la infancia. La clasificación se debe realizar en subtipos de neuroblastoma: subtipo indiferenciado (UD), subtipo pobremente diferenciado (PD), subtipo diferenciador (D) y el tipo normal de tejidos no neoplásicos (UN), para lo cual se usan redes neuronales en el análisis de estas imágenes que contienen la patología. Comienzan realizando una extracción de las características para generar las múltiples vistas con algoritmos de bajo nivel como lo son LBP y DSIFT, luego se lleva a cabo una función de clasificación basada en una función discriminante de las múltiples vistas en la parte del clasificador MVMED, donde fusionan las clasificaciones respectivas y se obtiene el resultado final de la clasificación, tal como se muestra en la Figura 16.

Continuando con el análisis de imágenes centradas en la detección, clasificación y diagnóstico médico, se presenta el artículo titulado “Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation” [43]. Este trabajo se enfoca en tomar imágenes y datos de una tomografía computarizada de tórax para detectar los nódulos pulmonares en formato tridimensional. La segmentación de estas imágenes en 3D de los pulmones es un desafío abordado en el artículo, dado que generalmente se trabajan con imágenes en 2D, por esta razón, se propone el

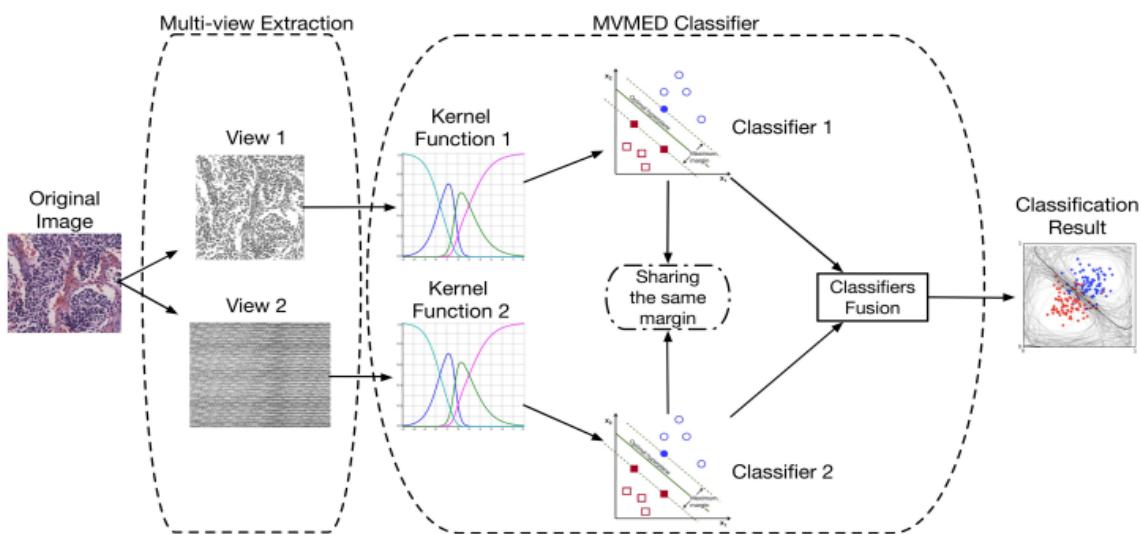
método “multi-view secondary input residual (MV-SIR)”, donde la imagen 3D se convierte en una imagen 2D para su clasificación generando múltiples vistas de esta imagen en planos sagital, coronal y axial. Estas vistas se introducen en módulos de reconocimiento de nódulos para luego combinar los resultados de cada vista y determinar la presencia o ausencia de un nódulo pulmonar. Todo esto se realiza en conjunto con una red neuronal para mejorar la segmentación de las imágenes que contienen nódulos pulmonares. Esta estructura se puede visualizar en la Figura 17.

Un reto que se ha visto en la clasificación de imágenes de múltiples vistas es cuando se trata de relacionar múltiples vistas tomadas desde diferentes ángulos ya que se

debe cruzar coincidencias entre las diferente vistas, para ello se propone un modelo en el trabajo “Multi-view Deep Network for Cross-view Classification” [44], este modelo es llamado “multi-view deep network (MvDN)” en el cual se busca un discriminante no lineal e invariable de una sola vista para las múltiples vistas que se trabajan, para ello se plantearon dos subredes que intentan eliminar las variaciones de las vistas y establecer una representación común y compartida de todas las vistas y así a partir de múltiples muestras poder guiar el aprendizaje en estas redes.

En el artículo “On Deep Multi-View Representation Learning: Objectives and Optimization” [45], los autores hacen una revisión de varias técnicas que ellos han

Figura 16. Muestra de la estructura de la arquitectura del MVMED [42].

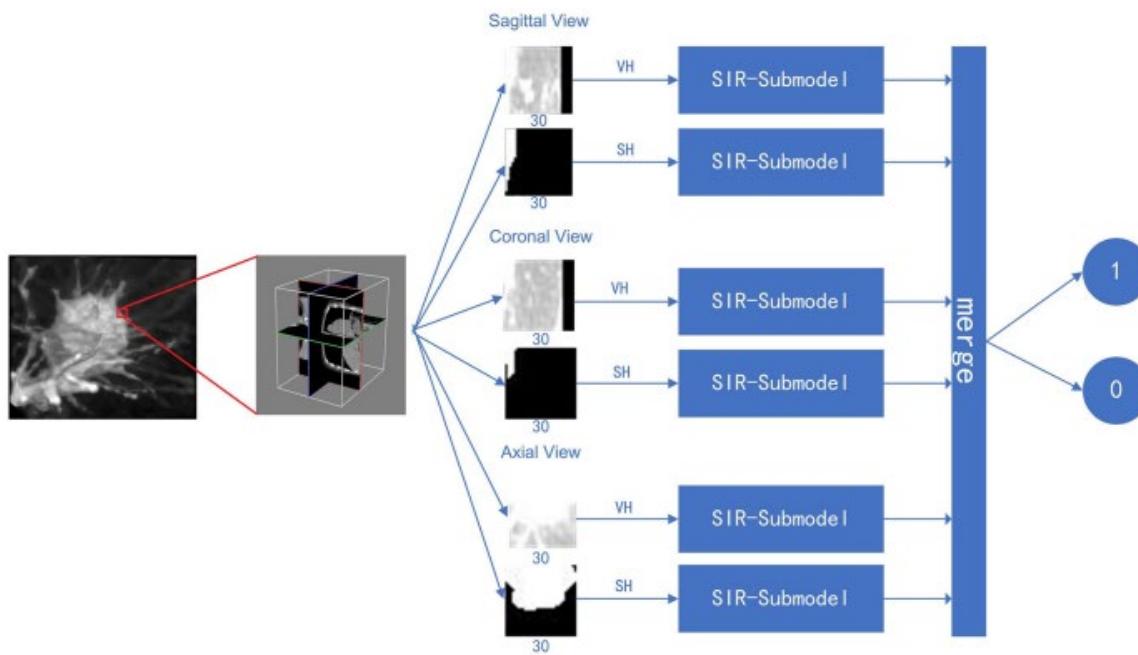


ido trabajando como variantes de las representación de las múltiples vistas a las cuales les han agregado un etiquetado, entre ellas se caracterizan varias técnicas de redes neuronales entre otras técnicas y algoritmos de coherencia, aproximación y correspondencia, entre información de múltiples vistas para así obtener indicadores de evaluación y clasificación, estos posteriormente son comparados contra el modelo propuesto por los autores denominado “deep canonically correlated autoencoders (DCCAE)” en el cual se hace el estudio de una optimización no lineal de algoritmo CCA (Canonically correlated autoencoders) para disminuir tiempos de

ejecución obtener una mejora en comparación con los algoritmos trabajados en todo el artículo.

Conforme a la clasificación de imágenes, algunas de ellas se encuentran pseudoetiquetadas es por ello que en el trabajo titulado “Improving Image Classification Quality Using Multi-View Learning” [46], utilizan el principio de máxima entropía para poder asignar una etiqueta de categoría adecuada a una imagen que no tiene etiqueta usando clasificadores específicos de vista que han sido entrenado de manera óptima, haciendo que no se requiera tener una

Figura 17. Muestra de la estructura del modelo MV-SIR y proceso de entrenamiento del modelo [43].



etiqueta en específico para asignar sino que de forma automática los clasificadores asignan la etiqueta que consideran acorde a la vista.

También los conjuntos de imágenes a clasificar se enfocan en una temática como lo pueden ser los rostros, lo cual es una rama de clasificación como tal ya que esta representa retos para establecer con precisión la identidad de alguien basado en el rostro, esto se trabaja en el artículo “Multi-View Perceptron: a Deep Model for Learning Face Identity and View Representations” [47], en el cual se propone una red neuronal profunda la cual tenga un perceptrón de vista múltiple en el cual se puede identificar características de identidad para cada vista y poder concatenarla en una sola vista logrando interpolar y predecir la identidad de la persona a pesar de que se trabajen diferentes puntos de vista de una persona.

Al realizar la revisión se encontró un trabajo reciente titulado “A Multi-view Convolutional Neural Network Approach for Image Data Classification” [48] , el cual centra su investigación en una red convolucional de múltiple vista enfocado a las múltiples vistas para la extracción y clasificación de las características de las imágenes “Multi-view convolutional neural network (MvCNN)” el cual busca una mejora en el rendimiento de las redes

neuronales convolucionales que se han trabajado de la mano con MVL esto se comparó con “single-view learning of CNN (SvCNN)” y se establece la metodología trabajada para la mejora y comparación entre MvCNN vs SvCNN.

Como se ha presentado a lo largo de esta revisión se suelen usar etiquetas para las múltiples vistas, esto genera que al hacer los procesos se exploren subespacios en común para combinar las vistas y obtener la información que sea más útil, en el trabajo denominado “Multi-View Multi-Label Learning with View-Specific Information Extraction” [49], lo que realizan es la utilización de la información específica que se extraen de los subespacios de cada vista para posteriormente minimizar la perdida de esta información y que las etiquetas múltiples contengan más información que pueda ser útil, esto imponiendo una serie de restricciones para así utilizar esta información discriminatoria de la vista que se está evaluando.

Una gran implementación de MVL enfocada a la clasificación de imágenes es la que se realizó en el artículo “Cross-View Representation Learning for Multi-View Logo Classification with Information Bottleneck” [50], en la cual se enfrentan al problema de las pequeñas variaciones que puede tener una apariencia de un logotipo,

y los algoritmos usuales de MVL se enfocan en hallar la máxima correlación entre las diferentes vistas, es aquí donde los autores se enfocan investigando los posibles cuellos de botella en la información a la hora de extraer las diferentes características comunes de la vista y posteriormente maximizaron la información mutua que había entre dos vistas para preservar las características clave de la clasificación, y al mismo tiempo realizan una eliminación de la información redundante que no se comparte entre las dos vistas.

Y ya para finalizar la revisión de artículos con temáticas enfocadas al MVL en imágenes y sus temáticas asociadas, se revisó el artículo “Embedding Regularizer Learning for Multi-View Semi-Supervised Classification” [51], en el cual su propósito es establecer un marco de trabajo para la integración de un regularización de información esto para poder integrar información complementaria a las vistas y de esta forma basarse en una matriz de etiquetas para la clasificación de múltiples vistas, esta técnica se denominó “Embedding regularizer learning scheme for multi-view semi-supervised classification (ERL-MVSC)”, e incluyen una aspecto de optimización conjunta de un regularización de incrustación de información y la matriz de información

predicha mediante un método de manejo de coordenadas por descenso.

4. Conclusiones

La investigación estructurada que se llevó a cabo basada en la metodología propuesta, permitió establecer las bases conceptuales más relevantes a tener en cuenta al seleccionar los artículos relacionados con MVL en imágenes. Esto se logró mediante la clasificación y depuración de los datos extraídos de diversas fuentes académicas consultadas.

Durante la revisión de los artículos de MVL se recopiló información suficiente para construir una muestra representativa de los temas más relevantes. Se identificaron las temáticas en las que se está enfocando la investigación actual, destacando cuáles son las más recientes y aquellas que están siendo abordadas en el ámbito académico, especialmente en relación con la clasificación de imágenes.

5. Agradecimientos

En reconocimiento al grupo de investigación GIIRA, al cual pertenecen los autores, este artículo se deriva de la

investigación de la tesis doctoral “Modelo computacional cognitivo para la predicción de preferencias de usuario, basado en análisis de emociones y preclasificación de contenido, orientado a la visualización de datos”, de igual manera sirve como base para la tesis de maestría “modelo de software para el manejo de imágenes culturales basado en multiview learning”, la cual es también se derivada de la tesis doctoral anteriormente nombrada.

Para obtener más detalles sobre la metodología que estructuró el presente artículo, se recomienda revisar el artículo denominado “Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización” [52].

Referencias

- [1] J. Zhao, X. Xijong, X. Xin y S. Shiliang, “Multi-view learning overview: Recent progress and new challenges and new challenges”, *Information Fusion*, pp. 43-54, 2017. <https://doi.org/10.1016/j.inffus.2017.02.007>
- [2] Ministerio de Ciencias de Colombia, “Modelo de clasificación de revistas científicas nacionales”, 2021. <https://minciencias.gov.co/pagina/modelo-de-clasificacion-de-revistas-cientificas-nacionales>
- [3] Scimago Lab, “Scimago Journal & Country Rank” (2021, September 07). [Online]. Available: <https://www.scimagojr.com/>
- [4] L. Zhang, H. Shum, L. Liu, G. Guo y L. Shao, “Multiview discriminative marginal metric learning for makeup face verification”, *Neurocomputing*, pp. 339-350, 2019. <https://doi.org/10.1016/j.neucom.2018.12.003>
- [5] S. Sun, Y. Liu y L. Mao, “Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features discrimination and deep features”, *Information Fusion*, pp. 43-53, 2018. <https://doi.org/10.1016/j.inffus.2018.10.004>
- [6] C. Xu, T. Dacheng y C. Xu, “A Survey on Multi-view Learning” arXiv.org, 2013.
- [7] C. Zhang, J. Cheng y Q. Tian, “Multi-View Image Classification with Visual, Semantic and View Consistency”, *IEEE Transactions on Image Processing*, vol. 29, pp. 617-627, 2019. <https://doi.org/10.1109/TIP.2019.2934576>
- [8] S. Zhu, X. Sun y D. Jin, “Multi-view semi-supervised learning for image

- classification", *Neurocomputing*, pp. 136-142, 2016.
<https://doi.org/10.1016/j.neucom.2016.02.072>
- [9] X. Zhang, J. Cheng, C. Xu, H. Lu y S. Ma, "Multi-view multi-label active learning for image classification", *2009 IEEE International Conference on Multimedia and Expo*, 03 07 2009.
- [10] M. Liu, Y. Luo, D. Tao, C. Xu y Y. Wen, "Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification", *Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 29, nº 1, 2015.
<https://doi.org/10.1609/aaai.v29i1.9547>
- [11] H. Su, S. Maji, E. Kalogerakis y E. L. Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition", *arXiv*, 2015.
<https://doi.org/10.1109/ICCV.2015.114>
- [12] M. Seeland y P. Mäder, "Multi-view classification with convolutional neural networks" *PLOS ONE*, 2020.
<https://doi.org/10.1371/journal.pone.0245230>
- [13] K. Jia, J. Lin, M. Tan y D. Tao, "Deep Multi-View Learning using Neuron-Wise Correlation-Maximizing Regularizers", 2019.
<https://doi.org/10.1109/TIP.2019.2912356>
- [14] Y. Tuozhong, W. Wenfeng y G. Yuhong, "A Deep Multiview Active Learning for Large-Scale Image Classification", *Mathematical Problems in Engineering*, 2020.
<https://doi.org/10.1155/2020/6639503>
- [15] L. Sun, J. Wang, Z. Hu, Y. Xu y Z. Cui, "Multi-View Convolutional Neural Networks for Mammographic Image Classification" 2019.
<https://doi.org/10.1109/ACCESS.2019.2939167>
- [16] M. Ángel Barrera Pérez, N. Y. Serrato Losada, E. Rojas Sánchez, G. Mancilla Gaona, "State of the art in software defined networking (SDN)", *Visión electrónica*, vol. 13, no. 1, pp. 178–194, 2019.
<https://doi.org/10.14483/22484728.14424>
- [17] S. Sun, J. Shawe-Taylor y L. Mao, "PAC-Bayes analysis of multi-view learning", *Information Fusion*, pp. 117-131, 2016.
<https://doi.org/10.1016/j.inffus.2016.09.008>
- [18] Z. Xu y S. Sun, "Multi-source Transfer Learning with Multi-view Adaboost", *de International Conference on Neural Information Processing, Doha, Qatar*, 2012. https://doi.org/10.1007/978-3-642-34487-9_41
- [19] Z. Ding, M. Shao y Y. Fu, "Robust Multi-view Representation: A Unified

- Perspective from Multi-view Learning to Domain Adaption", de *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018.
<https://doi.org/10.24963/ijcai.2018/767>
- [20] F. Wu, X. Y. Jing, X. You, D. Yue, R. Hu y J. Y. Yang, "Multi-view low-rank dictionary learning for image classification," *Pattern Recognition*, pp. 143-154, 2015.
<https://doi.org/10.1016/j.patcog.2015.08.012>
- [21] Y.-M. Xu, C.-D. Wang y J.-H. Lai, "Weighted Multi-view Clustering with Feature Selection", *Pattern Recognition*, pp. 25-35, 2015.
<https://doi.org/10.1016/j.patcog.2015.12.007>
- [22] D. Yu, S. Ji, J. Liu y S. Wei, "Automatic 3D building reconstruction from multi-view aerial images with deep learning", *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 155-170, 2021.
<https://doi.org/10.1016/j.isprsjprs.2020.11.011>
- [23] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu y Y. Xu, "Forest Fire Recognition Based on Feature Extraction from Multi-View Images", *Traitemet du Signal*, pp. 775-783, 2021.
<https://doi.org/10.18280/ts.380324>
- [24] N. Khan, M. Kim y J. Tompkin, "Differentiable Diffusion for Dense Depth Estimation from Multi-view Images", *arXiv*, 2021.
<https://doi.org/10.1109/CVPR46437.2021.00880>
- [25] X. Yan, S. Hu, Y. Mao, Y. Ye y H. Yu, "Deep multi-view learning methods: A review", *Neurocomputing*, pp. 106-129, 2021.
<https://doi.org/10.1016/j.neucom.2021.03.090>
- [26] J. N. Oliver, "Multi-View Imaging: Depth Estimation and Enhancement", 2019. [Online]. Available: <https://dspace.uib.es/xmlui/handle/11201/150580>
- [27] J. García, A. Gardel, I. Bravo y J. L. Lázaro, "Multiple View Oriented Matching Algorithm for People Reidentification", *IEEE Transactions on Industrial Informatics*, pp. 1841 - 1851, 2014.
<https://doi.org/10.1109/TII.2014.2330976>
- [28] B. Liu, A. Yu, X. Yu, R. Wang, K. Gao y W. Guo, "Deep Multiview Learning for Hyperspectral Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 7758-7772, 2021.
<https://doi.org/10.1109/TGRS.2020.3034133>
- [29] Y. Zhao, Y.-m. Cheung, X. You, Q. Peng, J. Peng, P. Yuan y Y. Shi, "Hyperspectral Image Classification via Spatial

- Window-Based Multiview Intact Feature Learning", *IEEE Transactions on Geoscience and Remote Sensing*, pp. 2294-2306, 03 03 2021.
<https://doi.org/10.1109/TGRS.2020.3004858>
- [30] K. Wong, E. Sinkovskaya, A. Abuhamad y T. Mahmood, "Multiview and Multiclass Image Segmentation using Deep Learning in Fetal Echocardiography" *Proceedings of SPIE - The International Society for Optical Engineering*, 2021.
<https://doi.org/10.1117/12.2582191>
- [31] A. Li, J. Chen, D. Chen y G. Sun, "Multiview Similarity Learning for Robust Visual Clustering" de *Computer Vision - ACCV 2020 Workshops: 15th Asian Conference on Computer Vision, Kyoto, Japan, Tokio, 2020.*
https://doi.org/10.1007/978-3-030-69756-3_12
- [32] F. Nie, G. Cai y X. Li, "Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours" *IEEE Transactions on Knowledge and Data Engineering*, pp. 2389 - 2400, 2020.
<https://doi.org/10.1109/TKDE.2019.2920985>
- [33] W. Liu y D. Tao, "Multiview Hessian Regularization for Image Annotation" *IEEE Transactions on Image Processing*, 2019.
- [34] D. Li y Y. Tian, "Multi-view metric learning for multi-instance image classification" *Expert Systems with Applications*, 2021.
- [35] F. Nie, G. Cai, J. Li y X. Li, "Auto-Weighted Multi-View Learning for Image Clustering and Semi-Supervised Classification" *IEEE Transactions on Image Processing*, pp. 1501-1511, 2018.
<https://doi.org/10.1109/TIP.2017.2754939>
- [36] C. Du, C. Du, L. Huang y H. He, "Reconstructing Perceived Images from Human Brain Activities with Bayesian Deep Multiview Learning" *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-14, 2018.
- [37] M. Mayo y E. Frank, "Experiments with Multi-view Multi-instance Learning for Supervised Image Classification", 2011.
- [38] C. Yan, B. Gong, Y. Wei y Y. Gao, "Deep Multi-View Enhancement Hashing for Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [39] J. Yu, J. Li, Z. Yu y Q. Huang, "Multimodal Transformer with Multi-View Visual Representation for Image Captioning", *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 4467-4480, 2020.
<https://doi.org/10.1109/TCSVT.2019.2947482>

- [40] Z. Xue, G. Li y Q. Huang, "Joint Multi-View Representation Learning and Image Tagging", de Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016.
<https://doi.org/10.1609/aaai.v30i1.1014>
- [41] J. Li, B. Zhang, G. Lu y D. Zhang, "Generative multi-view and multi-feature learning for classification", *Information Fusion*, pp. 215-226, 2019.
<https://doi.org/10.1016/j.inffus.2018.02.005>
- [42] Y. Liu, S. Sun y M. Yin, "Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis", de *Pacific Rim International Conference on Artificial Intelligence*, 2018.
https://doi.org/10.1007/978-3-319-97304-3_42
- [43] X. Dong, S. Xu, Y. Liu, A. Wang, M. I. Saripan, L. Li, X. Zhang y L. Lu, "Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation", *Cancer imaging*, 2020.
<https://doi.org/10.1186/s40644-020-00331-0>
- [44] M. Kan, S. Shan y X. Chen, "Multi-view Deep Network for Cross-view Classification", de 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
<https://doi.org/10.1109/CVPR.2016.524>
- [45] W. Wang, R. Arora, K. Livescu y J. Bilmes, "On Deep Multi-View Representation Learning: Objectives and Optimization", de *ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning*, 2015.
- [46] B. Neva Camacho, O. Najar Sánchez, "Comprensión de la lectura crítica con el Edublog Leanet", *Revista Vínculos*, vol. 16, no. 2, pp. 219–231, 2019.
<https://doi.org/10.14483/2322939X.15750>
- [47] Z. Zhu, P. Luo, X. Wang y X. Tang, "Multi-View Perceptron: A Deep Model for Learning Face Identity and View Representations", de *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2015.
- [48] M. T. Alam, V. Kumar y A. Kumar, "A Multi-view Convolutional Neural Network Approach for Image Data Classification", de 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021.
<https://doi.org/10.1109/ICCICT50803.2021.9509943>
- [49] X. Wu, Q.-G. Chen, Y. Hu, D. Wang y X. Chang, "Multi-View Multi-Label Learning with View-Specific Information Extraction", de *IJCAI'19: Proceedings of the 28th International Joint Conference*

on Artificial Intelligence, 2019.
<https://doi.org/10.24963/ijcai.2019/539>

<https://doi.org/10.1109/TIP.2021.3101917>

- [50] J. Wang, Y. Zheng, J. Song y S. Hou, “Cross-View Representation Learning for Multi-View Logo Classification with Information Bottleneck”, 2021.
<https://doi.org/10.1145/3474085.3475704>

- [51] A. Huang, Z. Wang, Y. Zheng, T. Zhao y C.-W. Lin, “Embedding Regularizer Learning for Multi-View Semi-Supervised Classification”, pp. 6997-7011, 2021.

- [52] E. Gómez Luna, D. Fernando Navas, G. Aponte Mayor y L. A. Betancourt Buitrago, “Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización”, *Dyna*, vol. 81, nº 184, pp. 158-163, 2014.
<https://doi.org/10.15446/dyna.v81n184.37066>