



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

**APRENDIZAJE MULTIETIQUETA DE PATRONES GEOMÉTRICOS EN  
OBJETOS DE HERENCIA CULTURAL**

TESIS PARA OPTAR AL GRADO DE  
MAGÍSTER EN CIENCIA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL EN COMPUTACIÓN

MATÍAS JESÚS VERGARA SILVA

PROFESOR GUÍA:  
BENJAMÍN BUSTOS CÁRDENAS

PROFESOR CO-GUÍA:  
IVÁN SIPIRÁN MENDOZA

MIEMBROS DE LA COMISIÓN:  
BENJAMÍN BUSTOS CÁRDENAS  
IVAN SIPIRÁN MENDOZA  
MARCELO MENDOZA ROCHA

SANTIAGO DE CHILE  
2022

RESUMEN DE LA TESIS PARA OPTAR  
AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS  
RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN  
POR: MATÍAS JESÚS VERGARA SILVA  
FECHA: 2022  
PROF. GUÍA: BENJAMÍN BUSTOS

## **APRENDIZAJE MULTIETIQUETA DE PATRONES GEOMÉTRICOS EN OBJETOS DE HERENCIA CULTURAL**

La presente investigación se enmarca en el estudio de objetos de herencia cultural, área de la Arqueología que guarda relación con la categorización y comprensión del patrimonio cultural tangible y que lleva por objetivo facilitar la *musealización* de este al mismo tiempo que se rescata información contextual del objeto y su cultura subyacente. En particular se aborda el etiquetado de patrones geométricos, proceso a través del cual las formas en la superficie de vasijas son asociadas a múltiples etiquetas que sintetizan sus características geométricas más relevantes.

Más específicamente, el problema que motiva esta investigación es la naturaleza tediosa de dicho proceso. En este sentido, la posibilidad de simplificar la tarea permitiría ahorrar tiempo experto, representando así una oportunidad de aportar valor a la Arqueología desde la Computación y en particular desde la Ciencia de Datos.

La solución propuesta es una herramienta de apoyo al etiquetado, la cual ante una imagen de un patrón geométrico responda con una serie de etiquetas a forma de sugerencia para el experto, reemplazando así parte de la carga relativa a idear etiquetas desde cero por la tarea de seleccionar sugerencias apropiadas.

Tal solución se lleva a cabo mediante técnicas de Aprendizaje Multietiqueta (MLC). En particular, se exploran técnicas tanto desde un enfoque de MLC Tradicional como desde un enfoque de *Extreme MLC* (XMLC), las cuales se conjugan además con distintas técnicas de tratamiento de datos y de generación de entradas sintéticas.

La investigación toma como entrada los patrones y etiquetas presentes en el libro *Ornamente Geometrischen Vasen: Ein Kompendium* [28], obra del arqueólogo Norbert Kunisch.

Los resultados experimentales muestran que los mejores métodos para tal propósito son *Binary Relevance* con *Logistic Regression* en el caso Tradicional y *Threshold Dependent Neural Network* en el caso XMLC. En ambos, el uso de múltiples técnicas para la generación de datos sintéticos y de pesos para tratamiento al desbalance habría resultado fundamental.

La herramienta final cuenta con propiedades tales como el predecir al menos 1 etiqueta correcta para el 95 % de los ejemplos, 3 para el 60 % y 5 para el 40 %, con una *Micro Precision* que varía entre un 0,85 y un 0,20 según la cantidad de etiquetas posibles, pero con un *Micro Recall* que se mantiene siempre por sobre 0,65.

Finalmente se propone como lineamiento para iteraciones futuras el comenzar por obtener retroalimentación usuaria de la herramienta actual, así como por el estudio de la capacidad de generalización de la misma ante conjuntos de datos de distinto origen.

*A mi padre en su lucha, a mi madre en su entrega.*

# Agradecimientos

A mi Padre Celestial, por todas las bendiciones que me han permitido llegar hasta acá.

A mi madre, Eliana, por enseñarme amor y disciplina. Por ser mi mejor refugio y también mi más severo juez, tu cariño y enseñanzas guían cada uno de mis pasos.

A mi padre, Cornelio, ejemplo de fe y valentía. Por demostrarle cada día, a través de pequeños gestos, que estás a mi lado ante cualquier desafío. Gracias por estar para mí.

A mi hermano, Eduardo, compañero y guardián. Por ser un amigo y guía en cada etapa, un abrazo en cada éxito y derrota. Gracias por tantos momentos bellos de la vida junto a ti.

A mi hermana, Evelyn, amiga y consejera. Por ofrecer siempre un oído atento y una conversación sincera, y por el valor que le das a mis palabras. Gracias por confiar en mí.

A Luna-Vittoria, mi compañera de alma. Por haber estado a mi lado desde el primer día, por tu amor incondicional y tu apoyo permanente. Gracias por escogerme cada día.

A mis sobrinos Mario, Benjamín, Florencia, Letizia, Renato y Leonor. Ser un ejemplo para ustedes es la mejor motivación para buscar ser mejor en la vida.

Al profesor Benjamín, quien sembrase en mí el interés por la computación durante mi primer año de estudios y me guiase hasta el final del camino, y al profesor Iván, quien me contagiase de pasión y entusiasmo en cada reunión. Gracias por creer en mí.

A Cristóbal, Diemen, Esteban, Geraldine y Rodrigo, mis grandes amistades. Por cada palabra de apoyo, por las noches en Discord y por haber crecido juntos estos 6 años. Gracias por su amistad.

Al equipo de *Technische Universität Graz*, en Austria, por facilitar los datos para la investigación. A los profesores Felipe Bravo y Juan Manuel Barrios, por proveer los equipos necesarios para los experimentos.

A Chimuelo, Carlota y Nadja. Por quedarse a mi lado cada vez que trasnoché, por subirle el volumen al ronroneo para que el cansancio no se escuchase. Este trabajo también es suyo.

*“La educación es la diferencia que existe entre desear poder ayudar a otras personas y el ser capaces de ayudarlas”.*

*Russell M. Nelson*

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Contexto</b>	<b>5</b>
2.1. Etiquetado de patrones en Arqueología . . . . .	5
2.2. Descripción de los datos a utilizar . . . . .	6
<b>3. Revisión de la Literatura</b>	<b>10</b>
3.1. Aprendizaje multietiqueta y sus desafíos . . . . .	10
3.2. Tipos de aprendizaje multietiqueta . . . . .	13
3.2.1. Esquema general . . . . .	13
3.2.2. <i>Online MLC</i> . . . . .	13
3.2.3. <i>MLC with Limited Supervision</i> . . . . .	15
3.2.4. <i>Traditional MLC</i> . . . . .	17
3.2.5. <i>Extreme MLC</i> . . . . .	19
3.3. Comentarios . . . . .	21
<b>4. Metodología</b>	<b>22</b>
4.1. Objetivos . . . . .	22
4.1.1. Objetivo general . . . . .	22
4.1.2. Objetivos específicos . . . . .	22
4.2. Supuestos . . . . .	23
4.3. Variables en estudio . . . . .	24
4.4. Selección de métodos . . . . .	26
4.4.1. Enfoque Tradicional . . . . .	26
4.4.2. Enfoque XMLC . . . . .	26
4.5. Experimentos a conducir . . . . .	27
4.5.1. Enfoque Tradicional . . . . .	27

4.5.2. Enfoque XMLC . . . . .	29
4.5.3. Incorporación de <i>K-Fold Cross Validation</i> . . . . .	30
4.6. Métricas de evaluación . . . . .	31
4.6.1. Métricas basadas en ejemplos . . . . .	31
4.6.2. Métricas basadas en etiquetas . . . . .	32
4.6.3. Relevancia e interpretación . . . . .	34
<b>5. Preparación de los datos</b>	<b>35</b>
5.1. Extracción y procesamiento de etiquetas . . . . .	35
5.1.1. Selección de idioma . . . . .	35
5.1.2. Normalización mediante NLP . . . . .	37
5.2. <i>Data Augmentation</i> . . . . .	39
5.2.1. Técnicas . . . . .	39
5.2.2. Conjuntos de datos . . . . .	43
<b>6. Enfoque Tradicional</b>	<b>45</b>
6.1. Técnicas . . . . .	45
6.1.1. Transformación del Problema . . . . .	45
6.1.2. Adaptación del Algoritmo . . . . .	47
6.2. Descriptores . . . . .	48
6.3. Experimentos . . . . .	50
6.3.1. Búsqueda de hiperparámetros . . . . .	50
6.3.2. E1: Evolución métodos sobre base, variando tratamiento al desbalance	51
6.3.3. E2: Evolución métodos sobre base, variando descriptores . . . . .	57
6.3.4. E3: Evolución mejores métodos, variando D.A. y descriptores . . . . .	59
<b>7. Enfoque XMLC</b>	<b>63</b>
7.1. Técnicas . . . . .	63
7.1.1. C2AE . . . . .	63
7.1.2. AlexNet TDNN . . . . .	69
7.1.3. CNN-RNN . . . . .	74
7.2. Descriptores . . . . .	81
7.3. Experimentos . . . . .	81
7.3.1. E4: C2AE, variando descriptores y conjunto de datos . . . . .	81
7.3.2. E5: AlexNet TDNN, variando tratamiento y conjunto . . . . .	83

7.3.3. E6: AlexNet TDNN, evolución en mejor conjunto . . . . .	85
7.3.4. E7: AlexNet TDNN, variando umbral de asignación . . . . .	87
7.3.5. E8: CNN-RNN, variando conjunto en tarea simplificada . . . . .	90
7.3.6. E9: CNN-RNN, evolución mejores conjuntos . . . . .	93
<b>8. Análisis de Resultados</b>	<b>95</b>
8.1. Comparación de enfoques . . . . .	95
8.1.1. Comparación según métricas basadas en etiquetas . . . . .	96
8.1.2. Comparación según métricas basadas en ejemplos . . . . .	98
8.2. Formulación de la herramienta de apoyo . . . . .	103
<b>9. Conclusiones</b>	<b>105</b>
9.1. Trabajo Futuro . . . . .	107
<b>Bibliografía</b>	<b>109</b>
<b>Anexos</b>	<b>116</b>
<b>Anexo A. Experimento 1</b>	<b>116</b>
<b>Anexo B. Experimento 2</b>	<b>132</b>
<b>Anexo C. Experimento 3</b>	<b>140</b>
<b>Anexo D. Experimento 4</b>	<b>144</b>
<b>Anexo E. Experimento 5</b>	<b>157</b>
<b>Anexo F. Experimento 6</b>	<b>159</b>
<b>Anexo G. Experimento 7</b>	<b>161</b>
<b>Anexo H. Experimento 9</b>	<b>163</b>

# Índice de Tablas

1.1.	Etiquetas según Kunisch para los patrones en las Figuras 1.1a y 1.1b . . . . .	2
2.1.	Cantidad de patrones por capítulo. . . . .	7
4.1.	Cantidades de etiquetas en estudio y umbral de frecuencia asociado . . . . .	28
5.1.	Etiquetas del patrón 17e, pág. 41 [28]. . . . .	35
5.2.	Etiquetas del patrón 7b, pág. 17 [28]. . . . .	36
5.3.	Etiquetas del patrón 7c, pág. 17 [28]. . . . .	36
5.4.	Etiquetas del patrón 1b, pág. 3 [28]. . . . .	37
5.5.	Etiquetas del patrón 31d, pág. 77 [28]. . . . .	38
5.6.	Ejemplos de etiquetas y su procesamiento . . . . .	38
5.7.	Cantidad de imágenes para entrenamiento por <i>fold</i> según conjunto de datos .	44
6.1.	<i>Accuracy</i> Arquitecturas ResNet variando técnica de D.A. . . . .	50
6.2.	Búsqueda de hiperparámetros . . . . .	51
6.3.	<i>Micro F<sub>2</sub>-Score</i> promedio para métodos de transformación del problema, según tratamiento al desbalance. . . . .	54
6.4.	<i>Micro F<sub>2</sub>-Score</i> promedio para métodos de adaptación del problema. . . . .	54
6.5.	<i>Micro F<sub>2</sub>-Score</i> promedio para 3 mejores métodos, según conjunto de descriptores. . . . .	58
6.6.	AUC de <i>Micro F<sub>2</sub>-Score</i> para 3 mejores métodos, variando conjunto de descriptores y técnica de D.A. . . . .	61
6.7.	Comparación mejores combinaciones Enfoque Tradicional . . . . .	61
7.1.	Detalle arquitectura AlexNet . . . . .	73
7.2.	Valores <i>Micro F<sub>2</sub>-Score</i> Experimento 7 . . . . .	88
7.3.	AUC <i>Micro F<sub>2</sub>-Score</i> Experimento 7 . . . . .	90
8.1.	AUC <i>Micro F<sub>2</sub>-Score</i> mejor método de cada enfoque . . . . .	96

8.2. AUC <i>Hamming Score</i> mejor método de cada enfoque . . . . .	99
8.3. AUC <i>K-Match Ratio</i> mejor método de cada enfoque . . . . .	102
8.4. Ejemplos de sugerencias de la herramienta con un universo de las 107 etiquetas más frecuentes. . . . .	104
A.1. Métricas Experimento 1 - Sin tratamiento al desbalance. . . . .	116
A.2. Métricas Experimento 1 - Con tratamiento al desbalance. . . . .	123
B.1. Métricas Experimento 2 - Descriptores ResNet50. . . . .	132
C.1. AUC por métrica - Experimento 3. . . . .	140
D.1. Métricas Experimento 4 - Descriptores AlexNet. . . . .	144
D.2. Métricas Experimento 4 - Descriptores Imágenes <i>flatten</i> . . . . .	150
E.1. Métricas Experimento 5. . . . .	157
F.1. Métricas Experimento 6. . . . .	159
G.1. Métricas Experimento 7. . . . .	161
H.1. Métricas Experimento 9. . . . .	163

# Índice de Ilustraciones

1.1. Ejemplos de patrones geométricos . . . . .	2
2.1. Sistema de etiquetado de estilos de vasijas mediante gamificación [29] . . . . .	6
2.2. Ejemplares por capítulo. . . . .	7
2.3. Cantidad de etiquetas y/s Frecuencia. . . . .	9
3.1. Ejemplo de clasificación binaria, multiclase y multietiqueta. . . . .	11
3.2. Tipos de MLC. . . . .	14
5.1. Patrón 15b, aumentado con Rotación . . . . .	40
5.2. Patrón 6e, aumentado con Reflexión . . . . .	40
5.3. Patrón 95l, aumentado con Lluvia . . . . .	41
5.4. Patrón 21f, aumentado con Distorsión elástica . . . . .	41
5.5. Patrón 44a, aumentado con Recorte . . . . .	42
5.6. Patrón 70m, aumentado con Difuminado . . . . .	42
5.7. Patrón 15b, aumentado con <i>RandAugment</i> . . . . .	43
6.1. Arquitecturas ResNet18 y ResNet50 en diagramas de bloques . . . . .	49
6.2. <i>Micro F<sub>2</sub>-Score</i> métodos sobre base, variando tratamiento del desbalance . . . . .	52
6.3. <i>Micro Precision</i> métodos sobre base, variando tratamiento al desbalance . . . . .	55
6.4. <i>Micro Recall</i> métodos sobre base, variando tratamiento al desbalance . . . . .	56
6.5. <i>Micro F<sub>2</sub>-Score</i> métodos sobre base, variando conjunto de descriptores . . . . .	57
6.6. Evolución <i>Micro F<sub>2</sub>-Score</i> para 3 mejores métodos en el espectro de etiquetas, variando conjunto de descriptores y técnica de D.A. . . . .	60
7.1. Arquitectura DCCA . . . . .	64
7.2. Ejemplo de arquitectura <i>autoencoder</i> . . . . .	65
7.3. Arquitectura C2AE . . . . .	67

7.4.	Configuración C2AE . . . . .	69
7.5.	Ejemplo de convolución . . . . .	70
7.6.	Ejemplo de <i>Max Pooling</i> en $t = 0$ . . . . .	71
7.7.	Ejemplo de conexión entre capas convolucionales y <i>fully-connected</i> . . . . .	71
7.8.	Diagrama de bloques AlexNet . . . . .	72
7.9.	Representación abstracta de una RNN. . . . .	75
7.10.	Ejemplo de <i>Beam Search</i> con $w = 2$ . . . . .	77
7.11.	Esquemas CNN-RNN. . . . .	79
7.12.	<i>Micro F<sub>2</sub>-Score C2AE</i> , variando conjunto de descriptores. . . . .	82
7.13.	<i>Micro F<sub>2</sub>-Score</i> en 107 etiquetas, variando tratamiento al desbalance, según conjunto de datos. . . . .	84
7.14.	<i>Micro F<sub>2</sub>-Score</i> , múltiples conjuntos . . . . .	86
7.15.	Múltiples métricas, conjunto 13 . . . . .	87
7.16.	<i>Micro F<sub>2</sub>-Score</i> conjunto 13 variando umbral de asignación. . . . .	88
7.17.	<i>Micro Precision</i> y <i>Micro Recall</i> conjunto 13 variando umbral de asignación. . . . .	89
7.18.	<i>Micro F<sub>2</sub>-Score</i> , <i>Recall</i> , <i>Precision</i> y <i>Hamming Score</i> por conjunto en 131 etiquetas, Experimento 8 . . . . .	91
7.19.	Primeras 35 matrices de confusión en 107 etiquetas, obtenidas a partir del <i>fold</i> 1, Experimento 8 . . . . .	92
7.20.	<i>Micro F<sub>2</sub>-Score</i> , múltiples conjuntos . . . . .	93
8.1.	<i>Micro F<sub>2</sub>-Score</i> . . . . .	96
8.2.	Comparación entre enfoques según métricas basadas en etiquetas . . . . .	97
8.3.	Comparación entre enfoques, métricas por ejemplo . . . . .	98
8.4.	Selección de matrices de confusión para etiquetas con al menos 15 eventos, según enfoque. . . . .	100
8.5.	Selección de matrices de confusión para etiquetas con a lo más 2 eventos, según enfoque. . . . .	100
8.6.	<i>Comparación entre enfoques, 1-Match Ratio</i> . . . . .	101
8.7.	Comparación entre enfoques, <i>K-Match Ratio</i> para $K \in [3, 5]$ . . . . .	102

# Capítulo 1

## Introducción

Definida como aquel patrimonio de bienes tangibles e intangibles de un grupo o sociedad que se hereda de generaciones pasadas [31], la herencia cultural representa un área de estudio presente en ciencias tales como la arqueología, paleontología y antropología, y la cual ha cobrado especial relevancia en las últimas décadas gracias a la conformación de acuerdos internacionales relativos a su reconocimiento y conservación [35][36].

Existe además un término acuñado especialmente para referirse a los bienes tangibles. Se trata de los objetos de herencia cultural, definidos como cualquier bien mueble de importancia cultural que requiera protección - trátese de artefactos arqueológicos, especímenes paleontológicos/geológicos, meteoritos o cualquier otro objeto con un significado cultural [34] -. Dichos bienes son importante de conservación y catalogación, en cuanto su estudio permite alcanzar una mayor comprensión del contexto que les subyace.

Tales objetos se hacen particularmente presentes en la arqueología, ciencia que busca estudiar la diversidad humana a través del registro material y que, como consecuencia, da lugar a numerosas excavaciones e incursiones en búsqueda de vestigios de culturas pasadas, encontrándose así con múltiples objetos de herencia cultural los cuales son extraídos, descritos y, una vez rescatada toda la información posible, puestos a disposición de instituciones de conservación.

Durante todo el proceso anterior, la etapa de descripción es probablemente la más compleja, en cuanto involucra múltiples tareas según el tipo de objeto, su origen y las características en estudio (la época a la que perteneció, qué propósito tenía, el significado de sus formas y figuras, entre otras). Esto requiere a menudo no solo de un amplio saber en arqueología, si no también de una basta experiencia con el escenario particular en el que se encuentre el objeto y de conocimiento previo sobre las posibles culturas relacionadas. Un ejemplo cercano de dicha complejidad se encuentra en los múltiples esfuerzos realizados por construir una tipología y seriación de la cerámica de la cultura Chimú [7], estado andino que se estableciese en el actual Perú entre los siglos X y XV y para el cual la catalogación de su cerámica es aún material de investigación.

Existe sin embargo una tarea transversal a múltiples escenarios, aunque particularmente desafiante. Se trata del etiquetado de patrones geométricos, proceso realizado por expertos en el cual se analiza la superficie de objetos de herencia cultural (con frecuencia vasijas y/o murales) en búsqueda de patrones geométricos para posteriormente asociar, a cada uno

de ellos, un conjunto de “etiquetas”: sustantivos y adjetivos que responden a las distintas propiedades geométricas del mismo. Ejemplos de ello se presentan a través de la Tabla 1.1, la cual incorpora las etiquetas asignadas por Norbert Kunisch, arqueólogo de renombre mundial, a los patrones reproducidos en las Figuras 1.1a y 1.1b.

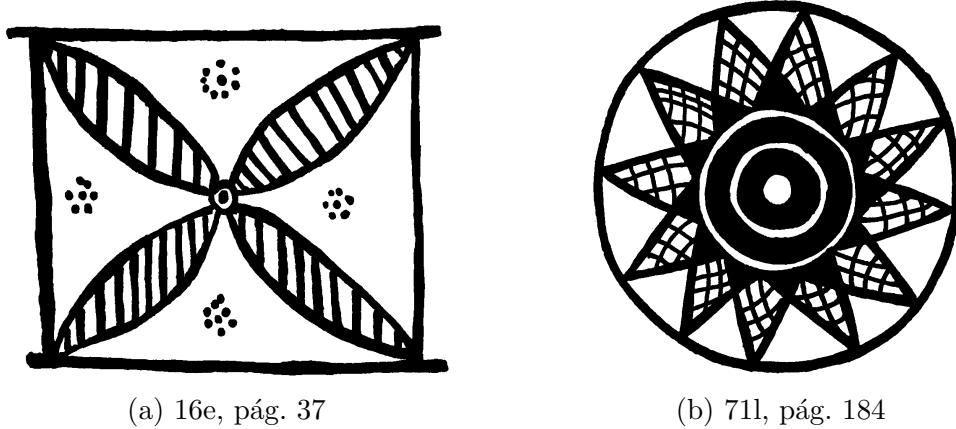


Figura 1.1: Ejemplos de patrones geométricos. **Observación:** los patrones presentados corresponden a la reproducción manual de bocetos realizados por Norbert Kunisch en su obra *Ornamente Geometrischen Vasen: Ein Kompendium* [28]. Las leyendas individuales incluyen el índice y la página correspondiente a cada patrón en el libro.

Tabla 1.1: Etiquetas según Kunisch para los patrones en las Figuras 1.1a y 1.1b.

**Observación:** las referencias corresponden al identificador y página de las etiquetas en el libro *Ornamente Geometrischen Vasen: Ein Kompendium* [28], obra de Norbert Kunisch.

Figura	Etiquetas	Referencia
1.1a	Quatrefoil, hatched, with central dotted circle, metopal	Patrón 16e, pág. 37
1.1b	Circle, enclosing a star of 12 points, cross-hatched, floor ornament	Patrón 71l, pág. 184

Es en dicho proceso y en la dificultad del mismo donde la presente investigación encuentra su motivación. Más en detalle, el problema que se busca resolver consiste en la inexistencia de una herramienta de apoyo al etiquetado, la cual simplifique la tarea para el experto. Las razones por las cuales esto representa una necesidad son, por un lado, lo tediosa que la tarea resulta dado el amplio universo de etiquetas posibles y la consecuente necesidad de invertir una gran cantidad de horas hombre en su desarrollo y, por otro lado, la falta de acuerdo entre expertos con respecto a qué etiquetas utilizar y en qué ocasiones. Ambos aspectos se detallarán a continuación.

En primer lugar, la complejidad del proceso se debe principalmente a la gran cantidad de información geométrica que un mismo patrón puede contener, lo cual se traduce en un trabajo extensivo y eventualmente incompleto, en donde los expertos pueden invertir horas tratando de asignar todas las etiquetas necesarias y aún así no lograr reflejar todas las características

importantes del patrón, dejando así información sin ser rescatada y abriendo la posibilidad a una catalogación errónea del objeto en estudio.

En segundo lugar, la falta de acuerdo entre etiquetadores encuentra sus orígenes en una diferencia de trasfondo, lo cual a menudo da lugar a variaciones en la consistencia de los datos [29]. Por ejemplo, un patrón podría recibir las etiquetas *ave*, *vertical* y *metopa* por un experto *A*, mientras que un experto *B* podría asignarle las etiquetas *pájaro* y *vertical*, evidenciando así dos problemas: el uso de sinónimos para referirse a la misma característica (*ave* y *pájaro*), y la ausencia de la etiqueta *metopa* en el etiquetado de *B*.

Ahora bien, existe además una tercera razón por la cual el desarrollo de una herramienta de apoyo al etiquetado de patrones geométricos resulta de interés. Se trata de que, si bien en esta investigación el problema se aborda mediante ejemplares provenientes de Grecia (como se verá en secciones posteriores), dicho proceso no es ajeno a nuestra región, donde las múltiples culturas que han habitado la zona durante el pasar de los siglos han dado lugar a diversas fuentes de objetos con patrones geométricos en su superficie - tales como los descritos por Paola Gonzalez en *Arte y Cultura Diaguita Chilena: Simetría, Simbolismo e Identidad* [21] -, permitiendo así que el eventual desarrollo represente también una contribución a la arqueología chilena y andina.

Ante las razones descritas surge entonces la pregunta de por qué realizar un etiquetado correcto resulta tan importante. La respuesta se encuentra en el hecho de que el etiquetado refleja una clasificación tipológica en la cual la cantidad de objetos individuales con un cierto contexto arqueológico resulta esencial para deducir aspectos socioculturales, económicos o conductuales de dichos contextos [5], por lo cual un etiquetado completo y sin ambigüedad es más indicativo del significado que el patrón en estudio pudo tener en el pasado, así como de la utilidad del objeto que lo contiene. Ambos resultados (significado del patrón y utilidad del objeto) contribuyen finalmente a aumentar el entendimiento de la cultura subyacente, facilitando además la *musealización* del objeto.

En este sentido, lo que se busca es desarrollar una herramienta que sirva de apoyo al experto mediante la sugerencia de etiquetas relevantes con nombres estandarizados, atacando así los dos problemas detectados: la complejidad del problema, al reducir la cantidad de etiquetas que el experto debe idear por sí solo, y la falta de acuerdo, al sugerir etiquetas de una misma taxonomía.

Dicho desafío es abordado en esta investigación a través del Aprendizaje multietiqueta, área del *Machine Learning* (ML) que ha presentado un importante crecimiento en los últimos años [30]. A fin de introducir dicho enfoque, resulta primero necesario describir brevemente lo que es la tarea de *clasificación en aprendizaje supervisado*.

Una necesidad frecuente en ML es el de una herramienta que, al recibir cierta entrada, responda asignando una clase  $\ell$  dentro de un conjunto  $L$  de clases tal que  $|L| > 1$ . Esta tarea recibe el nombre de “*clasificación*” y acostumbra ser aproximada desde un enfoque de aprendizaje supervisado, técnica en la cual se entrenaan algoritmos y modelos computacionales para aprender a entregar la clase correcta a partir de observar un conjunto de datos cuya clase es conocida y posteriormente usar dicha información para intentar predecir la clase de ejemplos nuevos, en un proceso de prueba y error. Existen múltiples variaciones de dicha tarea, siendo el ejemplo quizás más clásico aquel en donde  $|L| = 2$ , caso que recibe el nombre de *clasificación binaria*. Un caso más complejo es el de la *clasificación multiclasa*, que se halla

cuando  $|L| > 2$ . En ambos casos, una característica se mantiene invariante: cada ejemplo pertenece y es asignado a una única clase, tratándose así de clasificación *singular* [46].

Existe, sin embargo, una familia de problemas hermanos a la clasificación singular, menos explorados y ampliamente desafiantes: se trata de la clasificación *multietiqueta*, donde en lugar de asociar cada entrada a una clase  $\ell \in L$ , se asocia a un conjunto de clases  $Y \subseteq L$ . Un ejemplo de ello sería la clasificación de enfermedades a partir de diagnósticos médicos, donde un mismo examen podría revelar que un paciente sufre de diabetes e hipercolesterolemia, al mismo tiempo [46]. Otro ejemplo se encuentra en la genética, donde un mismo gen puede ser responsable de codificar múltiples características de una especie.

Es a través de este tipo de clasificación multietiqueta, más desafiante pero también más cercana al mundo real, que la presente investigación busca dar lugar a una herramienta de apoyo al etiquetado de patrones, tras identificar la evidente relación entre el proceso de asignar múltiples etiquetas a un patrón y el aprendizaje multietiqueta. Se pretende que, tras una selección y tratamiento metódico de un conjunto reducido de datos, se logre llegar a entradas que permitan, a través de las bondades de la programación, entrenar distintos algoritmos y/o modelos de clasificación multietiqueta para patrones geométricos, logrando así - en un panorama optimista - la automatización del etiquetado.

Ahora bien, existen dos variables cuyo comportamiento *a priori* es desconocido y que podrían incidir ampliamente en el nivel de automatización alcanzado. La primera de ellas guarda relación con la factibilidad de obtener un conjunto de datos lo suficientemente grande y representativo, variable la cual está restringida para esta investigación a una única fuente de datos (que se discutirá más adelante). La otra variable guarda relación con la existencia de una taxonomía real en las etiquetas, pues son este tipo de características - agrupaciones ordenadas, construidas basándose en relaciones "naturales" - las más factibles de aprender mediante el enfoque propuesto.

Existe, sin embargo, una razón adicional para llevar a cabo la investigación aún sin tener certeza de dichas variables. Se trata de la inexistencia a la fecha de una solución similar, lo cual tiene por efecto el que no sea necesario alcanzar las sugerencias perfectas o el automatizado total del proceso para que signifique un aporte significativo al área, sino que, por el contrario, el solo hecho de desarrollar una herramienta de apoyo - que sugiera un conjunto de etiquetas a evaluar y/o corregir por el experto - significaría de por sí un avance enorme.

De esta manera, la meta del trabajo a realizar se define como el concebir una herramienta de apoyo al etiquetado de patrones geométricos, la cual actúe como un *sistema recomendador* que permita hacer del etiquetado un proceso más eficiente y eficaz mediante la sugerencia de etiquetas razonables al experto y esperando que ello siente las bases para, en un futuro, poner miras en la automatización total del proceso.

# Capítulo 2

## Contexto

### 2.1. Etiquetado de patrones en Arqueología

En la actualidad, el proceso del etiquetado de patrones se lleva a cabo de forma completamente manual: el arqueólogo accede a un objeto sin etiquetar (generalmente a partir de excavaciones), lo estudia en búsqueda de patrones interesantes y, de encontrarlos, realiza un *sketch* de ellos para posteriormente dedicarse a la asignación de etiquetas, lo cual puede tomar desde minutos hasta horas dependiendo de la complejidad de los patrones y la motivación que el objeto signifique para el etiquetador. En todo el transcurso no existe ninguna herramienta de apoyo al experto.

Sin embargo, existen otros procesos propios de la arqueología donde la computación ha logrado contribuir con herramientas de apoyo, especialmente en lo que a computación gráfica y el análisis de formas tridimensionales se refiere - área que ha presentado una simbiosis histórica con el dominio de los objetos culturales [42] -. Ejemplos de ello se encuentran en la digitalización 3D, la reconstrucción y restauración visual de objetos, el desarrollo de interfaces inmersivas, entre otros.

Un ejemplo quizá más cercano al tema de esta propuesta se encuentra en el etiquetado de estilos de vasijas de herencia cultural, donde se desarrollase una herramienta que buscaría, a partir de un enfoque de gamificación, reducir la probabilidad de observar vasijas con etiquetados incompletos al aumentar la motivación del experto mediante un juego. Si bien dicha herramienta habría logrado cumplir su objetivo, el estudio de sus resultados indicaría que también habría aumentado el nivel de desacuerdo entre expertos [29]. La Figura 2.1 muestra una captura de dicho sistema.

Es así como, pese a lo inexplorado del etiquetado de patrones geométricos desde un enfoque de aprendizaje de máquinas, el potencial que guarda incorporar herramientas computacionales en los procesos arqueológicos es bien conocido para el área, así como lo es también el problema del etiquetado incompleto o inconsistente, el cual resulta además transversal a los múltiples catálogos posibles (de estilo, de forma, de patrones, entre otros).

En el caso particular del etiquetado de patrones geométricos, dicho problema ha servido de motivación para que los expertos busquen, a través de su obra, sentar una taxonomía base sobre la cual sus colegas puedan guiarse. Uno de los intentos más famosos es el libro *Orna-*



Figura 2.1: Sistema de etiquetado de estilos de vasijas mediante gamificación [29]

mente *Geometrischen Vasen: Ein Kompendium* [28], autoría del Prof. Dr. Norbert Kunisch, arqueólogo de renombre mundial que buscó establecer una taxonomía de etiquetas a través de ejemplos provenientes de excavaciones en Grecia durante el siglo XX. Dicho libro lograría reconocimiento en el área, más no habría alcanzado a establecerse como una referencia del etiquetado. Una posible razón para ello se encuentra en la misma naturaleza tediosa del proceso, donde el incorporar una tarea adicional - tal como la revisión de la obra de un tercero a modo de guía - no hace más que disminuir la motivación del experto por alcanzar un trabajo completo [29].

De esta manera, la investigación se plantea como una innovación sin precedentes para la el etiquetado de patrones geométricos, pero con el amplio respaldo de aplicaciones similares en otras áreas de la arqueología. El problema que abarca, por otro lado, se muestra como una dificultad conocida y con un historial de intentos de solución sin éxito, lo cual refleja a la vez la importancia de encontrar una solución y la alta complejidad del mismo.

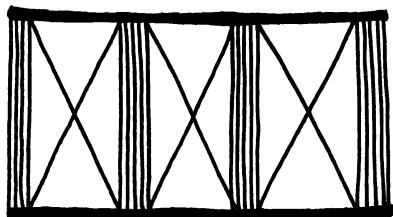
## 2.2. Descripción de los datos a utilizar

La investigación encuentra sus entradas en el ya mencionado libro *Ornamente Geometrischen Vasen: Ein Konpendium* del Dr. Norbert Kunisch. Dicho libro contiene un total de 776 patrones etiquetados en cinco idiomas (inglés, francés, italiano, griego y alemán) y clasificados en 6 capítulos según la forma predominante, los cuales se presentan en la Tabla 2.1. La Figura 2.2, por otro lado, muestra un ejemplar de patrón por cada uno.

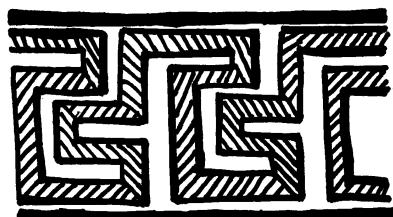
Una característica a destacar en este punto es que, si bien los capítulos buscan clasificar los patrones según su motivo principal, esta clasificación no es excluyente. Un buen ejemplo de ello se encuentra en la Figura 2.2e, que si bien presenta una forma global circular que le otorga su pertenencia al capítulo *Circular ornaments*, también presenta en su interior motivos triangulares muy similares a los de *Triangular ornaments* y mallas de líneas propias del capítulo *Strokes and lines*.

Tabla 2.1: Cantidad de patrones por capítulo.

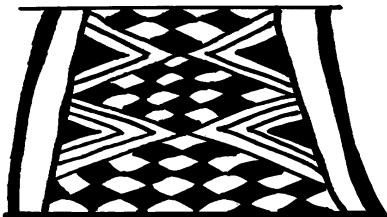
Capítulo	Patrones
Strokes and lines	70
Rectangular ornaments	179
Lozenge	101
Triangular ornaments	174
Circular ornaments	190
Pictographics	62



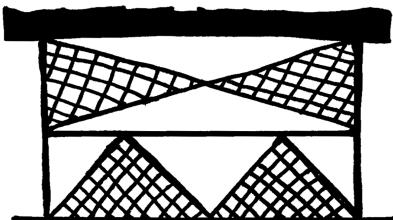
(a) *Strokes and lines*, patrón 3h (pág. 9).



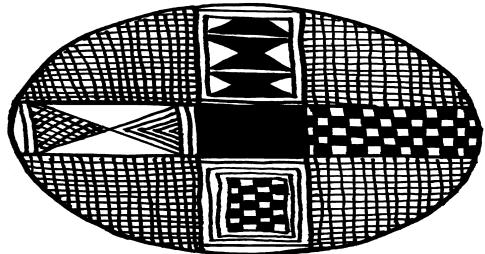
(b) *Rectangular ornaments*, patrón 24d (pág. 59).



(c) *Lozenge*, patrón 42c (pág. 106).



(d) *Triangular ornaments*, patrón 56c (pág. 141).



(e) *Circular ornaments*, patrón 79a (pág. 205).



(f) *Pictographics*, patrón 95j.

Figura 2.2: Ejemplares por capítulo. **Observación:** los patrones presentados corresponden a la reproducción manual de bocetos realizados por Norbert Kunisch en su obra *Ornamente Geometrischen Vasen: Ein Kompendium* [28]. Las leyendas individuales incluyen el índice y la página correspondiente a cada patrón en la obra del autor.

En lo que respecta a las etiquetas, por otro lado, se tiene un total de 586 entradas distintas, las cuales se presentan a una tasa promedio de 3 asignaciones por patrón y dan lugar a 595 combinaciones diferentes, entendiendo por combinación la presencia simultánea de múltiples etiquetas en un mismo patrón. Estos números levantan una alerta importante, en cuanto se tienen cantidades similares de patrones (que serán los puntos de entrenamiento) y de etiquetas a predecir, una situación en general desfavorable para cualquier problema de aprendizaje en ML. A fin de cuantificar este fenómeno, conviene introducir el concepto de *Label Density*.

*Label Density* (en adelante LD) corresponde a una métrica ampliamente utilizada para la descripción de conjuntos de datos multietiqueta, la cual busca responder a la pregunta “*¿Qué tan multietiqueta es un dataset?*” [45] mediante un índice que toma un valor 1 en aquel caso en que todos los patrones interactúan con todas las etiquetas y un valor 0 cuando ningún patrón tiene etiquetas asignadas. Su formulación matemática se presenta en la Ecuación 2.1, donde  $D$  representa al conjunto de ejemplos (patrones en este caso),  $Y_i$  al conjunto de etiquetas reales para un ejemplo  $i$  y  $L$  al total de etiquetas en los datos.

$$\text{Label Density} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i|}{|L|} \quad (2.1)$$

De esta manera, LD captura la cantidad promedio de etiquetas por patrón sobre la cantidad total de etiquetas, entregando así una información que resulta determinante para la dificultad del problema [6]: qué tanto desbalance entre etiquetas positivas y negativas hay en los datos, entendiendo por positivo el caso donde la asignación de la etiqueta para un ejemplo es correcta y por negativo cuando una etiqueta no corresponde.

Al aplicar este concepto a los datos en estudio se obtiene un LD de 0,0051, valor extremadamente bajo y que refleja entonces un enorme desbalance a favor de los casos negativos, situación que encuentra sus razones en una propiedad particular del proceso descriptivo: la existencia de muchas etiquetas “raras” con baja frecuencia, lo cual da lugar a una distribución de cola muy alargada, la cual se presenta en la Figura 2.3.

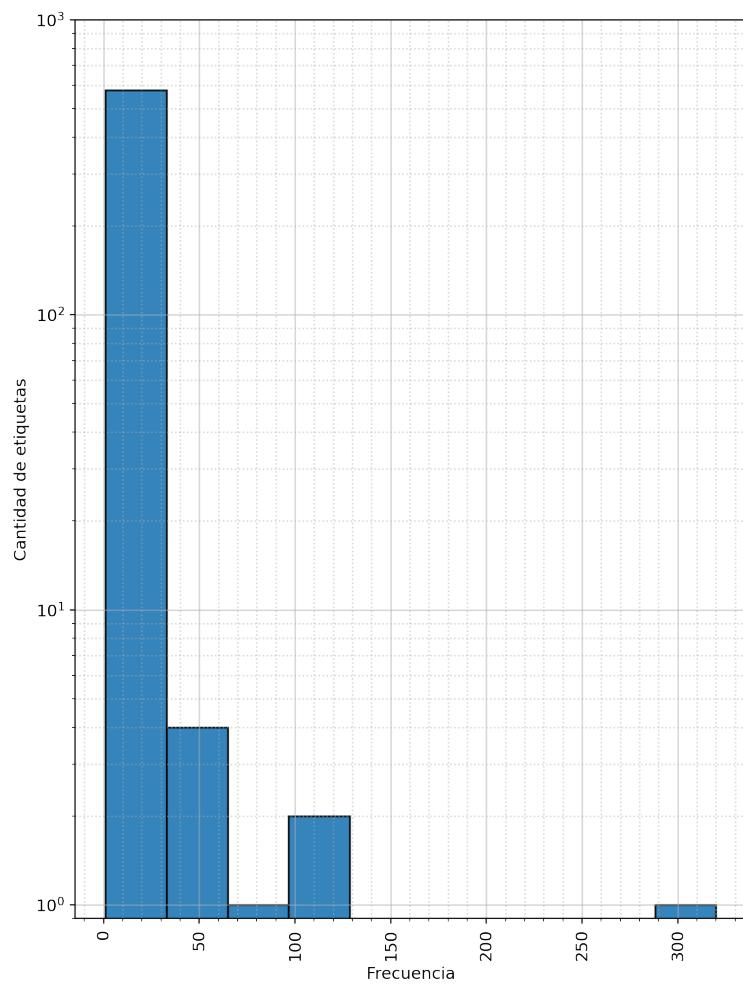


Figura 2.3: Cantidad de etiquetas v/s Frecuencia.

Dicho fenómeno, calificado por la literatura como un indicador de un proceso de aprendizaje más complejo [3], servirá entonces como motivación para realizar los experimentos no solo en un acercamiento de método contra método, sino también evaluando los impactos de reducir la cantidad de etiquetas en estudio según su frecuencia en los datos, esperando que esto contribuya a obtener mejores resultados y que no resulte necesaria una reducción excesiva para obtener buenas predicciones, en cuanto tal situación podría llevar a una trivialización del problema y en consecuencia a una solución sin valor real para el área.

# Capítulo 3

## Revisión de la Literatura

En este capítulo se presentan los conceptos y técnicas más fundamentales del aprendizaje multietiqueta. Dicho estudio se gesta en dos secciones: una primera en la cual se introduce el aprendizaje multietiqueta y sus principales desafíos (Sección 3.1) y una segunda dedicada a los distintos tipos de aprendizaje multietiqueta, sus aplicaciones y técnicas más recientes (3.2). Finalmente, se incorpora una breve sección de comentarios (3.3) en la cual se busca posicionar el problema en estudio dentro de las categorías presentadas.

### 3.1. Aprendizaje multietiqueta y sus desafíos

Resulta conveniente comenzar el estudio aclarando conceptos para dos términos a menudo usados de forma indiferente en la literatura: aprendizaje multietiqueta (o MLL, por sus siglas en inglés) y clasificación multietiqueta (MLC). Mientras el segundo es una aplicación del primero - caso en que a cada ejemplo se le asigna un subconjunto de etiquetas  $Y \subset L$  -, este último puede referirse también al ranking de etiquetas (LR), tarea hermana al MLC en la cual se predice el conjunto  $L$  completo pero en un determinado orden según relevancia. En esta investigación se adopta la convención más tradicional, es decir, utilizar de forma equivalente los términos aprendizaje y clasificación, en cuanto será la clasificación la tarea a través de la cual se evaluará el aprendizaje.

En la actualidad, el aprendizaje multietiqueta representa un área de investigación en rápido crecimiento gracias a dos grandes razones: por un lado, la increíble velocidad con la cual los seres humanos generamos datos - estimada a millones de exabytes por día [30] - y, por el otro, el valor que el abarcar dichos datos desde un enfoque de aprendizaje multietiqueta puede aportar, en cuanto responde de manera precisa a problemas tales como la clasificación de funciones de proteínas, la categorización de música y/o películas, la descripción de escenas en imágenes de redes sociales, entre otras tareas donde la salida o “clase” no es necesariamente única, si no más bien múltiple [30].

En muchos casos, dicho tipo de clasificación resulta incluso más natural que la clasificación singular. Por ejemplo, una foto de animales domésticos - ejemplo por excelencia de algoritmos de clasificación singular - podría no solo incluir o no un *perro* (caso binario) o una opción entre *gato* y *perro* (caso multiclasa), sino también a ambos al mismo tiempo, y también a una *planta* y/o un *bus*. Naturalmente, extraer toda esta información resultaría mucho más útil

que solo indicar la presencia o dominancia de una especie en la imagen, escenario que se ve ilustrado en la Figura 3.1 con una imagen de un gato y una planta. Nótese que, para el caso binario y multiclase, la salida (recuadro en verde) es única y corresponde a la clase con mayor probabilidad, mientras que para el caso multietiqueta la salida es múltiple (*Gato* y *Planta*) y corresponde a las clases con probabilidad mayor a cierto umbral (0,50, por ejemplo).

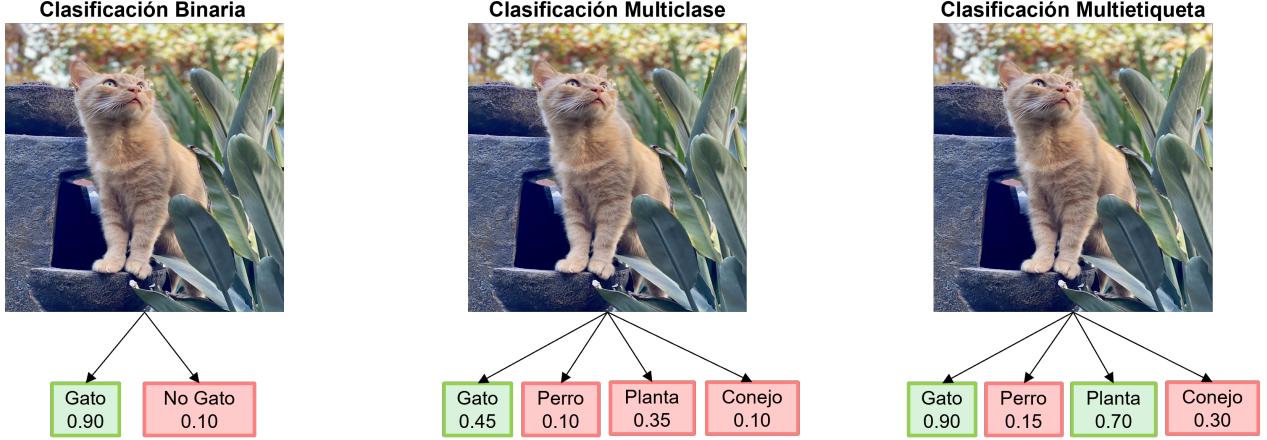


Figura 3.1: Ejemplo de clasificación binaria, multiclase y multietiqueta.

Dichas ventajas de la clasificación multietiqueta vienen, sin embargo, acompañadas de nuevos desafíos. Aún más, al tratarse de una tarea que contiene a la clasificación singular, esta mantiene sus dificultades e incorpora otras adicionales. Entre ellas:

- Se mantiene el compromiso eficiencia/rendimiento, en cuanto se requiere de clasificadores con alta capacidad de generalización pero que no requieran de cantidades excesivas de tiempo y/o espacio para obtener resultados, especialmente en cuanto la etapa de inferencia se refiere. En este sentido, se incorpora además la necesidad de métodos que no escalen con el número de etiquetas y que sean capaces de aprender con pocos ejemplos por cada etiqueta (en un contexto de *Few-Shot Learning* [53]), explotando además las eventuales relaciones que puedan presentarse entre ellas.
- Se mantienen los desafíos relacionados a la dificultad de obtener datos correctamente etiquetados, lo cual se ve además amplificado por la cantidad de etiquetas en estudio y la frecuente necesidad de conocimiento experto para su asignación.
- Se mantiene la necesidad de tratar cuidadosamente el desbalance de clases (o de etiquetas, en este caso) a fin de obtener un aprendizaje óptimo. En este sentido, se agrega también una inherente dificultad en cuanto la pertenencia de cada ejemplo a múltiples clases hace que las técnicas de balanceo clásicas - tales como el muestreo estratificado - pierdan sentido al no poder tratar con una única clase a la vez.
- Se incorpora un desafío usualmente resuelto para la clasificación singular: la evaluación de resultados, cuya dificultad radica en el hecho de que las métricas para clasificación singular dan lugar frecuentemente a evaluaciones demasiado estrictas en el caso múltiple.

- Se incorpora el ya introducido concepto de densidad del etiquetado (LD), en cuanto ya no basta con que un conjunto de datos incorpore muchos puntos de entrenamiento para ser considerado apto para el aprendizaje si no que también se requiere un cierto nivel de presencia de cada etiqueta en los datos.

Son estas nuevas dificultades las que han servido de motivación para la definición de técnicas especialmente diseñadas para el aprendizaje multietiqueta. En el caso del desbalance, por ejemplo, la no aplicabilidad de las técnicas tradicionales ha llevado al planteamiento de sistemas de asignación de pesos para el aprendizaje y al desarrollo de funciones de pérdida específicas al MLC [24].

En el caso de las métricas, por otro lado, la necesidad de sistemas de evaluación apropiados ha propiciado la definición de nuevas métricas, las cuales en general se categorizan en dos grandes grupos: aquellas basadas en ejemplos [17][18] y aquellas basadas en etiquetas [47]. Al respecto:

- Las métricas basadas en ejemplos, por un lado, trabajan evaluando la eficacia del sistema de aprendizaje en cada entrada por separado, y luego retornando el promedio sobre el conjunto de entradas. Ejemplos de este tipo de métricas se encuentran en el *Hamming Score* [13] y la versión promediada por ejemplos de las métricas clásicas (*precision*, *recall*, *F-score*, etc).

A modo de ejemplo, la Ecuación 3.1 presenta la formulación matemática del *Hamming Score*, donde  $D$  corresponde al conjunto de ejemplos e  $Y_i$  y  $Z_i$  a los conjuntos de etiquetas reales y predichas para un ejemplo  $i$ , respectivamente.

$$\text{Hamming-Score} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (3.1)$$

- Las métricas basadas en etiquetas, por otro lado, trabajan evaluando la eficacia del sistema en cada etiqueta por separado, y luego retornando el promedio *Micro* - cálculo global, contando los aciertos y desaciertos de todo el sistema para posteriormente calcular la métrica final - o *Macro* - calculando métricas sobre cada etiqueta y luego promediando dichas métricas para obtener un valor final -. En tal caso las métricas de clasificación singular se adaptan de buena forma, pero la selección de estrategia para el promedio debe hacerse con cuidado, en especial en casos con un alto desbalance de etiquetas.

Lo anterior podría llevar a pensar con justa razón que el desafío más grande con la clasificación multietiqueta es entonces la elección de métricas de evaluación adecuadas. Sin embargo, dicho título le pertenece a la última dificultad listada: el qué tantas etiquetas se busca predecir y el nivel de presencia de estas en los datos, dificultad que ha motivado a la introducción de métricas especiales para la descripción de conjuntos de datos multietiqueta tales como el *Label Average* [45] (LA, promedio simple de etiquetas asignadas por entrada y cuya formulación matemática se presenta en la Ecuación 3.2) y el ya introducido *Label Density* (LD), ambas las cuales son ampliamente utilizadas para jerarquizar conjuntos de

datos en cuanto a su dificultad, entendiendo que valores más bajos reflejan un aprendizaje más complejo.

$$\text{Label Average} = \frac{1}{|D|} \sum_{i=1}^{|D|} |Y_i| \quad (3.2)$$

Es en particular este desafío sobre la densidad del etiquetado el que, conjugado con distintos escenarios sobre la calidad de las etiquetas y el contexto de la tarea a resolver, ha dado lugar a diferentes tipos de aprendizaje multietiqueta, cada uno con sus aplicaciones y técnicas específicas. Dichos tipos se presentan a continuación.

## 3.2. Tipos de aprendizaje multietiqueta

### 3.2.1. Esquema general

En la actualidad, los problemas de aprendizaje multietiqueta se hacen presentes en distintos contextos y escenarios, según los cuales las técnicas recomendadas y la naturaleza del problema mismo varía ampliamente. Es por esta razón que la literatura propone una clasificación según tres grandes aspectos [30], la cual se esquematiza a través de la Figura 3.2 y se detalla a continuación:

1. La primera pregunta es si los datos son estáticos, es decir, si se tiene un conjunto fijo de datos para el aprendizaje o si estos se reciben en forma de *stream* de datos. En este último caso, el problema pasa a denominarse *Online MLC* (OMLC).
2. Si los datos son estáticos, viene entonces la segunda pregunta, que guarda relación con la calidad de las etiquetas: cuando estas presentan algún grado de error o de incompletitud, el problema pasa a pertenecer a la categoría de *MLC with Limited Supervision*. En el caso contrario, se procede con la tercera pregunta.
3. Por último, queda la pregunta si se trabaja con una cantidad “grande” de etiquetas. Si bien no existe un parámetro específico a partir del cual se determine dicha cantidad, un buen acercamiento es considerar que la cantidad de ejemplos para el entrenamiento debe ser de al menos dos ordenes de magnitud más alta que la de etiquetas a predecir - cayendo así en un caso de *Traditional MLC* - o, de lo contrario, se estaría hablando de un problema de *Extreme MLC* (XMLC).

A continuación se describirán brevemente cada uno de los escenarios de MLC mencionados, así como sus principales técnicas propuestas en el estado del arte.

### 3.2.2. *Online MLC*

Tratándose del caso en el cual los datos no son estáticos, *Online MLC* es aquel escenario en donde los datos se reciben a través de un *stream* en el tiempo [30]. Su existencia como

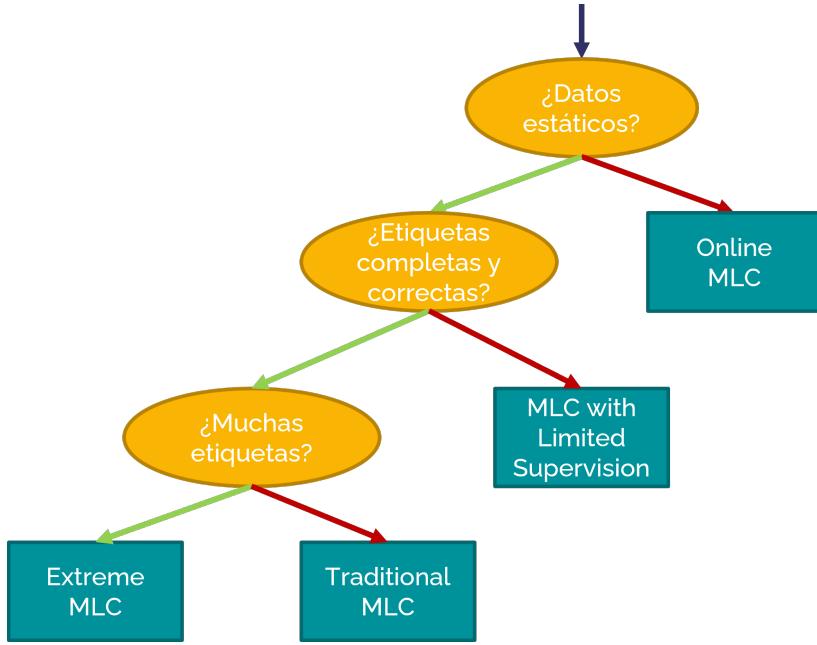


Figura 3.2: Tipos de MLC.

categoría responde a la dificultad que adapta los métodos *offline* a tal situación presenta, en cuanto en general estos últimos asumen que (1) los datos pueden ser cargados en memoria en su totalidad y (2) no consideran la posibilidad de que los datos expiren con el paso del tiempo, fenómeno que sí se hace presente en el caso *online* y que añade restricciones sobre la necesidad de un aprendizaje rápido y la capacidad de incorporar información nueva sobre la marcha, haciendo de este escenario un caso particularmente complejo de MLC.

En lo que respecta a las aplicaciones de OMLC, muchas de ellas se encuentran relacionadas con el estudio de interacciones en redes sociales. Por ejemplo:

- El estudio de intereses de usuarios a través de sus publicaciones en redes sociales y los metadatos que estos generan, a fin de ajustar la publicidad web que estos reciben. Este caso guarda además la particularidad de que la información, de tomar demasiado tiempo en ser obtenida, deja de ser útil, pues en general los intereses de usuarios evolucionan rápidamente.
- El estudio de tendencias en *Twitter*, a través del etiquetado múltiple de las publicaciones que se generan en dicha red social. Nuevamente, un procesamiento lento del flujo de datos podría significar el reaccionar de forma tardía ante una tendencia.

Sin embargo, el OMLC no se restringe solamente a aplicaciones sobre redes sociales: otros usos pueden darse también en el procesamiento de señales, datos de sensores, alertas astronómicas, entre otros.

Las técnicas utilizadas para tratar este tipo de problemas, por otro lado, responden principalmente a esfuerzos dedicados al abordaje del OMLC desde un enfoque de aprendizaje profundo, a saber:

- *Online Universal Classifier* (OUC) [15], técnica que busca responder a través de un mismo clasificador online al caso binario, multiclase y multietiqueta,

- OSML-ELM [48], modelo neuronal *feedforward* de capa oculta única que busca responder al caso multietiqueta a través de la transformación del caso singular y la incorporación de pesos en el entrenamiento,
- OLANGSD [37], método basado en la tarea de ranking de etiquetas, donde las funciones de ranking son aprendidas a través de la optimización de la *ranking loss* y que innova en cuanto incorpora la técnica de *Nesterov Smoothing* (AGD) como una alternativa de menor costo al descenso de gradiente estocástico (SGD) y que permite por ende una reducción de los tiempos de entrenamiento e inferencia para el caso online.

Dichas técnicas presentan, sin embargo, una importante deficiencia que limita la calidad de sus resultados y que se encuentra en el no uso de las eventuales relaciones entre etiquetas que, de encontrarse y ser explotadas, podrían contribuir ampliamente a mejorar las clasificaciones (descubriendo, por ejemplo, relaciones de exclusividad mutua entre una etiqueta y otra). Esto ha llevado al desarrollo de trabajos dirigidos particularmente a incorporar este conocimiento en los clasificadores, entre ellos:

- *Cost-sensitive dynamic principal projection* (CS-DPP) [11], método basado en la reducción del espacio de etiquetas a un espacio de menor dimensionalidad a través del cual se busca codificar las relaciones entre las mismas y con el cual la tarea de predecir se delega en un regresor a cargo de la tarea inversa,
- El trabajo propuesto en Gong et al. [19], el cual presenta un nuevo paradigma de aprendizaje de métricas para el OMLC a través de la proyección conjunta de *features* y etiquetas en un mismo espacio de *embedding* sobre el cual posteriormente se entranan métricas de distancia según restricciones específicas, y cuyos resultados experimentales muestran superioridad sobre los métodos anteriormente descritos.

En los últimos años se han planteado además nuevos trabajos que buscan abordar la reunión del caso de OMLC con otros tipos de MLC, tales como el de supervisión limitada y el de Extreme MLC. En este sentido, técnicas tales como OnSeML [20] han planteado un primer acercamiento más dejan aún un amplio campo por explorar, especialmente en lo que al trabajo con datos parcialmente etiquetados y al manejo de grandes volúmenes en contextos de memoria limitada se refiere, direcciones que aún requieren de mayores esfuerzos [30].

### **3.2.3. *MLC with Limited Supervision***

En la práctica, el proceso de construir un conjunto de datos acostumbra resultar costoso: conseguir entradas correctamente etiquetadas puede ser complejo, en especial cuando dichas etiquetas deben ser asignadas por un experto del dominio. Tal complejidad, propia de los problemas de clasificación singular, se ve ampliamente agravada en el caso multietiqueta, donde para cada entrada ya no basta con identificar una clase dominante, sino más bien quien etiqueta debe evaluar la pertenencia de la entrada a cada una de las etiquetas en estudio, que pueden ir desde unas pocas unidades hasta miles de ellas - lo cual con frecuencia se vuelve una tarea intratable para los anotadores [30] -. Es ahí donde nace esta nueva categoría de MLC, entendiendo que dicho proceso a menudo da lugar a etiquetados incompletos o parcialmente correctos y que requieren de técnicas robustas ante tal situación.

En este escenario, distintas configuraciones dan lugar a diferentes sub-problemas de supervisión limitada. Entre ellos:

- Aprendizaje multietiqueta con etiquetas faltantes (MLML por sus siglas en inglés), caso en el cual el etiquetado es correcto pero parcialmente completo: solo un subconjunto de las etiquetas correspondientes son asignadas,
- Clasificación multietiqueta semi-supervisada (SS-MLC), caso en el que el conjunto de datos se compone a su vez de un conjunto de datos etiquetados correcta y completamente y de un conjunto sin etiquetar,
- Aprendizaje multietiqueta parcial (PML) [56], que busca abordar el caso en que ciertas etiquetas son demasiado difíciles de asignar por un modelo (por ejemplo, es difícil en general diferenciar un *husky* de un *malamute*) y para lo cual se propone un enfoque más relajado, a través del cual cada instancia  $x$  es asociada a un conjunto de etiquetas *candidatas*  $S$ , dentro de las cuales algunas corresponden a las etiquetas correctas y el resto pasan a denominarse *etiquetas falso positivo* o *etiquetas distractoras*.
- Aprendizaje multietiqueta con etiquetas ruidosas (Noisy-MLC), que responde al caso en que se entregan etiquetas para todas las instancias pero estas no son necesariamente correctas ni completas,
- Aprendizaje multietiqueta con aprendizaje activo (MLAL), en el cual se seleccionan cuidadosamente las instancias más informativas para el etiquetado de manera tal que mejores modelos puedan ser entrenados con menos esfuerzo en la recopilación de datos,
- Clasificación multietiqueta con múltiples instancias (MIML), configuración en la cual se asume que cada ejemplo está conformado a su vez por múltiples instancias que corresponden a múltiples etiquetas binarias, es decir, el ejemplo completo es una “composición” de elementos de clasificación singular a ser identificados y así dar lugar a la salida multietiqueta.

Múltiples trabajos se han realizado buscando dar lugar a técnicas de clasificación aptas para cada escenario. Entre ellas, destacan:

- Métodos basados en la recuperación de una matriz de etiquetas *low-rank*, los cuales basan su funcionamiento en el supuesto de que la matriz de etiquetas es de bajo grado, es decir, presenta pocas columnas linealmente independientes y, en tal condición, buscan recuperar dicha matriz (incorporando la información de etiquetas faltantes) desde un enfoque de *matrix completion* [58] asistido por las *features* como matrices de información adicional. Ejemplos se encuentran en LEML [65], ColEmbed [22] y COCO [59].
- Métodos de *embedding*, los cuales mantienen el supuesto de *low-rank* pero cambian la estrategia al proyectar las etiquetas a una dimensión menor a la original y posteriormente utilizar uno o más regresores para obtener la decodificación completa de las mismas, incluyendo las etiquetas eventualmente faltantes. Ejemplos se encuentran en C2AE [63] - que incorpora además un *embedding* conjunto de *features* y etiquetas - y REFDHF [51], que incorpora conceptos de *hypergraph fusion* para explorar y utilizar la información complementaria entre el espacio de *features* y etiquetas.

- Métodos basados en grafos, los cuales asumen suavidad a nivel de *features* y de etiquetas, es decir, que si las *features* de dos ejemplos  $x_i$  y  $y_i$  son cercanas entre sí, también lo son sus vectores de etiquetas y, similarmente, si dos vectores incompletos de etiquetas  $y_i$  y  $y_j$  son semánticamente similares, también lo son sus vectores completos (a predecir) [30]. Bajo tal supuesto, estos métodos construyen uno o más grafos de etiquetas con pesos ajustados en base a la información que las *features* proveen para, posteriormente basar la inferencia en algoritmos de búsqueda de vecinos más cercanos, tales como *kNN*. Ejemplos se encuentran en WELL [43], MLMG-SL [54] y GLOCAL [73], entre otros.
- Métodos de aprendizaje en dos etapas, los cuales apuntan abordar el caso de PML mediante una primera fase de exploración de correlación entre etiquetas - la cual da como resultado un conjunto de etiquetas candidatas, además de un valor de confianza para cada una - seguida de una segunda fase de identificación de las etiquetas reales a través de la aplicación de técnicas de umbral sobre la confianza de conjuntos de etiquetas, construida en base a las confianzas individuales. Dos ejemplos recientes de dicho enfoque se encuentran en PARTICLE [66] y DRAMA [49].
- Métodos basados en redes neuronales, los cuales frecuentemente apuntan al caso de múltiples instancias (MIML) al buscar abordar la presencia de múltiples objetos con significados individuales mediante las capacidades de atención dirigida que ciertas arquitecturas ofrecen. En este sentido, MIML-FCN+ [60] ofrece un buen ejemplo en el cual se explora el uso de redes convolucionales *fully-connected* con información adicional (*privileged information, PI*).

Es importante mencionar que, si bien en esta sección se han presentado los escenarios y métodos más frecuentes, existen otras configuraciones y áreas de investigación en el campo del MLC con supervisión limitada que aún no han sido exploradas en profundidad, y que por ende requieren de mayor atención [30].

### 3.2.4. *Traditional MLC*

Cuando se cuenta con datos completos, bien etiquetados y una cantidad de etiquetas razonable (al menos 2 órdenes de magnitud menos que la cantidad de ejemplos disponibles para entrenamiento) cobra sentido la categoría de MLC Tradicional, en cuanto los problemas de dicha clase por lo general son abordables desde las técnicas más tradicionales y/o simples, adquiriendo el beneficio de contar con bajos costos de entrenamiento y predicción.

Dichas técnicas se basan en un proceso de *feature engineering* previo, es decir, en la determinación y codificación manual de los aspectos de los datos más importantes para la predicción, y se clasifican en general en dos grandes categorías: aquellas técnicas de transformación del problema y aquellas de adaptación del algoritmo [72].

Los métodos de transformación del problema, por un lado, corresponden al mapeo del problema multietiqueta hacia múltiples problemas singulares (uno por cada etiqueta) a resolver por “clásificadores base” (clásificadores binarios o multiclas), conformando así la forma quizás más intuitiva de lidiar con tal desafío - a menudo denominada como métodos *One vs All* [30] -. Al respecto:

- El representante por excelencia de este enfoque es el algoritmo de *Binary Relevance* (BR) [67], el cual consiste en entrenar un clasificador binario para cada etiqueta y predecir con la concatenación de la salida de cada clasificador. Dicho método, aunque poderoso, es también denominado en ocasiones como un método ingenuo, en cuanto realiza un supuesto que puede resultar demasiado grande: la inexistencia (o irrelevancia) de relaciones entre las etiquetas.
- Existen sin embargo otros métodos de la misma familia que no requieren de supuestos tan extremos, y que sin embargo presentan otras debilidades. Un ejemplo de ello es *Classifier Chain* (CC) [39], el cual a través de una transformación a múltiples problemas binarios encadenados permite rescatar información relativa a relaciones entre etiquetas, mas bajo un importante costo en complejidad espacial que lo vuelve inapropiado para grandes conjuntos de etiquetas y una fuerte dependencia del orden en que las etiquetas se presenten (orden de la cadena).
- Otras alternativas que pertenecen también a la categoría de transformación del problema se encuentran en los métodos que no mapean el multietiquetado a múltiples problemas binarios si no a un único problema multiclasa. En este caso el ejemplo por excelencia es *Label Powerset* (LP), algoritmo en el cual el objeto de predicción no son las etiquetas por sí solas si no más bien las combinaciones de etiquetas - conformando así un problema multiclasa donde cada combinación es una clase -. Dicho enfoque logra captar las relaciones entre etiquetas, mas sufre de dos grandes debilidades: la necesidad de requerir que todas las combinaciones posibles de etiquetas se encuentren en los datos de entrenamiento (lo cual en general es difícil de conseguir y da lugar a una generalización débil) y una alta probabilidad de subajuste en espacios de alta dimensión.
- Una versión mejorada de LP se halla en *Random k-Labelsets* (RakelID) [47], método el cual busca hacer frente a las debilidades de dicha técnica mediante la construcción de particiones aleatorias del espacio de etiquetado y su posterior uso como combinaciones para LP, presentando sin embargo una fuerte dependencia con la probabilidad de observar las particiones aleatorias en los datos reales. *Ovelapped Random k-Labelsets* (RakelO), por otro lado, busca atacar este último punto al convertir las particiones en subconjuntos potencialmente sobrepuertos.

En la otra vereda se encuentran los métodos de adaptación del algoritmo, los cuales buscan extender algoritmos de clasificación singular para tratar con la clasificación multietiquetada.

- Ejemplos de métodos pertenecientes a esta categoría son *Multilabel k-Nearest Neighbors* (MLkNN) [72] y *Binary Relevance k-Nearest Neighbors* (BRkNN) [57], los cuales mantienen una idea general de ampliar el funcionamiento clásico de *kNN* e incorporan alguna técnica adicional que permita traducir los vecindarios hacia las múltiples salidas esperadas (inferencia bayesiana en el caso de MLkNN y mediante BR en el de BRkNN).
- Otro ejemplo se encuentra en *Multilabel Twin Support Vector Machines* (MLTSVM) [9], el cual busca capturar la información multietiquetada a través de la determinación de múltiples hiperplanos, uno para cada etiqueta.

- Nuevamente, cada método tiene sus ventajas y desventajas, siendo una característica transversal la condición de que la distancia entre etiquetas sea un buen predictor para su asignación.

Existe sin embargo una característica propia de las técnicas tradicionales que pone en desventaja a ambas familias de métodos. Se trata de la fuerte dependencia que se genera con la calidad de los descriptores o *features* [23], la cual viene determinada por un proceso previo el cual no se retroalimenta de forma automática y que puede ser significativamente demandante en tiempo experto. Es esta desventaja la que ha motivado al surgimiento de estrategias *End-To-End*, donde la generación de *features* se presente como parte del método mismo y no como un paso previo desconectado del aprendizaje multietiqueta. En este sentido, las técnicas tradicionales se han visto desplazadas por aquellas basadas en el entrenamiento de *embeddings* - tales como las estudiadas en las secciones previas - mas sin embargo el explorar la posibilidad de abarcar problemas desde este enfoque no deja de merecer al menos un par de experimentos, en cuanto no dejan de presentar una importante ventaja en costos cuando la alternativa se trata de, por ejemplo, redes profundas.

### 3.2.5. *Extreme MLC*

El último caso en estudio es el de *Extreme MLC*, aquel donde la meta es “aprender un clasificador capaz de anotar automáticamente un ejemplo con el subconjunto más relevante de etiquetas a partir de un número de etiquetas extremadamente largo” [30], situación en la cual el obtener ejemplos anotados es aún más difícil dada la cantidad de etiquetas a asignar y por ende se ve acompañada de volúmenes de datos a menudo insuficientes para un entrenamiento óptimo con las técnicas anteriores, lo cual lleva a la necesidad de desarrollar métodos específicos para este caso que sean capaces de explotar al máximo la información disponible para obtener mejores resultados.

A modo de ejemplo, supóngase el caso de Instagram, red social donde diariamente se suben millones de *selfies* y fotos grupales. Basado en las primeras, podría resultar interesante construir un clasificador que reconozca quien aparece en las segundas y de esta forma entregue sugerencias de etiquetado en publicaciones. En tal caso, los nombres de las personas serían las etiquetas - millones de ellas - y una misma entrada podría relacionarse con más de una etiqueta - por ejemplo, en una foto familiar -. Tal desafío sería interesante de abarcar y, sin embargo, hasta el momento solo está cubierto para un número reducido de etiquetas, a menudo presente en aplicaciones de galería para dispositivos móviles (las cuales “detectan” a las personas y ofrecen la posibilidad de ordenar las fotos basándose en ello). Desde luego, este ejemplo podría también verse conectado con el caso de OMLC, haciendo de él una tarea aún más desafiante.

Para abordar dichos desafíos, múltiples trabajos han propuesto distintos acercamientos. Entre ellos, destacan:

- La aplicación de métodos tradicionales, que sin embargo han presentado dificultades para adaptarse a un escenario con un volumen de etiquetas tan grande debido a los costos computacionales de entrenar un clasificador para cada una. En este sentido, técnicas tales como PD-Sparse [64] han logrado sobrelevar este desafío pero al costo de no explotar las correlaciones entre etiquetas [30].

- Métodos de *embedding*, que amplían los supuestos de *low-rank* al caso de XMLC para de esta manera abordar las tareas mediante la compresión - proyección de las etiquetas a un espacio de menor dimensionalidad - y posterior descompresión mediante uno o más regresores, siendo SLEEC [4] el principal exponente.
- Métodos basados en árboles, los cuales buscan mapear el problema de gran escala a uno abordable con los métodos tradicionales mediante la partición jerárquica de los conjuntos de *features* y etiquetas. Ejemplos de ello se encuentran en técnicas tales como FastXML [38] y GBDT-SPARSE [41].

En últimos años, por otro lado, un nuevo acercamiento al XMLC ha tomado lugar. Se trata del XMLC a través de *Deep Learning*, enfoque que, gracias a la poderosa capacidad de aprendizaje de las redes neuronales y la capacidad de incluir funciones de pérdida específicas a cada caso, ha logrado posicionarse como el estado del arte en muchas aplicaciones reales de MLC [30]. En este sentido, ha resultado clave encontrar métodos que exploten dicha capacidad para capturar las relaciones entre etiquetas, lo cual ha llevado a la propuesta de distintas técnicas:

- Las *Binary Relevance Neural Networks* (BRNN), que corresponden a la traducción del algoritmo de *Binary Relevance* tradicional al *Deep Learning*, utilizando ahora redes profundas de clasificación binaria como clasificador base. Si bien este enfoque mantiene la desventaja de no explotar relaciones, adquiere ahora la bondad de aprender *features* específicas a cada etiqueta - lo cual facilita la decisión binaria - aunque a un costo de espacio y tiempo mucho mayor que el del caso tradicional, pues requiere del entrenamiento de tantas redes profundas como etiquetas a predecir.
- Las *Threshold Dependent Neural Networks* (TDNN), que basan su funcionamiento en el abordar el problema mediante una red profunda con un vector de salida del tamaño de la cantidad de etiquetas en estudio y posteriormente utilizar una técnica de umbral sobre la normalización del mismo para, a partir de cierto valor, asignar las clases correspondientes - logrando así una clasificación multietiqueta -. En este sentido, las TDNN ganan la virtud de explotar relaciones y economizar costos con respecto a las BRNN, pero incorporan la desventaja de incluir un mecanismo adicional de *thresholding* cuyo ajuste de por sí representa un desafío [69].
- Métodos de *Deep Embedding*, que buscan alcanzar una mejor representación de la entrada a través de la aplicación de redes neuronales entrenadas con funciones de pérdida especialmente diseñadas para el caso multietiqueta y su posterior clasificación con clasificadores tradicionales. Ejemplos se encuentran en BP-MLL [70], el ya introducido C2AE [63] y DCSPE [52].
- Métodos basados en redes convolucionales (CNN) en conjunto a redes recurrentes (RNN). Este enfoque propone utilizar las CNN como *embedding* de *features* y, al mismo tiempo, utilizar RNN como *embedding* de etiquetas, alcanzando así un *embedding* conjunto entrenable de extremo a extremo [50].
- Métodos basados en redes de grafos convolucionales (GCN), las cuales construyen un grafo dirigido sobre las etiquetas y posteriormente modelan la dependencia entre las

mismas a través de un proceso de convolución de grafos para finalmente mapear la representación de etiquetas a clasificadores interdependientes. Ejemplos se encuentran en GCN [61] y SSGRL [8].

Un punto importante a considerar para la investigación es que, si bien las arquitecturas profundas más avanzadas presentan una poderosa capacidad de aprendizaje y por ende más potencial para el problema de clasificación multietiqueta, vienen normalmente acompañadas por una alta complejidad en términos de costos de entrenamiento y predicción, además de la necesidad de ajustar un amplio espectro de parámetros (funciones de pérdida, optimizadores, tamaños y número de capas, entre otros). Es por esta razón que la literatura recomienda comenzar explorando el diseño de arquitecturas ligeras que permitan tener un entrenamiento y predicción eficientes y, de esta manera, experimentar con métodos más profundos de forma progresiva.

### 3.3. Comentarios

Una vez descrito el estado del arte para las distintas áreas de MLC, cabe preguntarse a qué área en específico corresponde el problema en estudio. En este sentido, es importante destacar:

- Aunque la herramienta de apoyo a concebir podría verse en un futuro implementada como una plataforma web capaz de retroalimentarse a sí misma en base a la experiencia usuaria - acercándose de esta manera al caso de OMLC -, dicha posibilidad queda aún fuera del alcance de esta investigación, en cuanto se trata de la primera exploración de la tarea desde un enfoque de MLC.
- Similarmente, la puesta en marcha de una aplicación que utilice el sistema a concebir podría poner en evidencia situaciones de aprendizaje con supervisión limitada, por ejemplo, si los usuarios complementasen la información entregada por el sistema con etiquetas no vistas en los datos a utilizar. Al respecto, se opta por dejar dicha posibilidad fuera de alcance, pues requiere de una solución previamente concebida (que no es el caso de esta investigación). Sin embargo, se reconoce como un posible trabajo futuro.
- Por último, al estudiar la relación entre etiquetas y patrones disponibles para el entrenamiento - ambas cantidades en el orden de las centenas - se detecta de inmediato un caso de XMLC, y en particular uno con condiciones bastante desfavorables.

De esta manera, la presente investigación aborda el problema del etiquetado de patrones geométricos desde un Enfoque XMLC, tomando, sin embargo, el tiempo de explorar también el desafío mediante un Enfoque Tradicional, a fin de contar con una base sobre la cual evaluar los resultados que los métodos más avanzados puedan entregar.

# Capítulo 4

## Metodología

Este capítulo tiene por fin el presentar la metodología sobre la cual se conducirán los experimentos a realizar. Para ello, en primer lugar se formalizan los objetivos de la investigación (Sección 4.1) y los supuestos sobre los cuales esta se basará (4.2) para luego presentar las distintas variables en estudio, distinguiendo aquellas a fijar y a controlar (4.3). Posteriormente se lista la selección de métodos a poner a prueba para cada enfoque (4.4) y se profundiza en los experimentos a conducir (4.5), concluyendo con la introducción de las métricas a utilizar para evaluar los resultados (4.6).

### 4.1. Objetivos

#### 4.1.1. Objetivo general

Concebir una herramienta computacional que, a partir de un patrón geométrico proveniente de un objeto de herencia cultural, entregue una serie de etiquetas en base a las características geométricas más importantes del mismo, contribuyendo así al proceso de etiquetado manual de patrones por parte de expertos de la arqueología como una herramienta de apoyo al proceso, sentando en mismo tiempo un precedente para una futura automatización del proceso.

#### 4.1.2. Objetivos específicos

- OE1. Construir un conjunto de datos apto para ser utilizado como entrada de métodos de aprendizaje multietiqueta a partir de la extracción de los patrones geométricos etiquetados disponibles en *Geometrischen Ornamente Vasen: Ein Kompendium* [28].
- OE2. Encontrar, mediante la experimentación con múltiples métodos, un modelo de aprendizaje multietiqueta que permita asignar etiquetas probables a los patrones geométricos, utilizando para ello el conjunto de datos mencionado.
- OE3. Seleccionar un conjunto de métricas apropiadas para el problema, tomando en consideración supuestos relativos a las aristas más importantes para los potenciales usuarios.

- OE4. Reconocer los métodos que mejor se adaptan al problema y formular hipótesis sobre las razones detrás de dicho fenómeno, a fin de construir lineamientos para el trabajo futuro.
- OE5. Reconocer cómo el tamaño del universo de etiquetas posibles y la cantidad de ejemplos disponibles para cada etiqueta se relaciona con la calidad de los resultados de los distintos métodos.
- OE6. Aplicar distintas técnicas de *Data Augmentation* e identificar el impacto de su uso en la calidad de los resultados.
- OE7. Aplicar técnicas para el tratamiento al desbalance de casos positivos (asignación) y negativos (no asignación) de cada etiqueta y evaluar su efecto en un contexto de aprendizaje multietiqueta.
- OE8. Traducir los resultados más importantes a un conjunto de afirmaciones que faciliten la comprensión de los mismos por parte de expertos del dominio en estudio.

## 4.2. Supuestos

La investigación toma cuatro supuestos sobre la naturaleza del problema. Estos son:

- S1. La representatividad del desafío con respecto a otros contextos de etiquetado de patrones geométricos en objetos de herencia cultural, en particular, en lo que respecta a objetos con un origen distinto a aquel de los disponibles en *Die Ornamente* (que corresponden principalmente a excavaciones en Grecia del siglo XX).
- S2. La cantidad mínima por patrón de etiquetas a predecir de forma tal que signifique un aporte para el área. En este sentido, se considerará que el solo hecho de asignar al menos una etiqueta correcta por patrón ya representa un apoyo para el etiquetador, considerando que en promedio estos reciben 3 etiquetas.
- S3. La importancia de las etiquetas menos frecuentes. Tal como se introdujo en la Descripción de los datos (Sección 2.2), existen muchas etiquetas con baja frecuencia y muy pocas con una frecuencia alta, lo que se asume podría guardar relación con el principio del mínimo esfuerzo [74]. Al respecto, se considerará la existencia de una correlación inversa entre frecuencia y relevancia, asumiendo que las etiquetas menos frecuentes (y por ende, que más esfuerzo representan) son también las más interesantes o que más información aportan a la descripción del patrón.
- S4. Se considerará además que el trabajo de descartar una cantidad razonable de sugerencias erróneas a cambio de incorporar alguna sugerencia correcta es aceptable, en cuanto para los expertos resultaría más sencillo reconocer etiquetas incorrectas que idear etiquetas apropiadas.

En lo que respecta a S1, cabe destacar que se opta por asumir la representatividad del desafío y no del conjunto de datos como tal, pues esto último representaría un supuesto quizá demasiado grande. De esta manera, no se espera que los patrones y etiquetas en estudio coincidan, por ejemplo, con lo que podría encontrarse en excavaciones en Roma o en Latinoamérica.

Por otra parte, el supuesto S2 responde a la necesidad de establecer un desempeño a partir del cual se considere valor agregado al proceso. Desde luego, dicho desempeño representa solo un mínimo aceptable, mas se buscará alcanzar los mejores resultados posibles.

El supuesto S3, por otro lado, complementa el supuesto anterior al declarar la importancia de alcanzar los mínimos impuestos no solo en un escenario de etiquetas más frecuentes, sino también en aquellos escenarios más complejos donde se cuenta con apenas un par de ejemplos para cada etiqueta.

Finalmente, el supuesto S4 responde a dos fenómenos particulares: en primer lugar, la inherente dificultad de recomendar *solo* las etiquetas correctas y, por otro lado, la eventual posibilidad de que los patrones disponibles para el estudio no estén del todo etiquetados, es decir, que existan etiquetas que el autor podría haber ignorado y que el sistema recomendador busque “completar” entregando etiquetas que para las métricas serían falsos positivos, pero para un experto podrían ser verdaderas (encontrándose así además un caso de MLML).

### 4.3. Variables en estudio

Los experimentos a conducir buscarán responder progresivamente a cinco grandes preguntas guía que a su vez apuntan a distintos objetivos específicos. Estas son:

- PG1. En virtud del conjunto de datos a utilizar, ¿Es posible formular una herramienta de apoyo al etiquetado mediante aprendizaje multietiqueta, entendiendo por apoyo el reducir la carga efectiva de trabajo de los etiquetadores al replicar parte de su labor? (OE1, OE2).
- PG2. Si la respuesta anterior es afirmativa, ¿Con cuántas etiquetas sería capaz de trabajar la herramienta de apoyo? ¿Cuántos ejemplos de cada etiqueta se requerirían para que el sistema fuese capaz de aprenderlas? (OE4, OE5).
- PG3. De la mano con lo anterior, ¿Podrían obtenerse mejores resultados si se contase con más datos? En particular, ¿Se encontraría una mejora al experimentar con datos sintéticos? ¿Qué técnica convendría utilizar para la generación de dichos datos? (OE6)
- PG4. Anteriormente se mencionó que el bajo valor de LD sugiere un importante desbalance entre casos positivos (asignación) y negativos (no asignación) para cada etiqueta, y se tiene la hipótesis de que esto representa un obstáculo para el aprendizaje. ¿Qué técnicas podrían utilizarse para lidiar con este aspecto? ¿Cómo influye su aplicación en los resultados? (OE7)
- PG5. Por último, ¿Con qué nivel de confianza podría el usuario final interpretar las sugerencias del sistema de apoyo final? ¿Podría eventualmente un etiquetador delegar responsabilidades en dicho sistema sin necesidad de hacer una revisión de las sugerencias? OE3, OE8)

Mientras la última pregunta encontrará su respuesta en la interpretación de las métricas obtenidas por cada experimento (las cuales se detallarán en la Sección 4.6), la respuesta de las otras vendrá dada por el manejo de distintas variables de control. Estas son:

- VC1. Los métodos a utilizar, en donde se probará con distintas propuestas tanto del MLC Tradicional como del XMLC a fin de responder con propiedad PG1.
- VC2. La aplicación de técnicas para el tratamiento del desbalance, lo cual vendrá condicionado por las posibilidades de tratamiento al desbalance que los distintos métodos admitan.
- VC3. La cantidad de etiquetas a predecir, aspecto el cual se abordará desde un enfoque de poda de etiquetas menos frecuentes (con menos eventos en los datos) para así reconocer los alcances de la herramienta y hacer frente también al ya mencionado obstáculo del LD, manipulando a través de esta variable la dificultad del problema.
- VC4. El origen de los descriptores a utilizar como entrada de las técnicas que requieran de ello, es decir, aquellas que no incorporen la generación de *features* como parte de su trabajo.
- VC5. La aplicación de distintas técnicas de *Data Augmentation* para generar datos sintéticos, aspecto para el cual se determinará un conjunto de técnicas a utilizar y se realizarán experimentos con cada una de ellas a fin de reconocer las que mejor se adapten al problema en estudio.

En este punto es importante destacar que, si bien en la Sección 3.2 se presentaron una gran variedad de métodos tanto para MLC Tradicional como para XMLC, abordar el problema desde todas las aristas posibles sería una tarea titánica y que escaparía de los alcances de esta tesis, en especial por el tiempo que requeriría implementar cada una de ellas y realizar los entrenamientos correspondientes. En su lugar, se opta por escoger un conjunto reducido de técnicas bajo ciertas hipótesis (cuyo detalle se presenta en la Sección 4.5) y dedicar tiempo a buscar su mejor expresión, variando tanto los hiperparámetros que estas incorporen como el escenario en el cual se desempeñan al manipular las distintas variables descritas.

Existe además una serie de variables cuyo aporte a los resultados no se estudiará en profundidad, si no más bien tomarán valores fijos con la intención de mantener una cantidad razonable de experimentos. Entre ellas, destacan:

- VF1. La cantidad de patrones en estudio (restringida a aquellos presentes en el libro de Kunisch) y la información de cada patrón a utilizar como entrada para los algoritmos, en cuanto se considerará solamente la imagen (*sketch*) de los mismos y no la información adicional que el libro contiene para algunos de ellos, tal como la ubicación en museos de los objetos de los cuales forman parte.
- VF2. Las configuraciones específicas a cada método, salvo que se indique lo contrario. Esto cobrará especial relevancia con ciertos métodos que incorporan un gran conjunto de parámetros, casos en los cuales será necesario fijar parte de ellos para evitar un incremento exponencial de los experimentos a conducir y de los costos asociados.

La eventual manipulación de dichas variables y los efectos que ello podría tener sobre los resultados se discutirá al cierre de la investigación como parte del Trabajo Futuro (9.1), con miras a iteraciones futuras de la herramienta.

## 4.4. Selección de métodos

Los métodos a poner a prueba se seleccionan con base en la hipótesis de que, si bien el problema en su estado original presenta claras características de un desafiante caso de XMLC (tales como una cantidad similar de patrones y etiquetas), la simplificación del mismo a través del podado de etiquetas podría convertirlo en un problema abordable desde el enfoque tradicional y, como consecuencia, hacer relevante la exploración desde sus métodos. Esto además permitiría dilucidar a partir de cuántas apariciones en los datos las etiquetas pasan a ser abordables desde un enfoque u otro, entregando conocimiento adicional que podría servir para la construcción de mejores conjuntos de datos en un futuro.

Bajo este razonamiento, se opta entonces por abordar el problema desde ambos enfoques, realizando una selección de métodos de cada uno según sus principales características. Dicha selección se presenta a continuación.

### 4.4.1. Enfoque Tradicional

En el caso del Enfoque MLC Tradicional, la selección es la siguiente:

- Para la familia de métodos de transformación del problema se seleccionan Binary Relevance (BR), Label Powerset (LP), Classifier Chain (CC) y RakelD, cada uno con Regresión logística (LR), Máquinas de Soporte Vectorial (SVC), Árboles de Decisión (DT) y Gaussian Naive-Bayes (GNB) como clasificador base.
- Para la familia de adaptación del algoritmo, por otro lado, se decide experimentar con Binary Relevance kNN (BRkNN), Multi-label kNN (MLkNN) y Multi-label Twin Support Vector Machine (MLTSVM).

Si bien el detalle de cada técnica se presentará en el capítulo dedicado a ello (6.1), cabe destacar que esta amplia selección de métodos permitirá abordar el enfoque tradicional tanto con métodos que no explotan las relaciones entre etiquetas (como es el caso de BR y BRkNN) como con otros que sí lo hacen y de distintas formas (CC, LP, MLTSVM), cubriendo de esta manera un amplio espectro de posibilidades. Es importante mencionar también que los experimentos con dichas técnicas se llevarán a cabo mediante las implementaciones disponibles en la librería de Python *scikit-multilearn* [44], la cual entrega una amplia gama de clasificadores multietiqueta debidamente documentados.

### 4.4.2. Enfoque XMLC

Para el Enfoque XMLC, por otro lado, se escogen las siguientes técnicas:

- Canonical Correlated Autoencoder (C2AE) [63], como representante de los métodos de *Deep Embedding*,
- La aplicación de AlexNet [27] como CNN multietiqueta a través de su implementación como *Threshold Dependent Neural Network* (TDNN),

- La aplicación de AlexNet para la descripción de los patrones y de celdas recurrentes LSTM para la descripción de las etiquetas, a fin de obtener un *embedding* conjunto según se describe en CNN-RNN [50].

De esta manera, la investigación incorpora experimentos de las principales familias de aprendizaje profundo para XMLC, dejando fuera solamente las *Binary Relevance Neural Networks* (BRNN) por considerarse una técnica demasiado demandante computacionalmente y poco prometedora dada la no explotación de relaciones entre etiquetas, y los métodos basados en redes de grafos convolucionales (GCN), los cuales se descartan por razones de tiempo pues, a diferencia del caso tradicional, en general no se cuenta con implementaciones públicas y/o mantenidas regularmente para estos métodos, lo cual obliga a invertir tiempo extra en desarrollar implementaciones propias y/o curar desarrollos de terceros.

## 4.5. Experimentos a conducir

Los experimentos a realizar vendrán dictados tanto por el método en juego como por el escenario, este último descrito por (1) la cantidad de etiquetas a predecir y (2) la técnica de *Data Augmentation* (D.A) en uso, además de (3) la eventual participación de técnicas de balance específicas a cada método. A continuación se presenta la organización de dichos experimentos.

### 4.5.1. Enfoque Tradicional

E1. En primer lugar se llevarán a cabo los experimentos del Enfoque Tradicional, comenzando por una comparación de todas las combinaciones método de transformación-clasificador base (un total de 16) y los métodos de adaptación del algoritmo (3) sobre el conjunto de datos original y variando la cantidad de etiquetas a predecir, con y sin el uso de técnicas para el balanceo. Al respecto:

- El experimento tomará como entrada descriptores obtenidos a partir de una red ResNet18. Tal decisión se justificará más en detalle en la sección correspondiente (6.1).
- El objetivo del experimento será obtener una primera noción de qué técnicas se adaptan mejor al problema (VC1), en qué rango de etiquetas lo hacen (VC3) y determinar el efecto de incorporar técnicas para el balanceo (VC2).
- La cantidad de etiquetas a predecir variará entre 5 y 330 con un paso “deseado” de 10.
  - El paso es deseado y no exacto en razón de que, al venir la cantidad de etiquetas dada por su frecuencia los datos, no es posible seleccionar exactamente una cantidad de etiquetas sin descartar injustificadamente algunas de ellas.
  - En su lugar, se opta por trabajar con los umbrales encontrados al aumentar la cantidad de etiquetas en estudio de 10 en 10 a partir de 5 hasta 330, lo cual da lugar a 13 umbrales diferentes, los cuales se presentan en la Tabla 4.1.

Tabla 4.1: Cantidades de etiquetas en estudio y umbral de frecuencia asociado

Cantidad de etiquetas	Umbral de frecuencia
5	85
14	37
26	21
34	17
54	9
63	8
72	7
82	6
91	5
107	4
131	3
169	2
281	1

E2. Luego se fijará el uso de la técnica de balanceo al caso que mejores resultados haya arrojado en el experimento anterior y se procederá a evaluar nuevamente el comportamiento de los distintos métodos al variar la cantidad de etiquetas pero esta vez cambiando los descriptores a utilizar (VC4). Sobre ello:

- Los nuevos descriptores serán obtenidos mediante ResNet50.
- El objetivo de este experimento será identificar si descriptores más grandes dan lugar a mejores resultados en el conjunto *base*, y si este fenómeno es transversal a los distintos métodos o se da con algunos en específico.

E3. Por último se fijará la técnica de balanceo y se seleccionarán los tres métodos que mejores resultados hayan entregado hasta el momento para proceder a estudiar su comportamiento al variar la técnica de D.A. en uso (VC5) y con ambos conjuntos de descriptores. Al respecto:

- El objetivo principal de este experimento será encontrar las mejores técnicas de D.A. para el escenario en cuestión, considerando tanto su impacto en el entrenamiento de las redes generadoras de descriptores como en el entrenamiento de clasificadores.
- La decisión de medir tal impacto solo sobre el método que mejores resultados haya presentado hasta el momento y no sobre todos responde a economía de recursos, pues de lo contrario se estarían evaluando 19 métodos contra 16 técnicas de D.A. sobre 2 orígenes de descriptores y en 13 cantidades de etiquetas, lo que se estima requeriría alrededor de 14 semanas para obtener resultados.
- Este experimento complementará al anterior en cuanto se considera que es probable que el contar con más datos para el entrenamiento contribuya a que descriptores más grandes obtengan mejores resultados.

- A esta altura los experimentos habrán permitido identificar la mejor configuración de las variables de control para el Enfoque Tradicional bajo la única salvedad de asumir que el mejor método de los experimentos E1 y E2 se mantiene frente a los cambios en E3.

#### 4.5.2. Enfoque XMLC

E4. Luego seguirán los experimentos del Enfoque XMLC. Para ello se comenzará con C2AE, técnica la cual será evaluada directamente ante el espectro completo de etiquetas, variando los descriptores entre la imagen convertida a vector *flatten* y las *features* entregadas por una red AlexNet preentrenada sobre ImageNet (la justificación de dicha elección se presentará más en detalle Sección 7.1). Cabe sin embargo destacar:

- El objetivo de este experimento será encontrar la técnica de D.A. (VC5) que mejor se adapte al problema mediante C2AE, así como el conjunto de descriptores más apropiado (VC4).
- La decisión de repetir el ejercicio de encontrar la mejor técnica de D.A. responde a la hipótesis de que dicha técnica no es necesariamente la misma para todos los métodos y/o enfoques en estudio.
- Cabe notar que, para el caso de C2AE, no se estudiará el efecto de aplicar técnicas de balanceo (VC2) pues el método no incorpora tal posibilidad.

E5. El siguiente experimento corresponderá a la aplicación de AlexNet como TDNN, caso en el que la entrada corresponderá directamente a las imágenes de los patrones. Para ello se iniciará buscando la mejor técnica de D.A. con la tarea simplificada a 107 etiquetas y variando la aplicación de tratamiento del desbalance.

- Al tratarse de un método que incorpora la generación de *features*, la variable de control relacionada al uso de descriptores (VC4) sale del estudio.
- Por otro lado, al tratarse la técnica de una *Threshold-Dependent Neural Network*, se suma una nueva variable de control relacionada al umbral de asignación de la red. Para este experimento particular, sin embargo, dicha variable se fijará a 0,5.
- De esta manera, el objetivo principal de este experimento será encontrar la mejor técnica de D.A. (VC5), tomando como supuesto que dicho orden se mantiene en el espectro completo (por motivos de economía de recursos) y agregando además el supuesto de que el orden se mantiene también al variar el umbral de asignación. Desde luego, este experimento también contribuirá a reconocer el efecto de aplicar tratamiento al desbalance (VC2).

E6. Posteriormente se estudiará la evolución de AlexNet TDNN al variar la cantidad de etiquetas para la mejor técnica de D.A., con la opción de tratamiento al desbalance que resulte más conveniente según E5. Al respecto:

- El objetivo de este experimento será fortalecer la conclusión de E5 por medio del estudio de la mejor configuración en 107 etiquetas ahora en el espectro completo (VC3), lo cual además permitirá comparar resultados con los obtenidos por el Enfoque Tradicional y C2AE.

- E7. El último experimento de AlexNet TDNN corresponderá a la repetición de E6 pero esta vez probando con distintos valores para el umbral de asignación, a fin de evaluar así el impacto de dicha variable (habiendo entonces cerrado todas las variables de control para TDNN).
- E8. Luego se realizarán los experimentos con CNN-RNN, siguiendo la implementación original basada en AlexNet y LSTM. El primer ejercicio corresponderá, al igual que con los experimentos anteriores, a la comparación de distintas técnicas de D.A. sobre la tarea simplificada a un número fijo de etiquetas.
- Cabe mencionar que, al igual que C2AE, esta técnica no incorpora alternativas para el tratamiento del desbalance y, como AlexNet TDNN, no utiliza descriptores previamente generados, por lo cual las únicas variables de control son la cantidad de etiquetas y las técnicas de D.A. a utilizar, siendo la definición de esta última el objetivo de este experimento.
- E9. Finalmente se estudiará la evolución de CNN-RNN al variar la cantidad de etiquetas, escogiendo para ello la mejor técnica de D.A. y comparando también con la evolución en el conjunto de datos original.

De esta manera, los experimentos permitirán responder a las preguntas de qué método se adapta mejor al problema (PG1), cuántas etiquetas es capaz de aprender (PG2), la influencia de utilizar datos sintéticos (PG3) y el efecto de incorporar técnicas para el balanceo de casos positivos y negativos en el estudio (PG4). La pregunta relativa a la confianza con la cual los resultados deben ser interpretados (PG5), por otra parte, se responderá a través de las métricas de evaluación, a detallar más adelante (Sección 4.6).

#### 4.5.3. Incorporación de *K-Fold Cross Validation*

Un punto importante a mencionar sobre la metodología experimental guarda relación con la forma de proceder ante cada experimento. A fin de incorporar robustez en los resultados y sus interpretaciones, se opta por llevar a cabo los experimentos utilizando *K-Fold Cross Validation* según se describe en *An Introduction to Statistical Learning* [25], adoptando un valor  $K = 4$ . Esto se traduce en los siguientes efectos:

- El conjunto de datos original da lugar a cuatro particiones distintas, cada una con una proporción de datos para prueba y entrenamiento de 0.25 y 0.75, respectivamente.
- En cada partición, el conjunto de prueba cambia para tomar otra cuarta parte de los datos, cambiando así también el conjunto de entrenamiento. Al par (conjunto de entrenamiento, conjunto de test) se le denominará un *fold* de los datos.
- Los conjuntos de prueba en cada *fold* son sin intersección, cubren todo el conjunto de datos (todos los patrones son eventualmente parte del entrenamiento y del test) y se generan de forma desordenada.
- En cada entrenamiento, el conjunto destinado a entrenar es separado nuevamente en un 0,10 del total para validación y un 0,65 para entrenamiento (a fin de contar, por ejemplo, con datos para monitorear el *early-stopping*).

- Desde luego, los experimentos sobre cada *fold* se realizan de forma aislada, evitando la transferencia de conocimiento entre iteraciones (a fin de mantener los conjuntos de prueba como desconocidos, evitando el *target leakage*).

Cabe destacar que el valor de  $K$  fue escogido principalmente con el objetivo de obtener conjuntos de prueba de un tamaño razonable, en cuanto experimentos preliminares mostrarían que usar conjuntos muy pequeños para la evaluación daría lugar a resultados de calidad arbitraria (fuertemente dependientes con la complejidad de los patrones en evaluación). Por otro lado, valores más altos de  $K$  implicarían realizar más veces cada experimento, lo cual aumentaría significativamente los tiempos de cómputo necesarios.

## 4.6. Métricas de evaluación

Por último resulta necesario especificar las métricas a barajar para la evaluación de los resultados. Para ello se considerarán dos tipos de métricas: aquellas basadas en ejemplos y aquellas basadas en etiquetas.

### 4.6.1. Métricas basadas en ejemplos

Las métricas basadas en ejemplos entregan una noción de qué tan bien se asignan las etiquetas de cada patrón, calculando algún tipo de puntaje para cada uno y posteriormente dividiendo por el total de patrones. La selección es la siguiente:

#### *Hamming Score*

La primera de las métricas a presentar corresponde al *Hamming Score*, nombre que recibe por su relación con la distancia de Hamming. Expresado como la intersección de las etiquetas predichas y las reales sobre la unión de las mismas, *Hamming Score* toma un valor de 1 en caso de una predicción perfecta, entre 0 y 1 si hay alguna etiqueta incorrectamente predicha (*prediction error*) o faltante (*missing error*) y 0 en caso de que ninguna etiqueta sea correctamente asignada. Su formulación matemática se presenta en la Ecuación 4.1.

$$\text{Hamming-Score} = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \quad (4.1)$$

Cabe mencionar que la inclusión de esta métrica responde a la necesidad de cuantificar la calidad de las predicciones teniendo en consideración que tanto el no asignar etiquetas reales como el asignar etiquetas incorrectas representan fenómenos indeseables, mas que su eventualidad no implica que toda la asignación de etiquetas para el patrón en estudio sea incorrecta. Esto en razón de que, si bien existen otras métricas tales como el *Accuracy* que reconocen también el error de predicción y de etiquetas faltantes, resultan demasiado estrictas pues, ante cualquier error, es el patrón completo el cual se considera como incorrectamente predicho (pese a que quizás el 90 % de las etiquetas asignadas son correctas). *Hamming Score*, en cambio, introduce la posibilidad de contar con salidas parcialmente correctas, lo cual lo convierte en una métrica más apropiada para el aprendizaje multietiquetado.

### ***Exact Match Ratio***

Si bien en la justificación de *Hamming Score* se comentó sobre la importancia de incorporar salidas parcialmente correctas en el aprendizaje multietiqueta, el si hay alguna fracción de patrones cuya asignación sea perfecta no deja de ser una pregunta por lo menos interesante. Por esta razón, se incluye también en el estudio la métrica *Exact Match Ratio*, extensión al problema multietiqueta del *Accuracy* tradicional y cuya formulación matemática se presenta en la Ecuación 4.2:

$$\text{Exact Match Ratio} = \frac{1}{|D|} \sum_{i=1}^{|D|} I(Y_i = Z_i) \quad (4.2)$$

donde  $I$  representa la función indicatriz.

### ***K-Match Ratio***

Las métricas anteriores servirán para cuantificar el nivel de acierto o de error de los distintos métodos, sin embargo, no aportan información con respecto al tipo de error presente. Dicha situación, en adición al supuesto S2 (mínimo aceptable de una asignación correcta), llevaría a la propuesta de una nueva medida, basada en el *Exact Match Ratio* pero mucho más relajada. Se trata del *K-Match Ratio*, una métrica parametrizada por el valor  $K$  cuya formulación se presenta a continuación (Ecuación 4.3):

$$K\text{-Match Ratio} = \frac{1}{|D|} \sum_{i=1}^{|D|} I(|Y_i \cap Z_i| \geq \min(|Z_i|, K)) \quad (4.3)$$

De esta forma, *K-Match Ratio* se plantea como una métrica a nivel de patrones que, para cada patrón, toma valor 1 si es que la asignación de etiquetas es de al menos  $K$  etiquetas correctas o bien incorpora todas las etiquetas reales en el caso de que la cardinalidad de dicho conjunto sea menor a  $K$ , y luego promedia sobre todos los patrones (asignando un valor 0 a aquellos en que la condición no se cumpla).

Se espera que *K-Match Ratio* permita responder a la pregunta “¿A qué fracción de los patrones se le está asignando al menos  $K$  etiquetas correctas?”, indicando así qué tan cerca o lejos se encuentran los resultados de satisfacer lo enunciado en S2 al probar con  $K = 1$  y, al mismo tiempo, contribuir a la detección del error por ausencia de etiquetas correctas, al probar con valores de  $K$  cercanos al promedio de etiquetas por patrón.

#### **4.6.2. Métricas basadas en etiquetas**

Las métricas basadas en etiquetas, por otro lado, son aquellas en donde el puntaje no se calcula sobre los patrones sino sobre cada etiqueta, y luego el promedio se calcula sobre el total de etiquetas. Su uso responde a la necesidad de entender cuántas etiquetas se están aprendiendo de manera apropiada (PG2), alcanzando así mayor entendimiento de la eventual utilidad de la herramienta de apoyo con patrones distintos, pero cuyas etiquetas respondan a la misma taxonomía.

## **Recall**

La primera métrica corresponde al *Recall*, entendida como la fracción de etiquetas reales correctamente predichas. Su formulación matemática, definida en términos de correctos positivos (TP), correctos negativos (TN), incorrectos positivos (FP) e incorrectos negativos (FN) se presenta a continuación (Ecuación 4.4):

$$Recall = \frac{TP}{TP + FN} \quad (4.4)$$

## **Precision**

Se utilizará también la métrica de Precision, la cual reporta la fracción de etiquetas predichas correctamente como positivas. Su fórmula se presenta en la Ecuación 4.5:

$$Precision = \frac{TP}{TP + FP} \quad (4.5)$$

## **$F_\beta$ -Score**

Por último se incorporará la métrica  $F_\beta$ -Score, la cual corresponde a un promedio de *Recall* y *Precision* donde en particular *Recall* recibe una relevancia de  $\beta$  veces *Precision*. De esta manera,  $F_1$ -Score corresponde a la media armónica entre ambas métricas, y  $F_2$ -Score a aquella en donde el *Recall* pondera el doble que la *Precision*, escenario que podría resultar de interés dado S4. La fórmula general de  $F_\beta$ -Score se presenta en la Ecuación 4.6:

$$F_\beta\text{-Score} = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall} \quad (4.6)$$

Un punto importante a mencionar sobre las métricas seleccionadas para el caso basado en etiquetas es la estrategia con la cual se obtendrán los indicadores TP, TN, FP y FN. A priori, existen cuatro maneras de proceder:

- El cálculo basado en ejemplos (*Samples*), según el cual los indicadores se calculan para cada patrón, se obtiene la métrica por patrón y posteriormente se promedia sobre todos los patrones. Esta aplicación corresponde a un caso de métrica basada en ejemplos, y no será de interés para la investigación.
- El cálculo *Micro*, en donde los indicadores se calculan de forma global (contando TP, TN, FP y FN sobre todas las predicciones) para posteriormente calcular la métrica, la cual obtiene la propiedad de robustez ante el desbalance de etiquetas.
- El cálculo *Macro*, que corresponde al caso en que la métrica se obtiene primero por cada etiqueta y posteriormente se calcula el promedio simple de las mismas. En este caso la métrica resultante pierde robustez ante el desbalance, pues todas las etiquetas ponderan por igual en el resultado (sin importar su frecuencia).
- El cálculo *Weighted*, que corresponde al caso *Macro* pero utilizando un promedio ponderado por el soporte de las etiquetas (la cantidad de instancias positivas) para dotar a la métrica de robustez ante el desbalance.

Llevando esta situación al desafío en cuestión, la decisión más natural es la de utilizar las estrategias *Micro* o *Weighted*, pues ofrecen robustez ante el desbalance (un aspecto ampliamente presente en los datos). Aún más, si bien se reconoce que las etiquetas menos frecuentes son probablemente las más informativas, usar una estrategia *Macro* daría lugar a métricas demasiado exigentes, pues la cantidad de etiquetas poco frecuentes supera ampliamente a las más frecuentes (el 80.12 % de las etiquetas no presenta más de 3 eventos en los datos).

En consideración de lo anterior, la estrategia escogida para los experimentos es la estrategia *Micro*. Su elección por sobre *Weighted* responde a que, si bien no se busca tratar por igual a las etiquetas más y menos frecuentes pues las últimas pueden resultar extremadamente difíciles de aprender, el darles una ponderación basada en su soporte resultaría en una infravaloración de las mismas que tampoco es lo deseado.

#### 4.6.3. Relevancia e interpretación

Ante la variedad de métricas presentadas, cabe preguntarse cómo estas se interpretarán y en particular cómo se decidirá la superioridad de un modelo por sobre otro. Al respecto:

- La métrica principal para escoger modelos será el *F<sub>2</sub>-Score*, a fin de considerar tanto *Precision* como *Recall* pero dando más importancia a este último. Tales métricas también serán estudiadas de forma individual, pero solo para alcanzar una mayor comprensión de lo que el *F<sub>2</sub>-Score* refleje.
- La elección de una métrica principal basada en etiquetas y no en patrones responde al interés por alcanzar un modelo que generalice a patrones distintos de los del libro, pero que guarden una taxonomía de etiquetado similar.
- *Hamming Score*, por otro lado, servirá como complemento al *F<sub>2</sub>-Score* desde la vereda del éxito por patrón, lo cual permitirá entender en mayor profundidad los resultados obtenidos al preguntarse, por ejemplo, si los cambios en una métrica se reflejan también en la otra y con qué magnitud. *Exact Match Ratio* y *Hamming Loss*, por su parte, entregarán información adicional para el análisis de los mejores modelos.
- Finalmente, *K-Match Ratio* se utilizará como una herramienta para la traducción de resultados desde métricas más abstractas hacia afirmaciones simples del estilo “el modelo asigna correctamente K etiquetas en el X % de los casos”, entendiendo que son dichas afirmaciones las que con mayor probabilidad despertarán interés en posibles usuarios del sistema final.

# Capítulo 5

## Preparación de los datos

En este capítulo se presentan los procesos realizados sobre los datos de forma previa a la experimentación y con el fin de alcanzar un escenario óptimo para la misma. Para ello, la Sección 5.1 comienza presentando algunos aspectos descubiertos al momento de extraer las etiquetas y cómo estos serían abordados, para posteriormente presentar las distintas técnicas de creación de datos sintéticos en la Sección 5.2.

### 5.1. Extracción y procesamiento de etiquetas

#### 5.1.1. Selección de idioma

Como se comentase en la Sección 2.2, el libro a partir del cual se construye el conjunto de datos presenta las etiquetas de cada patrón en cinco idiomas: alemán, inglés, francés, italiano y griego (véase Tabla 5.1). Esto llevaría al cuestionamiento de qué idioma sería más conveniente para la extracción, sin embargo, era de esperar que las etiquetas guardasen el mismo significado entre un idioma y otro, por lo que la decisión debía ser más bien irrelevante.

Tabla 5.1: Etiquetas del patrón 17e, pág. 41 [28].

Idioma	Etiquetas
Alemán	Blattrosette, sechsfach, schraffiert oder doppelt, in einer Metope
Inglés	Hexafoil, hatched or doubled, metopal
Francés	Rosette à 6 pétales, hachurée ou doublée, dans une métope
Italiano	Fiore a sei foglie, campito a tratteggi o a contorno doppio, in una metopa
Griego	Εξαψυλλος ρόδακας, διαγραμμια η διπλα πεταλα, εντός μετόπης

Sin embargo, al estudiar las etiquetas en detalle se descubrirían algunas incongruencias entre idiomas que refutarían tal hipótesis, dotando de importancia a la elección de idioma. En particular, se detectarían dos problemas críticos que estarían contribuyendo al bajo valor

de LD y por ende requerirían tratamiento, siendo estos la existencia de etiquetas correspondientes a un mismo concepto pero en distinto género gramatical, y de etiquetas compuestas en ciertos idiomas que se presentarían de forma individual en otros.

Por un lado, la presencia de ambos géneros gramaticales para un mismo concepto se habría detectado al estudiar las etiquetas en francés, idioma que junto al inglés conformaría los principales candidatos para la extracción dada su comprensión por parte del autor. La razón por la cual este fenómeno representaría un obstáculo guardaría relación con el hecho de que, al tener, por ejemplo, las etiquetas “separadas por triángulos negros” y “separados por triángulos negros” la cantidad de etiquetas aumentaría innecesariamente (una unidad más de lo necesario) y la frecuencia de lo que debía ser una única etiqueta se veía repartida en dos, agregando aún más dificultad a un problema que por sí ya se presentase como bastante complejo. Las Tablas 5.2 y 5.3 ilustran dicho fenómeno con las etiquetas en inglés y francés de los patrones 7b y 7c, respectivamente.

Tabla 5.2: Etiquetas del patrón 7b, pág. 17 [28].

Idioma	Etiquetas
Inglés	Opposed diagonals, separated by solid triangles, horizontal panel
Francés	Traits obliques affrontés, séparés par des triangles noirs, panneau horizontal

Tabla 5.3: Etiquetas del patrón 7c, pág. 17 [28].

Idioma	Etiquetas
Inglés	Opposed diagonals, separated by solid triangles, vertical panel
Francés	Traits obliques affrontés, séparées par des triangles noirs, panneau vertical

Nótese que, mientras en el patrón 7b figura la etiqueta *séparés par des triangles noirs*, en el caso del patrón 7c dicha etiqueta se presenta como *separées par des triangles noirs*.

Tal situación posicionaría al inglés como el idioma a escoger, en cuanto la propiedad de ser un lenguaje neutro al género aseguraba que esta dificultad no se haría presente. Sin embargo, al observar en más detalle las etiquetas en inglés se descubriría la segunda discrepancia entre lenguajes, relacionada con la existencia de etiquetas que se presentan de manera compuesta o individual según el idioma. A modo de Ejemplo, véase la cantidad de comas (interpretadas para la investigación como separador de etiquetas) por línea en la Tabla 5.4:

En particular, en el ejemplo anterior se identificarían tres formas distintas de etiquetar la misma característica: “Lágrimas, colgantes en grupos” para el alemán, “Lágrimas, colgantes, en grupos” para el inglés e italiano, y “Grupos de lágrimas colgantes” para el francés (se desconoce la traducción desde el griego). Si bien estos casos serían aislados, la relación entre idiomas se mantendría.

La pregunta entonces sería con qué nivel de atomicidad para las etiquetas convendría trabajar, pues si bien el significado global se mantendría, la forma en la que estas se presentarían para el aprendizaje sería diametralmente opuesta: mientras las etiquetas compuestas guardarían mayor expresividad al tratarse principalmente de oraciones, sufrirían de contar con frecuencias muy bajas por la misma razón (a más palabras, más posibilidad de que el uso

Tabla 5.4: Etiquetas del patrón 1b, pág. 3 [28].

Idioma	Etiquetas
Alemán	Tropfen, in Gruppen hängend, Horizontalfeld
Inglés	Teardrops, pendent, in groups, horizontal panel
Francés	Groupes de gouttes pendants, panneau horizontal
Italiano	Gocce, pendule, in gruppi, campo orizzontale
Griego	Εξαψυλος ρόδακας, διαγραμμια η διπλα πεταλα, εντός μετόπης

de un sinónimo de lugar a una nueva etiqueta) y, por otro lado, las etiquetas singulares gozarían de una mayor frecuencia, pero al costo de significados eventualmente interdependientes (como sucede con la etiqueta “en grupos” en el ejemplo anterior, la cual cobra significado al presentarse junto a “lágrimas”).

Finalmente, la decisión fue la de seleccionar el idioma inglés para la extracción de etiquetas, optando entonces por trabajar con las versiones más atómicas. De esta manera, se contaría con:

1. Neutralidad al género gramatical,
2. Una cantidad menor de etiquetas infrecuentes, al costo de menor expresividad,
3. Etiquetas en un idioma comprensible por el autor, lo cual resultaría importante para la comprensión de los resultados (y en especial para corregir implementaciones erróneas).

Cabe mencionar que medir el impacto exacto de dicha decisión habría requerido de la recopilación de las etiquetas en todos los idiomas para posteriormente realizar los cálculos respectivos, lo cual habría significado usar las horas de varias semanas solo para este fin (la extracción en inglés tomaría alrededor de 30 horas), por lo cual se opta por asumir la conveniencia aún sin valores exactos que la respalden.

### 5.1.2. Normalización mediante NLP

Una vez realizada la extracción de las etiquetas, el siguiente paso correspondería al procesamiento de las mismas con el fin de obtener un conjunto más apropiado para el aprendizaje. Las razones que motivarían a este procesamiento serían, en primer lugar, el hecho de que aún restarían algunas etiquetas con frases largas, de baja frecuencia y compuestas por una o más etiquetas individuales y, en segundo lugar, la presencia de etiquetas con preposiciones o sinónimos complementando una idea base, lo cual si bien contribuiría con mayor expresividad, daría lugar a etiquetas únicas.

A modo de ejemplo, considérese la primera etiqueta en idioma inglés del patrón 31d (Tabla 5.5): “*Meander-like pattern*”, la cual aparece una única vez en todo el conjunto de datos. Sería razonable asumir que el uso de complementos al concepto principal podría deberse a que el patrón en cuestión no calza al cien por ciento con *Meander*, lo cual justificaría su uso. Sin embargo, el costo de expresar dicha situación es el de separar esta etiqueta de los más de 20

casos de “*Meander*” como etiqueta singular, agregando mucha más dificultad a su aprendizaje (en comparación al de la etiqueta singular) y contribuyendo ampliamente al bajo valor de LD, en cuanto agrega un único caso positivo pero al costo de  $|D| - 1$  casos negativos.

Tabla 5.5: Etiquetas del patrón 31d, pág. 77 [28].

Idioma	Etiquetas
Ingles	Meander-like pattern, hatched, horizontal panel

Similar es el caso de la etiqueta “*Meander and battlement pattern*”, que reúne en una misma oración dos etiquetas singulares y se hace presente solo en los patrones 27c y 27d. Vale destacar que estas etiquetas no solo dificultan el aprendizaje en el sentido de agregar una etiqueta difícil y aumentar el LD, sino también reduciendo la potencial frecuencia de las etiquetas atómicas (en este caso, *Meander* y *Battlement*, las cuales podrían cada una tener dos entradas más).

Dichos aspectos serían abordados desde un enfoque de Programación de Lenguaje Natural (NLP), en particular a través de la aplicación de dos técnicas: la remoción de “palabras vacías” (*Stopwords removal*) en las etiquetas (a fin de quitar, por ejemplo, el *and* de *Meander and battlement pattern*), y la posterior lematización de las mismas, a fin de extraer los conceptos principales (que para el ejemplo corresponderían a *meander*, *battlement*, *pattern*). Tales tareas se llevarían a cabo mediante las librerías de Python *Gensim* [40], que contribuiría en la remoción, y *Natural Language Toolkit* (NLTK) [32], a través de la clase *WordNetLemmatizer*. La Tabla 5.6 muestra algunos ejemplos de etiquetas originales y su estado final.

Tabla 5.6: Ejemplos de etiquetas y su procesamiento

Patrón, Referencia	Etiquetas originales	Etiquetas tras Stopwords Removal	Etiquetas tras Lemmatization
31d pág. 41	meander-like pattern, hatched, horizontal panel	meander pattern, hatched, horizontal panel	meander, pattern, hatched, horizontal, panel
47d pág. 119	triangles, with bars at apice, hatched, horizontal panel	triangles, bars apice, hatched, horizontal panel	triangle, bar, apice, hatched, horizontal, panel
53f pág. 135	alternating antithetic rays, often accompanied by dots, horizontal panel	alternating antithetic rays, accompanied dots, horizontal panel	alternating, antithetic, ray, accompanied, dot, horizontal panel
68a pág. 173	scribble (“tadpoles”), vertical, between bars, horizontal panel	scribble tadpole, vertical, bars, horizontal panel	scribble, tadpole, vertical, bar, horizontal, panel
71a pág. 183	horizontal panel, concentric circles	horizontal panel, concentric circles	horizontal, panel, concentric, circle

Cabe destacar que, si bien el procesamiento llevado a cabo contribuiría a aumentar el valor de LD del conjunto de datos y por ende a atacar el principal problema detectado, se reconoce que la última etapa (correspondiente a la lematización) también habría dado lugar a etiquetas interdependientes - como *panel* y *vertical* u *horizontal* - y generado otras sin

un significado relevante, tales como *accompanied*. Sin embargo, el primer caso no debería significar una dificultad mayor, puesto que el descubrimiento de relaciones entre etiquetas es parte de las bondades ofrecidas por los algoritmos más avanzados. El segundo caso, por otra parte, podría requerir un procesamiento más fino, como por ejemplo la inclusión manual de dichas etiquetas en las palabras a remover (*stopwords*). No obstante, a fin de mantener la investigación lo más replicable posible, se opta por conservar la lista de palabras original y permitir que dichas etiquetas permanezcan en los datos, siendo entonces uno de las principales desventajas del procesamiento realizado junto a la pérdida de expresividad en las etiquetas, y representando así una posible variable para futuras investigaciones.

Más en específico, el efecto del procesamiento con respecto a las cantidades de interés en los datos fue:

1. Un cambio del número de etiquetas desde 585 a 330 (reducción del 43,58%),
2. Un aumento del *Label Density* a 0,0161, más de tres veces su valor original (0,0051),
3. Un aumento del promedio de etiquetas por patrón desde 3,0309 a 5,3144.

Los dos primeros puntos responden directamente al objetivo del procesamiento, a saber, conseguir un conjunto refinado de etiquetas con mejores características para el aprendizaje. El tercer punto, por otra parte, se trataría de un efecto no deseado pero que no debería representar un obstáculo mayor, más que deberá ser considerado al momento de analizar los resultados en cuanto cambia lo expresado en S2 sobre el promedio de 3 etiquetas por patrón, que se usase para justificar el mínimo aceptable de una etiqueta correctamente predicha.

## 5.2. *Data Augmentation*

El último proceso a realizar antes de proceder con los experimentos correspondería a la creación de datos sintéticos que complementasen a los originales, a fin de incluir más variedad y de esta manera facilitar el aprendizaje. Tal tarea sería además el primer paso a seguir para encontrar respuesta a PG3.

Para ello se optaría por crear múltiples funciones que, tras recibir la imagen de un patrón original como entrada, creasen una versión modificada de este bajo alguna transformación afín, para posteriormente asociar la imagen sintética con las mismas etiquetas del patrón original. Tal proceder requeriría, sin embargo, añadir el supuesto de que para cada una de las técnicas de D.A. a implementar, su aplicación no transgrede con la semántica asociada a cada etiqueta del patrón original ni da lugar a etiquetas nuevas.

### 5.2.1. Técnicas

A continuación se presenta cada una de las transformaciones realizadas, su descripción y ejemplos. Cabe mencionar que mientras parte de las transformaciones fue implementada manualmente a través de la librería Python *OpenCV*, para otras se descansó en las implementaciones disponibles en la librería *imgaug* [26], lo cual se indica según corresponda.

## Rotación

Corresponde a la aplicación de rotaciones en 90, 180 y 270 grados de la imagen original. La transformación toma con especial cuidado aquellos patrones con las etiquetas *horizontal* y/o *vertical*, casos en los cuales solo se aplica la rotación en 180 grados.

La Figura 5.1 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 15b del libro de Kunisch (pág. 35 [28]).

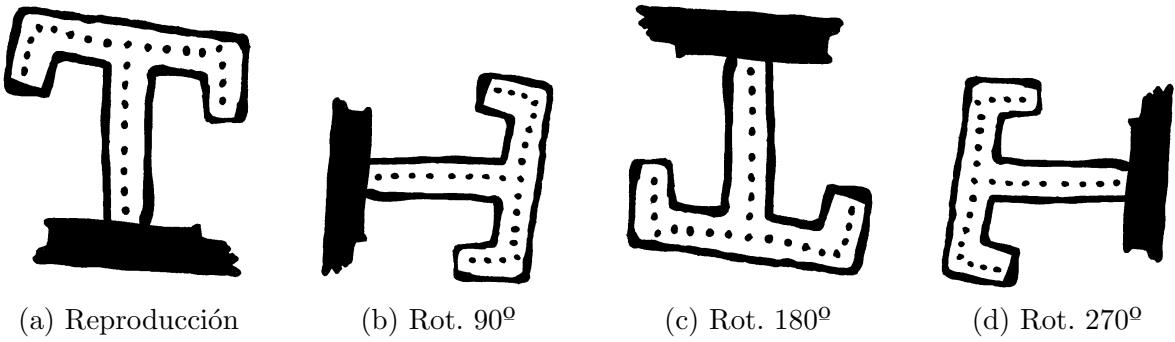


Figura 5.1: Patrón 15b, aumentado con Rotación

## Reflexión

Corresponde a la aplicación de reflexiones con respecto a los ejes X e Y de cada imagen. La Figura 5.2 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 6e del libro de Kunisch (pág. 14 [28]).

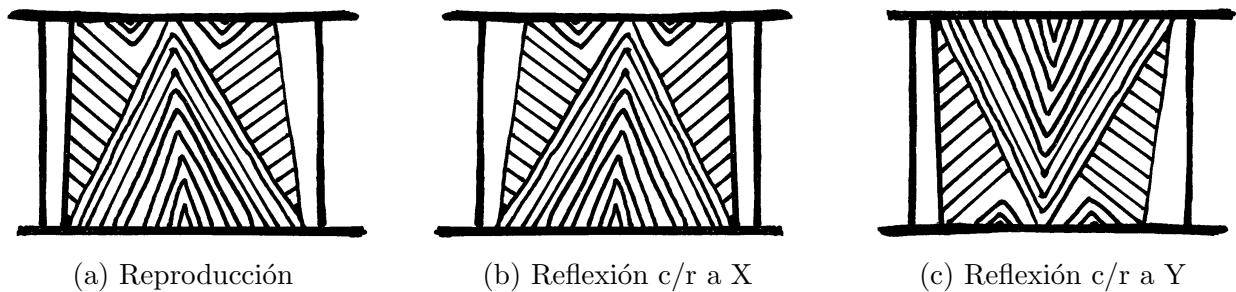


Figura 5.2: Patrón 6e, aumentado con Reflexión

## Lluvia

Corresponde a la aplicación de líneas blancas y finas a modo de “lluvia” sobre la imagen original. Se utiliza la implementación de *imgaug*.

La Figura 5.3 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 95l del libro de Kunisch (pág. 246 [28]).



(a) Reproducción



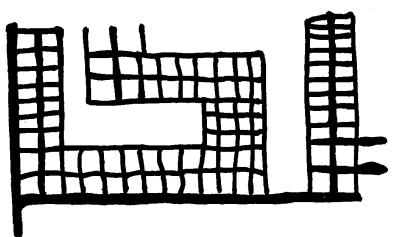
(b) Lluvia

Figura 5.3: Patrón 95l, aumentado con Lluvia

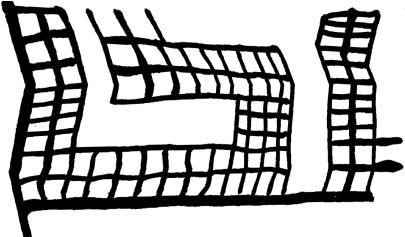
### Distorsión elástica

Implementado a través de *imgaug*, corresponde a la aplicación de una malla regular de puntos sobre la imagen para posteriormente mover vecindarios de puntos bajo transformaciones afines a trozos, generando distorsiones locales.

La Figura 5.4 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 21f del libro de Kunisch (pág. 52 [28]).



(a) Reproducción



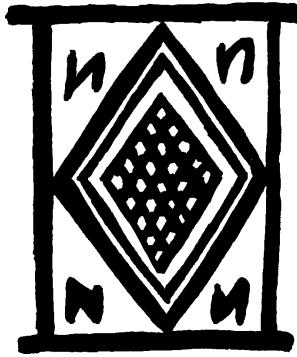
(b) Distorsión elástica

Figura 5.4: Patrón 21f, aumentado con Distorsión elástica

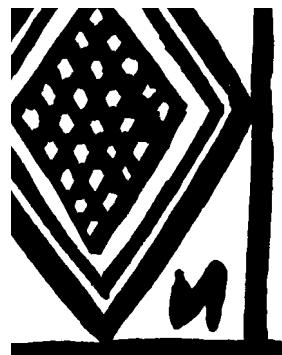
### Recorte

Corresponde a la extracción de una parte de la imagen original bajo la restricción de mantener al menos la mitad del ancho y largo. La restricción se incluye al reconocer que el recorte es la transformación con más probabilidad de transgredir semánticas relacionadas con etiquetas (por ejemplo, encuadres) y por ende se agrega para limitar a que al menos la mitad de la imagen permanezca.

La Figura 5.5 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 44a (pág. 113 [28]).



(a) Reproducción



(b) Recorte

Figura 5.5: Patrón 44a, aumentado con Recorte

### Difuminado (*blur*)

Implementado también mediante *imgaug*, esta transformación se presenta de tres formas: la primera corresponde a la aplicación de un *kernel* gaussiano con un sigma de distribución uniforme entre 2 y 10 sobre la imagen. La segunda, por otro lado, se trata de la aplicación de un promedio sobre vecindarios de tamaño aleatorio entre (2, 2) y (10, 10). Por último, la tercera forma corresponde a la aplicación de un filtro que simula el movimiento de la cámara u objeto, en este caso con un *kernel* de tamaño (15, 15).

La Figura 5.6 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 70m del libro de Kunisch (pág. 180 [28]). En particular, 5.6b presenta una aplicación de filtro gaussiano (*Gauss Blur*) mientras que 5.6c corresponde al caso de *blur* promedio (*Average Blur*) y 5.6d a *blur* por movimiento (*Motion Blur*).



(a) Reproducción



(b) *Gauss Blur*



(c) *Average Blur*



(d) *Motion Blur*

Figura 5.6: Patrón 70m, aumentado con Difuminado

### RandAugment

Corresponde a la aplicación de alguna(s) de las transformaciones aleatorias propuestas en RandAugment [12], las cuales mostraron contribuir a los resultados de arquitecturas competentes en *datasets* como SVHN e ImageNet. En particular, se usa la implementación disponible en *imgaug* y se parametriza por  $n = 2$  y  $m = 9$ , lo cual corresponde a aplicar dos

transformaciones por imagen (una tras la otra) y con una fuerza de 9 (entre un mínimo de 0 y máximo de 30), coincidiendo así con los parámetros recomendados por los autores para abordar ImageNet.

La Figura 5.7 muestra un ejemplo de dicha transformación aplicada sobre la reproducción del patrón correspondiente a la entrada 11h del libro de Kunisch (pág. 35 [28]).

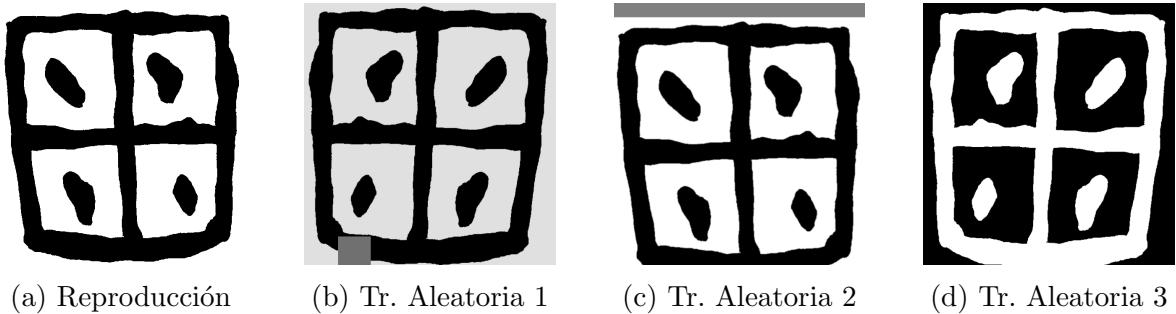


Figura 5.7: Patrón 15b, aumentado con *RandAugment*

### 5.2.2. Conjuntos de datos

Una vez descritas las técnicas de aumentado a considerar, sigue el detalle de cómo estas se usarán para los experimentos. Al respecto:

- Las técnicas de aumentado solo se aplicarán en los datos de entrenamiento de cada *fold*.
- Las técnicas se utilizarán, en primera instancia, de manera individual (una técnica a la vez) y luego de forma aditiva hasta llegar a un conjunto que las considere todas.

De esta forma, distintos conjuntos de datos son construidos:

- Conjunto *base*, denominación bajo la cual se referirá a los patrones originales y que formará parte de todos los otros conjuntos.
- Conjuntos *rotación*, *reflexión*, *recorte*, *elástico*, *lluvia* y *random*, para referirse al conjunto resultante de aplicar cada una de las transformaciones con dichos nombres por separado, sobre el conjunto *base*.
- Conjuntos *gauss blur*, *avg. blur* y *mtn. blur*, para referirse a los conjuntos relativos a cada una de las técnicas de difuminado.
- Cada uno de los conjuntos resultantes de agregar las técnicas en el orden en que fueron presentadas, con la salvedad de utilizar solo *avg. blur* para el caso del difuminado. Por ejemplo: *rotación+reflexión*, *rotación+reflexión+lluvia*, etc.

Cabe mencionar que la decisión de seleccionar solo el difuminado mediante *average blur* responde principalmente a limitaciones de cómputo, en cuanto los conjuntos más grandes (aquellos que reúnen múltiples técnicas) incorporan en su mayoría sobre 5000 imágenes y

dan lugar a entrenamientos que, según el modelo, requieren desde unas pocas horas hasta días de procesamiento.

A cada conjunto de datos se le asignará un identificador numérico único, a fin de economizar el espacio que listar todas las técnicas que conforman un conjunto agregado puede requerir. Tales identificadores se presentan en la Tabla 5.7, junto a la cantidad final de imágenes para entrenamiento por conjunto según cada *fold*. Nótese que si bien las cantidades entre *folds* son similares, no son necesariamente las mismas para todas las técnicas, pues la condición sobre la técnica de *rotación* da lugar a más o menos datos sintéticos según los patrones a los que se vea enfrentada (a fin de respetar las etiquetas *horizontal* y *vertical*).

Tabla 5.7: Cantidad de imágenes para entrenamiento por *fold* según conjunto de datos

Conjunto	Identificador	Fold 1	Fold 2	Fold 3	Fold 4
<i>base</i>	0	504	504	504	504
<i>rotación</i>	1	1532	1538	1514	1506
<i>reflexión</i>	2	1512	1512	1512	1512
<i>recorte</i>	3	1008	1008	1008	1008
<i>elástico</i>	4	1008	1008	1008	1008
<i>lluvia</i>	5	1008	1008	1008	1008
<i>random</i>	6	1008	1008	1008	1008
<i>gauss blur</i>	7	1008	1008	1008	1008
<i>avg. blur</i>	8	1008	1008	1008	1008
<i>mtn. blur</i>	9	1008	1008	1008	1008
<i>rotación + reflexión</i>	10	2540	2546	2522	2514
<i>rotación + reflexión + lluvia</i>	11	3044	3050	3026	3018
<i>rotación + reflexión + lluvia + elástico</i>	12	3548	3554	3530	3522
<i>rotación + reflexión + lluvia + elástico + avg. blur</i>	13	4052	4058	4038	4026
<i>rotación + reflexión + lluvia + elástico + avg. blur + recorte</i>	14	4556	4562	4538	4530
<i>rotación + reflexión + lluvia + elástico + avg. blur + recorte + random</i>	15	5060	5066	5042	5034

Desde luego, los conjuntos de test no se someten al proceso de D.A. y cuentan con un total fijo de 194 patrones.

# Capítulo 6

## Enfoque Tradicional

Este capítulo incorpora la exploración del problema desde un enfoque de MLC Tradicional, es decir, a través de métodos desarrollados para el caso con una cantidad reducida de etiquetas y un amplio número de ejemplos de entrenamiento. En primer lugar, se presenta el detalle técnico de los métodos a utilizar (Sección 6.1) y los descriptores a entregar como entrada de los mismos (6.2). Luego se presentan los experimentos y sus resultados, además de un breve análisis de los mismos (6.3).

### 6.1. Técnicas

Como se listase en la Selección de Métodos (Sección 4.4.1), el Enfoque Tradicional se aborda a través de dos grandes familias de técnicas: aquellas de transformación del problema y de adaptación del algoritmo. Dichas familias se detallan a continuación.

#### 6.1.1. Transformación del Problema

##### *Binary Relevance* (BR)

Corresponde a la separación del problema multietiqueta con  $N$  etiquetas en  $N$  problemas independientes de clasificación binaria.

*Binary Relevance* se favorece de ser un acercamiento simple, mas presenta importantes desventajas tanto en cuanto a las consecuencias de tratar las etiquetas por separado (obviando eventuales relaciones entre las mismas) y a la necesidad de entrenar un clasificador base por cada una, lo cual limita su aplicación a conjuntos con una cantidad limitada de etiquetas.

Este método no presenta parámetros salvo por el o los clasificadores base a utilizar. Nótese que si bien podrían usarse clasificadores distintos para cada etiqueta, la implementación sobre la cual se conducirán los experimentos (aquella de *scikit-multilearn*) limita este aspecto a un único clasificador base por experimento.

## ***Classifier Chain* (CC)**

*Classifier Chain* se propone como un método que busca dar solución a la no explotación de relaciones entre etiquetas presente en BR a través de un acercamiento de predicción secuencial: el clasificador  $i$  recibe la entrada original concatenada con las salidas de los clasificadores anteriores, tras lo cual realiza la clasificación teniendo conocimiento de las etiquetas ya asignadas o descartadas.

Si bien el objetivo de incorporar relaciones entre etiquetas se cumple, la calidad de los resultados adquiere una importante dependencia con el orden en el que las etiquetas se presenten para clasificación, pues el error de un clasificador se arrastra para todos los que le sigan. Para lidiar con este aspecto, se recomienda utilizar la regla de “Las etiquetas más frecuentes primero” en cuanto estas son en general las más fáciles de asignar.

Por otro lado, la desventaja con respecto a los costos de BR se mantiene y empeora, en cuanto ya no solo se requiere entrenar  $N$  clasificadores, sino que además se agrega la restricción de que el entrenamiento debe ser secuencial (sin posibilidad de paralelismo).

Este método mantiene el único parámetro de BR, a saber, el tipo de clasificador a utilizar como clasificador base.

## ***Label Powerset* (LP)**

A diferencia de los métodos anteriores que transforman el problema multietiqueta a múltiples problemas binarios, *Label Powerset* aborda el desafío mediante la transformación a un único problema multiclasificación, donde cada clase es una combinación de etiquetas presente en los datos.

Al estudiar las combinaciones de etiquetas como el objetivo de clasificación, LP adquiere la propiedad de explotar las relaciones más superficiales entre etiquetas (en particular, la co-ocurrencia). Sin embargo, guarda dos importantes desventajas: por un lado, la necesidad no solo de que las etiquetas tengan alta frecuencia, sino también de que sus combinaciones lo hagan y, por otro lado, el hecho de contar con una nula capacidad de generalización a combinaciones no vistas de etiquetas, pues requiere que todas las combinaciones estén presentes en los datos o de lo contrario no las asignará jamás.

Nuevamente, el único parámetro de este método es el tipo de clasificador a utilizar como clasificador base.

## ***RakelD***

*RakelD* (acrónimo para *Random k-Labelsets*) busca ser una versión mejorada de LP a través del uso de particiones aleatorias del espacio de etiquetado y su posterior uso como combinaciones para el aprendizaje multiclasificación, lo cual mejoraría la generalización pero mantendría una fuerte dependencia con la probabilidad de observar las combinaciones en los datos.

En este caso, el método incorpora un parámetro adicional al clasificador base: el tamaño deseado de las particiones, cuyo valor por defecto (y recomendado por sus creadores) es 3.

### 6.1.2. Adaptación del Algoritmo

#### *Binary Relevance k-Nearest Neighbors (BRkNN)*

BRkNN corresponde a otro acercamiento de BR al problema multietiqueta pero utilizando kNN como algoritmo de decisión en lugar de clasificadores base. Para ello se genera un modelo de  $k$  vecinos más cercanos para cada etiqueta de forma independiente usando los datos de entrenamiento, modelos los cuales son posteriormente consultados uno a uno en la etapa de inferencia a fin de determinar el conjunto de etiquetas a asignar.

Las desventajas de dicho método son las mismas del método de BR, a saber, la no explotación de relaciones entre etiquetas y un alto costo computacional, esta vez por la construcción de tantos modelos kNN como etiquetas se estudien. Sin embargo, los autores sugieren que el proceso podría hacerse más eficiente a través del uso de una matriz de similitud compartida entre los distintos modelos [57].

Por otro lado, BR-kNN requiere la estimación de un único parámetro, correspondiente al valor de  $k$  para la búsqueda de vecinos más cercanos.

#### *Multi-label k-Nearest Neighbors (MLkNN)*

MLkNN es otra adaptación de kNN al etiquetado múltiple, pero incorporando el principio de *Maximum a Posteriori* para la asignación de etiquetas. Más en detalle, para cada nueva instancia, en primer lugar se identifican los  $k$  vecinos más cercanos en el conjunto de entrenamiento, tras lo cual se utiliza MAP según las etiquetas de dichos vecinos para determinar el conjunto de etiquetas de la nueva instancia. El método incorpora dos parámetros: el número  $k$  de vecinos de cada instancia a tomar en cuenta y el parámetro de suavizado  $s$ , el cual controla la fuerza de la distribución prior uniforme y toma un valor por defecto de 1.0 (dando lugar al suavizado de Laplace).

La principal desventaja del método es la de tratar cada etiqueta de forma independiente, de manera análoga a BR-kNN. Sin embargo, se añade también el requerimiento de que la distancia entre ejemplos sea un buen predictor para la asignación de etiquetas.

#### *Multi-label Twin Support Vector Machine (MLTSVM)*

El último método a evaluar es MLTSVM, adaptación del algoritmo de clasificación binaria *Twin Support Vector Machine* (TWSVM) al caso multietiqueta mediante la determinación de múltiples hiperplanos para capturar la información multietiqueta en los datos.

En entrenamiento, MLTSVM construye múltiples hiperplanos no paralelos en una estrategia *one-vs-all*, con el propósito de explotar la información multietiqueta a través de la resolución de problemas de programación cuadrática (QPPs). Cada hiperplano resultante está asociado a una etiqueta de la cual es el más cercano y es a la vez tan lejano del resto como sea posible, lo cual hace que para la predicción baste con el diseño de una función de decisión de acuerdo a las distancias entre una instancia y los diferentes hiperplanos.

Dicho procedimiento puede además ser acelerado mediante el uso de *Sucessive Overrelaxation* (SOR) para la solución de los QPPs [9].

El método incorpora múltiples parámetros, a saber:

- $c_k$ , parámetro de *penalty* para el riesgo empírico que determina el intercambio entre términos de la función de pérdida y de valor por defecto 0,
- $sor_\omega$ , parámetro de suavizado para el algoritmo SOR y de valor por defecto 1.0,
- $threshold$ , correspondiente al umbral a partir del cual se asigna una etiqueta y de valor por defecto  $1e - 6$ ,
- $\lambda$ , parámetro de regularización para la optimización de los hiperplanos,
- $max_{iteration}$ , número máximo de iteraciones a usar en SOR y de valor por defecto 500.

## 6.2. Descriptores

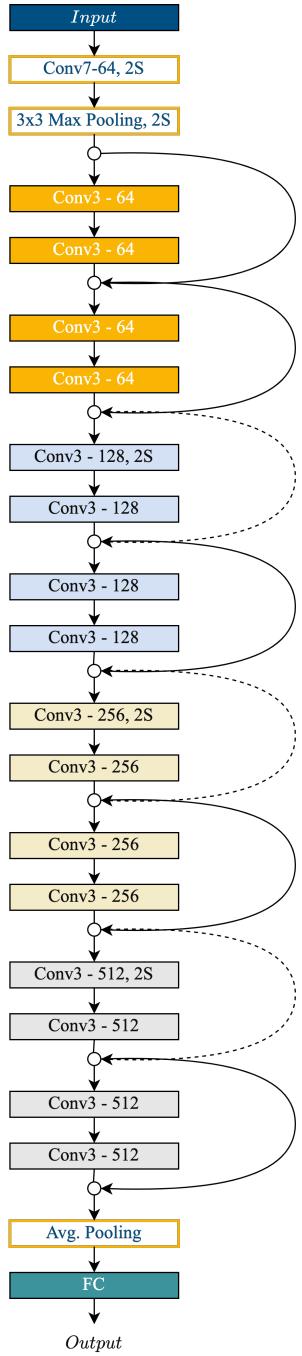
Una característica común de los métodos en estudio es la necesidad de recibir descripciones apropiadas de los datos como entrada, en cuanto son incapaces de generarlas por sí mismos. Tal fenómeno a menudo es tratado en la literatura mediante un proceso de *Feature Engineering*, en el cual se utiliza conocimiento experto para construir una serie de variables que rescatan las características más importantes de los elementos en estudio para posteriormente utilizar su codificación en atributos numéricos como descriptor para predecir alguna otra característica más compleja. El problema con tal enfoque en este caso está en que es precisamente el trabajo experto lo que se busca reducir, por lo cual añadir a los arqueólogos la tarea de generar descriptores resulta contradictorio.

Es por esta razón que, ante la ausencia de *features* previamente construidas, se opta por utilizar descripciones generadas por redes profundas entrenadas sobre una tarea similar, pero diferente: se trata de las redes convolucionales ResNet18 y ResNet50, las cuales son instanciadas con los pesos recomendados por *Pytorch* (mejor estado encontrado para la clasificación multiclase sobre ImageNet) y reentrenadas para el desafío de asignar el capítulo del libro correspondiente a cada patrón en un proceso de *fine-tuning*, a través del cual se busca ajustar las medias y varianzas de las capas de *Batch Normalization* al problema en estudio. La Figura 6.1 presenta diagramas de bloques con el detalle de dichas arquitecturas.

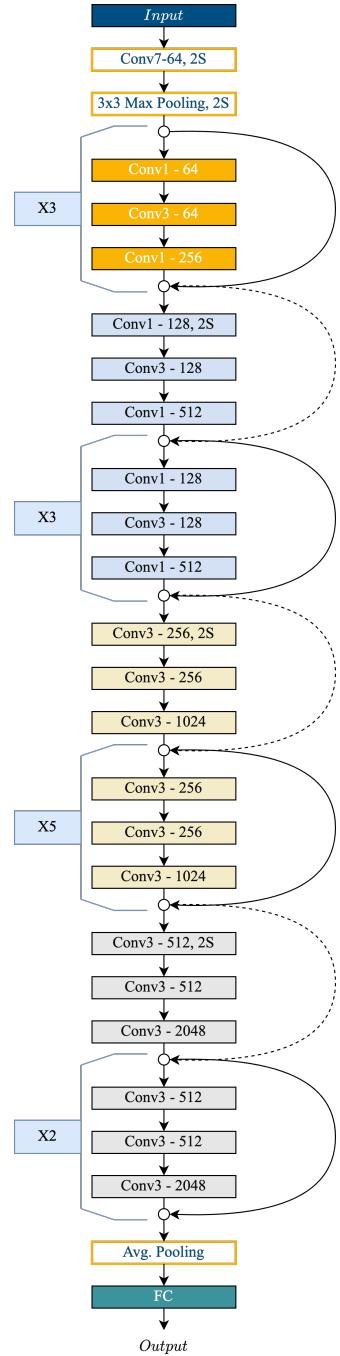
Si bien las redes convolucionales se revisitarán en mayor detalle en el Enfoque XMLC (Sección 7.1), por el momento basta considerar las arquitecturas presentadas como una alternativa al *feature engineering* para obtener descriptores con un alto nivel de abstracción.

La decisión de utilizar dichas arquitecturas responde al hecho de que la tarea para la cual se entrena habría sido el objetivo de una investigación previa, la cual hallaría que dichas configuraciones alcanzarían los mejores resultados en términos de *accuracy*, y por ende se podría esperar que sus representaciones guardasen información semántica de los patrones.

De esta manera, los descriptores se obtendrán a partir de la representación posterior a la última capa convolucional de cada arquitectura (entrada del bloque *Avg. Pooling*), correspondiente a un vector de dimensión 512 en ResNet18 y 2048 en ResNet50. Tales redes han sido reentrenadas sobre cada conjunto de datos, usando *early stopping* con una paciencia de 15 épocas, vigilando el *accuracy* y con una cantidad máxima de 300 épocas.



(a) ResNet18



(b) ResNet50.

Figura 6.1: Arquitecturas ResNet18 y ResNet50 en diagramas de bloques. **Observación:** Cada capa convolucional es seguida por una capa de *batch-normalization* y una función de activación ReLU, omitidas por economía de espacio. Una etiqueta *ConvY – X* denota un bloque de convolución con *Y* filtros de tamaño  $X \times X$ . Una etiqueta con el sufijo *NS* denota una operación con *stride* de *N*. Las flechas punteadas corresponden a conexiones residuales con ajuste de dimensionalidad mediante convolución. En el caso de ResNet50, las llaves celestes con la etiqueta *XN* rodeando conjuntos de bloques indican la repetición en secuencia de los mismos *N* veces.

La Tabla 6.1 presenta el *accuracy* en test obtenido por cada arquitectura tras ser entrenada con los distintos conjuntos. En general, todas las técnicas de D.A. contribuyen a que las redes alcancen puntajes más altos que los del caso base, salvo por recorte (que afecta negativamente a ambas arquitecturas) y *motion blur* (que disminuye levemente los resultados de ResNet50).

Tabla 6.1: *Accuracy* Arquitecturas ResNet variando técnica de D.A.

Conjunto (Identificador)	ResNet18	ResNet50
0	0.7629	0.5928
1	0.8557	0.7784
2	0.8454	0.7680
3	0.7577	0.5773
4	0.8505	0.7887
5	0.8404	0.7422
6	0.8351	0.6753
7	0.8041	0.6753
8	0.8454	0.8660
9	0.7887	0.5876
10	0.8454	0.7990
11	0.8557	0.8866
12	0.8454	0.8608
13	0.8660	0.8660
14	0.8402	0.8041
15	0.8247	0.8351

## 6.3. Experimentos

### 6.3.1. Búsqueda de hiperparámetros

Los experimentos inician utilizando el *fold* 1 para ajustar los distintos hiperparámetros de los métodos en estudio. Para ello se utiliza la función *GridSearchCV* de la librería *scikit-learn*, la cual permite determinar el mejor conjunto de hiperparámetros a partir de valores candidatos mediante búsqueda exhaustiva con validación cruzada y el uso de un *score* objetivo, en este caso el *F<sub>2</sub>-Score*.

La Tabla 6.2 lista los métodos que requirieron de dicha búsqueda, así como sus valores candidatos y finales para cada hiperparámetro.

Para el resto de los métodos (a saber, los de transformación del problema) no hay hiperparámetros a ajustar. Para sus clasificadores base, por otro lado, se opta por variar solamente la aplicación de pesos y mantener los valores por defecto para los otros parámetros.

Tabla 6.2: Búsqueda de hiperparámetros

Método	Parámetro	Valores candidatos	Valor final
BRkNN	$k$	[1, 2, 4, 8, 16, 32, 64]	1
MLkNN	$k$	[1, 2, 4, 8]	4
	$s$	[0.5, 1.0, 1.5]	1.0
MLTSVM	$c_k$	[1, 2]	2
	$sor_\omega$	[1.0, 1.5]	1.0
	$threshold$	[1e-5, 5e-5]	1e-5
	$\lambda$	[1.0, 1.5]	1.0
	$max_{iteration}$	[200, 500]	200

### 6.3.2. E1: Evolución métodos sobre base, variando tratamiento al desbalance

El primer experimento corresponde a la comparación de todos los métodos sobre el conjunto de datos original, utilizando los descriptores ResNet18 y variando la cantidad de etiquetas a predecir y el tratamiento al desbalance de los datos. Al respecto:

- La manipulación del tratamiento al desbalance resulta posible solamente para los métodos de transformación del problema, en cuanto los métodos de adaptación del algoritmo se ofrecen como naturalmente robustos ante tal fenómeno (sin depender de un parámetro u opción adicional).
- Tal manipulación se lleva a cabo mediante el uso del parámetro  $class\_weight = 'balanced'$  sobre los clasificadores base, lo cual corresponde al ajuste de pesos de manera inversamente proporcional a la frecuencia de cada clase (positiva o negativa) en cada tarea de clasificación binaria.

El experimento en cuestión tomaría alrededor de 6 horas de procesamiento en total. Cabe notar que *scikit-learn* no cuenta con las implementaciones necesarias para desarrollar experimentos utilizando procesadores gráficos, por lo cual utilizar CPU es la única opción.

El detalle de los resultados se encuentra disponible en el Apéndice A, donde las Tablas A.1 y A.2 corresponden al caso sin y con tratamiento al desbalance, respectivamente. A continuación se estudiarán dichos resultados a través de visualizaciones.

La Figura 6.2 muestra el *Micro F<sub>2</sub>-Score* obtenido por cada método a través de curvas de desempeño. En específico, la Figura 6.2a corresponde al puntaje obtenido por los métodos tradicionales al variar la cantidad de etiquetas sin utilizar tratamiento para el desbalance, mientras que la Figura 6.5b corresponde al mismo caso pero utilizando pesos en el entrenamiento para balancear los casos positivos y negativos.

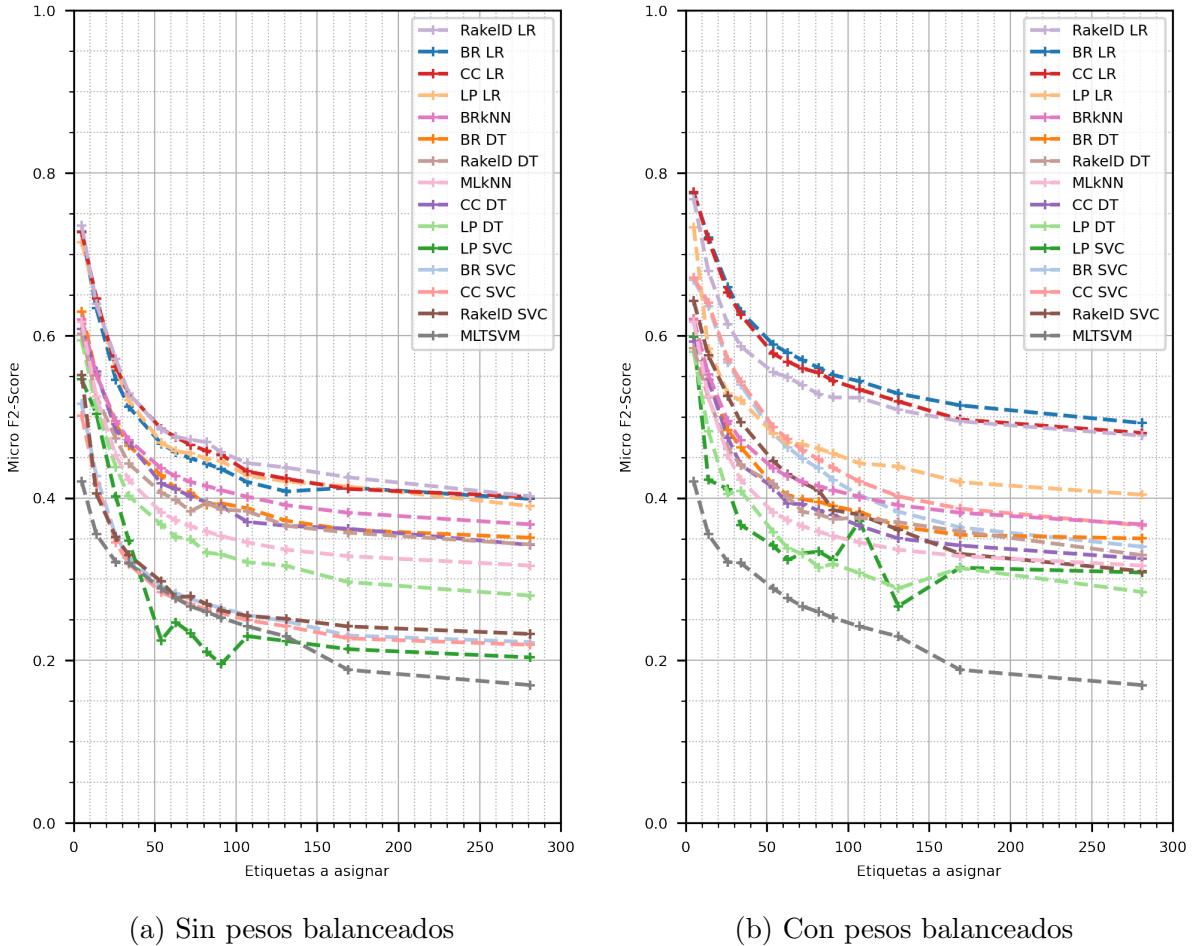


Figura 6.2: *Micro F<sub>2</sub>-Score* métodos sobre base, variando tratamiento del desbalance

Es preciso destacar que, si bien se realizaron experimentos utilizando 4 clasificadores base (LR, SVC, DT, GNB) para cada método de transformación del problema, las Figuras solo incorporan los resultados obtenidos con 3 de ellos, dejando fuera GNB para evitar una sobrecarga visual que complicase el análisis. Dichos resultados guardan un comportamiento similar al de MLkNN y su detalle permanece disponible en el Anexo A.

Diversos fenómenos destacan al estudiar las curvas anteriores, de los cuales el quizá más evidente es el decrecimiento prácticamente monótono de las curvas a medida que la cantidad de etiquetas en estudio aumenta. Tal fenómeno es razonable en cuanto aumentar la cantidad de etiquetas implica mayor dificultad y, sin embargo, existen también métodos que en ciertos pasos rompen con ello, siendo LP SVC el caso más drástico. Al respecto:

- Si bien el aumentar las etiquetas a asignar agrega dificultad sobre los métodos, también les entrega más información sobre las etiquetas.
- En específico, un paso pequeño (por ejemplo, de 82 a 91 etiquetas) podría incluir etiquetas con relación de coocurrencia con las previamente presentes, haciendo que ambas resulten más fáciles de predecir al tener mayor información y por ende aumentando las métricas de los métodos capaces de explotar tales relaciones.

- Tal hipótesis tendría sentido además al considerar los métodos que presentan dicho fenómeno, en cuanto son principalmente aquellos con LP como clasificador multietiqueta, método que efectivamente utiliza la información de relaciones entre etiquetas.

Ahora bien, profundizando el análisis y comenzando por el estudio del caso sin balance, se distinguen tres “grupos” de métodos: aquellos con un desempeño en general bajo (LP SVC, RakelD SVC, CC SVC, BR SVC y MLTSVM), otros con resultados de calidad intermedia (LP DT, RakelD DT, CC DT, BR DT, MLkNN y BRkNN) y finalmente los más competitivos, que disputan el primer lugar en distintas abscisas (RakelD LR, BR LR, CC LR y LP LR). Sobre ello:

- La pertenencia exclusiva de un mismo clasificador base a cada grupo pone en evidencia una relación de orden independiente al clasificador multietiqueta. En específico, los resultados indicarían que LR responde de mejor manera al problema binario de cada etiqueta que DT, y que a su vez DT arrojaría mejores resultados que SVC.
- Para los métodos de adaptación del algoritmo, por otro lado, se observa que el con mejor *Micro F<sub>2</sub>-Score* corresponde a BRkNN, el cual se posiciona como el mejor método del grupo con resultados de calidad intermedia, mas a una diferencia promedio de 0,0402 puntos de BR LR, siendo este último el método competitivo con promedio más bajo (véase Tablas 6.3 y 6.4).
- Cabe mencionar que dichas agrupaciones se han realizado por inspección del comportamiento de los métodos en todo el espectro, y con especial atención a los casos más complejos.
- Vale también aclarar que cuando se habla de métrica promedio se refiere al promedio de los puntajes obtenidos en cada una de las abscisas para las cuales hay información, y que dicha magnitud se utiliza solo como complemento a la inspección visual para el orden de los métodos. Para situaciones donde tal inspección no baste o no sea posible dada la proximidad entre curvas, se utilizará el *Area Under Curve* (AUC, véase Sección 6.3.4), métrica más robusta pero con un significado menos cercano que el promedio.

El siguiente fenómeno a comentar es el efecto de incorporar pesos, a saber, la comparación entre ambas visualizaciones. Al respecto:

- El cambio que más destaca es la evidente mejoría que los métodos RakelD LR, BR LR y CC LR presentan al incorporar pesos en el entrenamiento, posicionándose ampliamente por encima de los demás métodos en todo el espectro y en especial cuando la cantidad de etiquetas aumenta sobre 54.
- Destaca en particular BR LR, que pasase de ser uno de los cuatro métodos más competitivos a ser el mejor en casi todo el espectro (salvo con 5 etiquetas), alcanzando un Micro F<sub>2</sub> promedio de 0,5932 (0,1079 puntos por sobre el caso sin pesos, véase Tabla 6.3).
- Si bien los métodos con LR como clasificador base se mantienen como los mejores, los más impactados por la incorporación de pesos son aquellos con SVC, presentando incrementos desde 0,0718 puntos (caso LP SVC) hasta 0,1870 (caso CC SVC).

- Por otro lado, los métodos con DT como clasificador base se ven impactados de forma levemente negativa por la incorporación de pesos, siendo CC DT el método más afectado con una disminución de 0,0133 puntos.

Tabla 6.3: *Micro F<sub>2</sub>-Score* promedio para métodos de transformación del problema.

**Observación:** la cuarta columna corresponde a la diferencia entre el puntaje obtenido con y sin pesos, en ambos casos utilizando ResNet18 para generar descriptores.

Método	Sin pesos	Con pesos	Diferencia
RakelD LR	0,5028	0,5649	0,0621
BR LR	0,4853	0,5932	0,1079
CC LR	0,4975	0,5849	0,0874
LP LR	0,4896	0,4922	0,0026
RakelD DT	0,4199	0,4153	-0,0046
BR DT	0,4337	0,4280	-0,0057
CC DT	0,4284	0,4151	-0,0133
LP DT	0,3762	0,3639	-0,0123
RakelD SVC	0,3079	0,4388	0,1309
BR SVC	0,3048	0,4730	0,1682
CC SVC	0,2978	0,4848	0,1870
LP SVC	0,2908	0,3626	0,0718

Tabla 6.4: *Micro F<sub>2</sub>-Score* promedio para métodos de adaptación del problema.

**Observación:** A diferencia de la Tabla 6.3, esta vez no se presentan columnas según uso de pesos pues los métodos de adaptación no admiten tal tratamiento. Nuevamente, los resultados se obtienen a partir de descriptores generados mediante ResNet18.

Método	Promedio
BRkNN	0,4451
MLkNN	0,3980
MLTSVM	0,2762

Tales cambios alterarían el orden descubierto en el estudio del caso sin pesos para ahora reunir a los métodos con DT y SVC como clasificador base en un solo gran grupo de resultados intermedios, acotado superiormente por LP LR (que se separaría de los métodos competitivos) e inferiormente por MLTSVM, que se quedaría ampliamente atrás. Para el grupo competitivo, por otro lado, se observaría ahora un orden en el cual BR LR sería el mejor método, seguido de CC LR y finalmente RakelD, separados, sin embargo, por diferencias estrechas (por ejemplo, la diferencia promedio entre BR LR y CC LR es de 0,0083).

De esta manera, el experimento habría cumplido a cabalidad con sus objetivos, a saber:

- Obtener una primera noción de qué técnicas se adaptan mejor al problema, y en qué rango de etiquetas lo hacen. En este sentido, se habrían reconocido 3 técnicas que sobresalen ampliamente del resto y que mantienen su dominancia en todo el espectro de etiquetas, manteniendo un  $F_2$ -Score sobre 0,45 incluso en el caso más complejo.
- Determinar el efecto de aplicar técnicas para el tratamiento del desbalance. En este caso, los experimentos demostrarían que el impacto de incorporar pesos para balancear el entrenamiento de los clasificadores base contribuiría ampliamente a obtener mejores resultados, salvo por aquellos métodos que dependan de SVC como clasificador base (los cuales, sin embargo, no forman parte de los métodos competitivos).

En consecuencia, los próximos experimentos del Enfoque Tradicional se realizarán fijando el uso de tratamiento al desbalance, en cuanto el presente experimento habría demostrado sus beneficios. Para los métodos en estudio, por otro lado, se mantendrá la experimentación con todos ellos a fin de evaluar si el orden observado hasta ahora se mantiene al cambiar el conjunto de descriptores, sin embargo, el conocimiento adquirido hasta el momento será considerado para la elección de métodos a realizar en el Experimento 3 (Sección 6.3.4).

Antes de continuar, sin embargo, vale la pena revisar los componentes del  $F_2$ -Score en búsqueda de mayor comprensión de los resultados actuales. Para ello se presentan nuevas curvas de desempeño según *Precision* (Figura 6.3) y *Recall* (6.4).

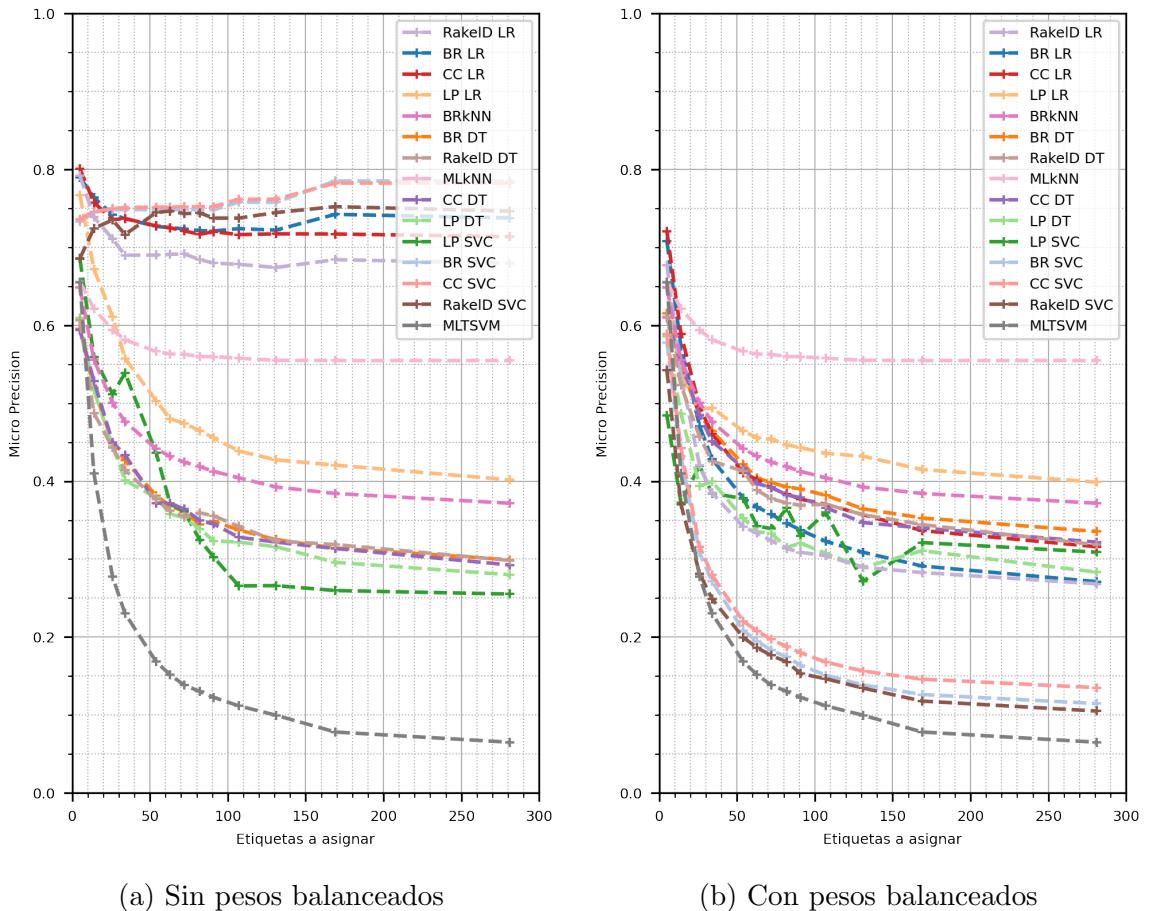


Figura 6.3: *Micro Precision* métodos sobre base, variando tratamiento al desbalance

En el estudio de la *Precision* destaca el hecho de que si bien parte de los métodos mantiene su comportamiento al incorporar el tratamiento de pesos, otros como CC LR, BR LR y RakelD SVC empeoran drásticamente con dicho cambio. Tal fenómeno es, sin embargo, razonable, en cuanto el fuerte desbalance contribuye a que la decisión de asignar muy pocas etiquetas permita alcanzar buenos valores de *Precision*, desbalance que el uso de pesos “oculta” a los clasificadores para obligarlos a “equivocarse más” en pos de ser más relajados con la asignación de etiquetas. Dicho *trade-off* se ve también reflejado en las curvas de *Recall*, donde los mismos métodos que caen en *Precision* elevan sus puntajes hasta más de 3 veces el valor obtenido al no utilizar pesos.

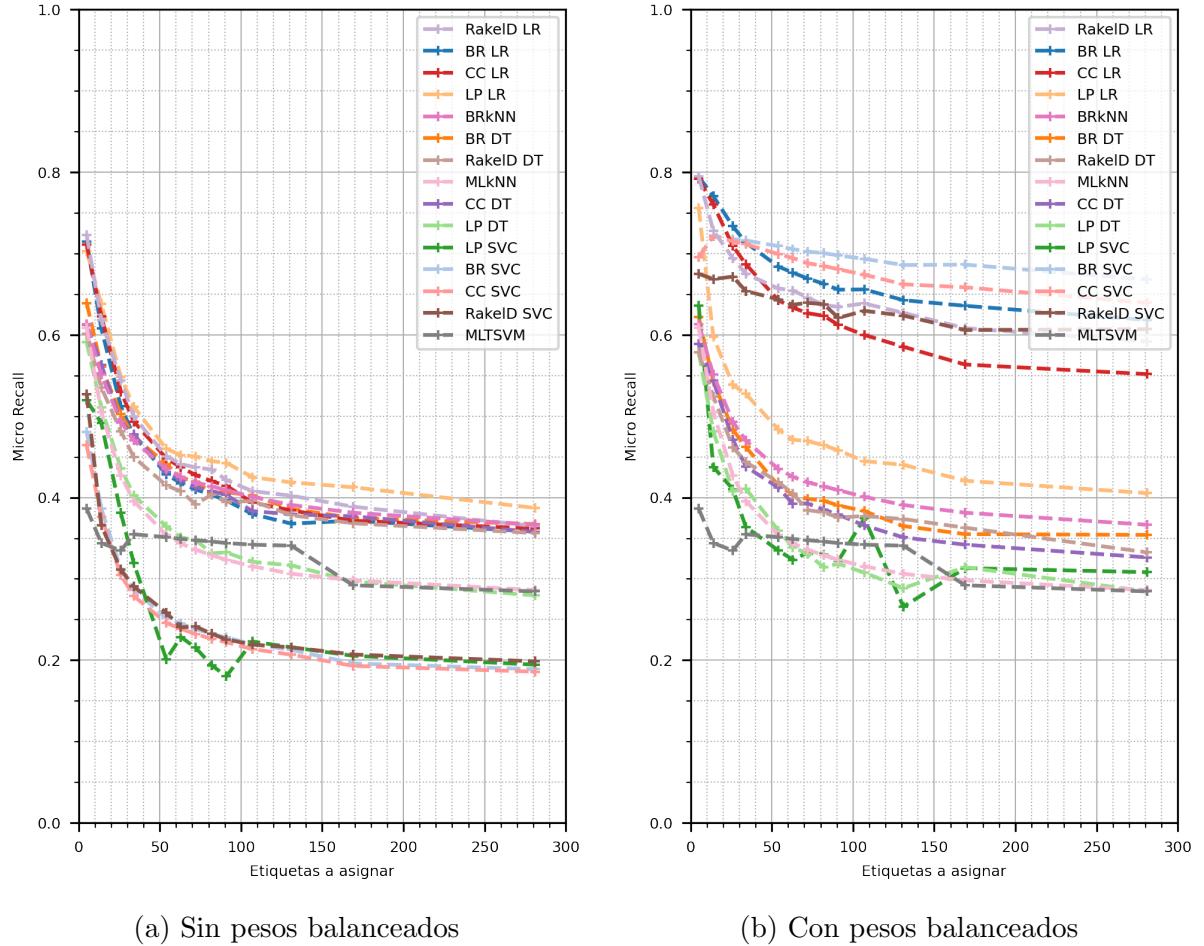


Figura 6.4: *Micro Recall* métodos sobre base, variando tratamiento al desbalance

Un último punto a comentar se encuentra en el hecho de que si bien aún resta por analizar las métricas basadas en ejemplos (*Hamming Score* y *K-Match Ratio*), dicho análisis se postergará hasta el Análisis de Resultados (Capítulo 8), donde se estudiarán en profundidad los mejores resultados de cada enfoque.

### 6.3.3. E2: Evolución métodos sobre base, variando descriptores

El siguiente experimento corresponde a la comparación de todos los métodos sobre el conjunto de datos original y en todo el espectro de etiquetas, fijando el tratamiento del desbalance y variando el origen de los descriptores entre ResNet18 y ResNet50. Al respecto:

- Si bien el desarrollo de los experimentos con descriptores a partir de ResNet18 tomaría en promedio 3 horas por opción de balanceo, al utilizar descriptores de ResNet50 el tiempo necesario habría aumentado alrededor de 15 horas.
- Lo anterior cobra sentido al considerar que los descriptores extraídos de ResNet50 son de dimensión 2048, 4 veces aquellos de ResNet18 (de dimensión 512).

A continuación se presenta el *Micro F<sub>2</sub>-Score* obtenido por cada método, una vez más a través de curvas de desempeño. La Figura 6.5a, por un lado, corresponde al caso con descriptores de ResNet18 (que coincide con lo presentado en 6.2b más se incluye nuevamente a fin de facilitar la comparación). La Figura 6.5b, por otro lado, presenta los resultados al utilizar descriptores generados mediante ResNet50.

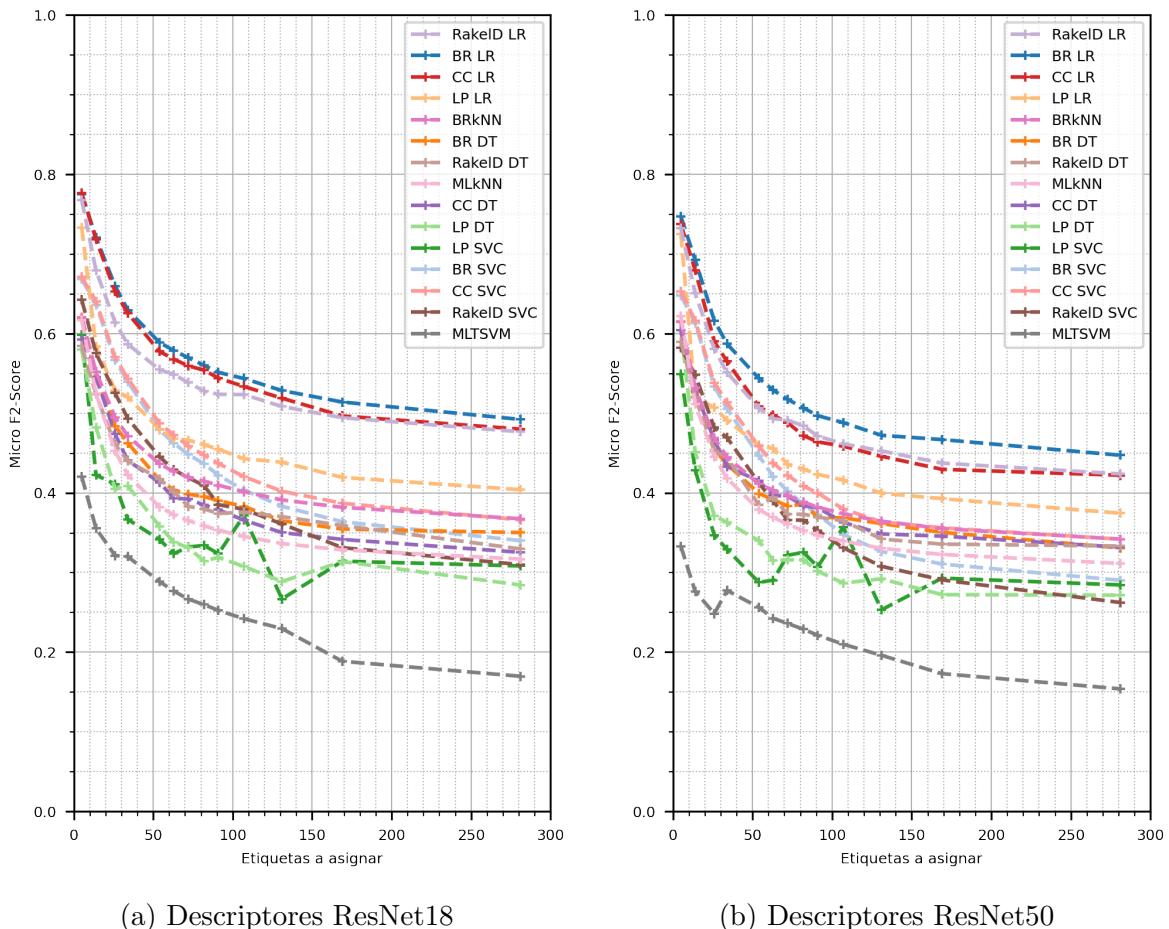


Figura 6.5: *Micro F<sub>2</sub>-Score* métodos sobre base, variando conjunto de descriptores

Nuevamente, el detalle de los resultados se encuentra disponible en los Anexos. En específico, la Tabla B.1 incorpora las diferentes métricas obtenidas por cada método al utilizar

descriptores obtenidos a partir de ResNet50, mientras que el caso con descriptores extraídos desde de ResNet18 se presenta en la Tabla A.2.

Al observar las visualizaciones anteriores destaca en primer lugar su similitud, en cuanto el comportamiento general de las curvas se mantiene y lo mismo sucede a grandes rasgos con el orden entre ellas. Sin embargo, al comparar las dos versiones de una misma curva se distingue que, en general, aquellas de la Figura 6.5b son levemente más bajas que las de 6.5a.

Aún más, al estudiar el promedio de *Micro F<sub>2</sub>-Score* obtenido por los tres mejores métodos sobre las distintas cantidades de etiquetas en estudio (Tabla 6.5), se observa que para dichos métodos la diferencia va desde los cuatro puntos porcentuales (0,0462 para BR LR y 0,0468 para RakelD LR) hasta casi siete (0,0651 para CC LR), siempre a favor de los experimentos utilizando descriptores obtenidos a partir de ResNet18.

Tabla 6.5: *Micro F<sub>2</sub>-Score* promedio para 3 mejores métodos, según conjunto de descriptores. **Observación:** la cuarta columna corresponde a la diferencia entre el puntaje obtenido con descriptores a partir de ResNet18 y ResNet50.

Método	ResNet18	ResNet50	Diferencia
BR LR	0,5932	0,5470	0,0462
CC LR	0,5849	0,5198	0,0651
RakelD LR	0,5649	0,5181	0,0468

Ahora bien, considerando que el objetivo principal de este experimento es el de identificar si descriptores más grandes dan lugar a mejores resultados en el conjunto *base*, la respuesta inmediata sería que, por el contrario, los resultados habrían empeorado en tal escenario, y que distintos métodos serían impactados en diferentes cantidades (por ejemplo, CC LR habría decaído bastante más que RakelD LR, quedando por debajo en gran parte del espectro).

Esto haría justicia con lo presentado en la Tabla 6.1 en cuanto ResNet50 obtendría un *accuracy* sobre *base* de 0,5928 contra un 0,7629 de ResNet18, lo cual sugeriría que esta última arquitectura captaría de mejor manera la semántica bajo la cual un patrón se asocia a un capítulo y, así mismo, sus descriptores estarían aportando con más información útil para el caso multietiqueta que aquellos de ResNet50.

Tal fenómeno podría llevar escoger en adelante el conjunto de descriptores de ResNet18 y trabajar solo con ellos. Dicha decisión sería razonable, más no estaría tomando en consideración toda la información que la ya comentada Tabla 6.1 aporta, a saber:

- Si bien en el conjunto *base* ResNet18 habría logrado mejor *test accuracy* en la tarea de descubrir el capítulo de un patrón, tal fenómeno no sería transversal a todos los conjuntos de datos.
- Aún más, la Tabla 6.1 pondría en evidencia que si bien en los conjuntos de datos más pequeños (aquellos correspondientes a una sola técnica de D.A.) ResNet18 llevaría ventaja, al aumentar la variabilidad en los patrones (conjuntos más grandes) ResNet50 tomaría la delantera.

- Esto tendría sentido, en cuanto ResNet50 podría corresponder a una arquitectura demasiado grande para los conjuntos más pequeños (dando lugar a un sobre ajuste) pero que se adapta mejor al caso con múltiples datos sintéticos.

Ahora bien, tal información sería particularmente útil si la causalidad entre un valor alto de *accuracy* en las redes para su tarea multiclas y la calidad de los descriptores para la tarea multietiqueta fuese clara, más hasta el momento solo se contaría con correlación en el caso base. Sin embargo, descartar una posible causalidad podría alejarnos de obtener mejores resultados.

Lo anterior llevaría a optar por no decantar los experimentos restantes a un solo conjunto de descriptores, sino más bien a continuar trabajando con ambos pero reduciendo el espectro de métodos a los tres que mejores resultados han mostrado hasta el momento, a saber, BR LR, CC LR y RakelD LR. En este sentido, el experimento habría servido para corroborar que independientemente al conjunto de descriptores a utilizar, la dominancia de tales métodos se mantendría y con amplia ventaja.

#### **6.3.4. E3: Evolución mejores métodos, variando D.A. y descriptores**

El último experimento a conducir para el Enfoque Tradicional corresponde a la comparación de los tres mejores métodos nuevamente sobre el espectro completo de etiquetas, pero esta vez variando el conjunto de datos a utilizar (en relación con las técnicas de D.A.) y la red utilizada para generar los descriptores. Al respecto:

- Si bien hasta el momento los experimentos han indicado que ResNet18 entrega mejores resultados, el variar la cantidad de datos disponibles para el entrenamiento de las redes generadoras (a través de las técnicas de D.A.) podría cambiar esta situación, bajo las razones expuestas como parte del análisis del Experimento 2 (Sección 6.3.3).
- Cabe notar que al incluir técnicas de D.A. el tiempo necesario para conducir los experimentos aumenta sustancialmente. Por ejemplo, mientras evaluar los 3 mejores métodos con descriptores obtenidos a partir de ResNet18 para el conjunto 0 (base) toma alrededor de 10 minutos, el mismo ejercicio con el conjunto 15 (todas las técnicas de D.A. simultáneamente) requiere poco más de 2 horas. Al replicar dicho ejercicio con ResNet50, por otro lado, el tiempo aumenta sobre las 8 horas.
- Dado lo anterior, la obtención de todos los resultados relativos a este experimento habría requerido un total aproximado de 38 horas, sin contar el tiempo asociado al reentrenamiento de las redes generadoras (que requiriese más de 200 horas de un procesador gráfico Nvidia RTX 3060).

A diferencia de los experimentos anteriores, en este caso la cantidad de curvas es excesiva para ser presentada en una única visualización (un total de 96 curvas correspondientes a 3 métodos sobre 16 conjuntos y para 2 orígenes de descriptores distintos). Por tal razón, se opta por estudiar las 15 con mejor AUC (*Area Under Curve*), manteniendo, sin embargo, el detalle en el Apéndice C. Cabe mencionar que en esta ocasión se escogió el AUC para

reemplazar al promedio como medida de orden entre curvas, por tratarse de un experimento donde la dominancia es más compleja de identificar, y por ende se requiere de una métrica más robusta.

La Figura 6.6 presenta los resultados obtenidos para las 15 mejores combinaciones de descriptor-conjunto-método según el AUC de las curvas  $F_2$ -Score. La Tabla 6.6, por otro lado, presenta dichos valores de AUC en orden descendiente.

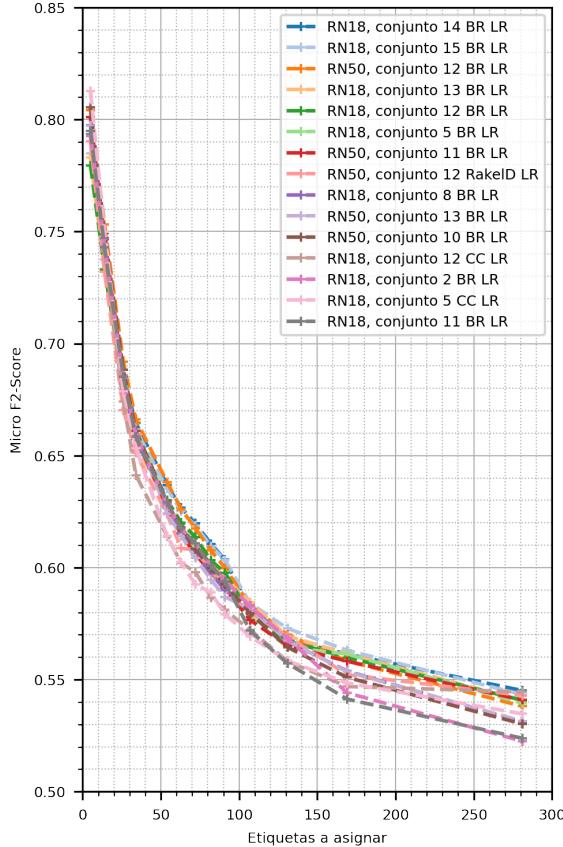


Figura 6.6: *Evolución Micro  $F_2$ -Score* para 3 mejores métodos en el espectro de etiquetas, variando conjunto de descriptores y técnica de D.A. **Observación:** el eje Y ha sido recortado al rango [0,50, 0,85] debido a la proximidad entre curvas.

A diferencia de los experimentos anteriores, esta vez las curvas son sumamente próximas (incluso tras un *zoom* en el eje Y) y muchas de ellas se cruzan en reiteradas ocasiones, lo cual complica el análisis visual y resta significado al estudio de los promedios. Al estudiar la Tabla 6.6, por otro lado, se desprenden distintas observaciones:

- De las 15 combinaciones con mejor AUC para  $F_2$ -Score, 10 de ellas utilizan ResNet18, incluida la mejor de todas.
- Un fenómeno similar ocurre con el conjunto de datos, donde 11 de 15 combinaciones utilizan conjuntos donde la cantidad de imágenes para entrenamiento por *fold* supera las 3000 (del conjunto 11 en adelante). En particular, la mejor combinación utilizaría el conjunto 14, correspondiente al uso simultáneo de todas las técnicas salvo *randAugment*.

- Para el método utilizado, por otro lado, BR LR se posiciona como el mejor no solo por estar presente en los tres primeros lugares, sino por corresponder al método utilizado por 11 de las 15 mejores combinaciones.

Tabla 6.6: AUC de *Micro F<sub>2</sub>-Score* para 3 mejores métodos, variando conjunto de descriptores y técnica de D.A.

Origen descriptores	Conjunto de datos	Método	AUC
ResNet18	14	BR LR	244,0482
ResNet18	15	BR LR	243,9790
ResNet50	12	BR LR	243,3054
ResNet18	13	BR LR	242,4738
ResNet18	8	BR LR	242,3744
ResNet18	11	BR LR	242,2549
ResNet50	12	RakelD LR	241,9622
ResNet50	8	BR LR	241,1392
ResNet18	13	BR LR	240,6531
ResNet50	10	BR LR	240,5909
ResNet50	12	CC LR	240,5013
ResNet18	2	BR LR	239,3588
ResNet18	5	CC LR	239,0618
ResNet18	11	BR LR	238,7767
ResNet18	11	RakelD LR	238,1474

Por otro lado, al comparar tales resultados con los mejores de experimentos anteriores (véase Tabla 6.7) se observa el efecto de utilizar datos sintéticos: mientras en *base* el mejor promedio y AUC habrían sido de 0,5932 y 225,2721 respectivamente, la mejor combinación encontrada en este experimento alcanza valores de 0,6340 y 244,0482, elevando el promedio en más de 4 puntos porcentuales.

Tabla 6.7: Comparación mejores métodos Enfoque Tradicional

Experimento	Método	Origen descriptores	Conjunto de datos	Promedio <i>Micro F<sub>2</sub></i>	AUC <i>Micro F<sub>2</sub></i>
1 & 2	BR LR	ResNet18	0	0,5932	225.2721
3	BR LR	ResNet18	14	0,6340	244,0482

De esta manera, los exhaustivos experimentos realizados para el Enfoque Tradicional habrían permitido finalmente identificar la mejor expresión de las variables en estudio para el problema, a saber:

1. El uso de *Binary Relevance* como clasificador multietiqueta, con Regresión Logística como clasificador base.
2. El uso de pesos en el clasificador base para balancear las clases positiva y negativa.
3. La aplicación de las técnicas de rotación, reflexión, lluvia, deformación elástica, difuminado (*avg. blur*) y recorte para generar datos sintéticos, todas al mismo tiempo.
4. El uso de ResNet18 como red generadora de descriptores.

Si bien tales resultados se analizarán en más detalle al momento de comparar las mejores técnicas de cada enfoque (Capítulo 8), hay algunos fenómenos que merece la pena comentar de inmediato. Entre ellos:

- Destaca en particular que el clasificador multietiqueta más sobresaliente haya sido *Binary Relevance*, en cuanto este trata cada etiqueta de forma independiente y por ende no explota las relaciones entre las mismas, una desventaja que en el Estudio de la Literatura (Capítulo 3) se habría planteado como crítica.
- Destaca también el hecho de que utilizar ResNet18 como red generadora de descriptores haya resultado más conveniente que utilizar su versión de mayor capacidad, ResNet50. Si bien este fenómeno parecía razonable en conjuntos pequeños (como se expuso en la Sección 6.3.3), el que siga ocurriendo en los conjuntos más grandes donde ResNet50 alcanza los mejores valores de *accuracy* es, al menos, llamativo. Se cree, sin embargo, que las razones podrían guardar relación con la posibilidad de que aún no haya suficiente variedad en los datos o bien que un nivel de aprendizaje alto en la tarea de asignar el capítulo correspondiente a un patrón - como el que alcanza ResNet50 - no resulte en mejores descriptores para la tarea multietiqueta, sino más bien que sea un conocimiento “general” en el caso multiclase el que mejor se transfiera a dicha tarea.
  - Esta última posibilidad contribuiría además a cerrar una de las interrogantes planteadas en el Experimento 2 (Sección 6.3.3), en cuanto no existiría causalidad entre un mejor *test accuracy* en la tarea multiclase y un valor de *F<sub>2</sub>-Score* más alto en la tarea multietiqueta.
- Para el uso de pesos, por otro lado, resulta razonable que aplicar tratamiento al desbalance sea la opción ganadora, en cuanto tal como se reconociese desde un inicio, el desbalance de casos negativos y positivos es uno de los desafíos del problema en estudio.
- Finalmente, el que sean los conjuntos con más técnicas de D.A. los que hayan entregado los mejores resultados resulta también razonable, en cuanto aumenta la información disponible para entrenamiento tanto de las redes generadoras como de los clasificadores.

Cabe mencionar que las métricas no presentadas en esta ocasión (detalle del *F<sub>2</sub>-Score* en términos de *Recall* y *Precision*) no presentan cambios mayores con respecto a las expuestas en el caso balanceado del Experimento 1 (Sección 6.3.2) y se explorarán en profundidad al momento de comparar con la mejor técnica del Enfoque XMLC (Capítulo 8, Análisis de Resultados).

# Capítulo 7

## Enfoque XMLC

Este capítulo presenta la exploración del problema desde un enfoque de XMLC, es decir, a través de métodos desarrollados para el caso con una cantidad amplia de etiquetas. En primer lugar, se presenta el detalle técnico de los métodos (Sección 7.1) y los descriptores a utilizar como entrada (7.2). Luego se presentan los experimentos realizados, sus resultados y un breve análisis de los mismos (7.3).

### 7.1. Técnicas

La Selección de métodos para el Enfoque XMLC (Sección 4.4.2) incluye tres formas distintas de abordar el problema: a través de C2AE como método de *Deep Embedding*, mediante AlexNet como *Threshold Dependent Neural Network* (TDNN) y por último a través de AlexNet en complemento a celdas recurrentes LSTM, como método de *Joint Embedding CNN-RNN*. El detalle técnico de dichas técnicas se aborda a continuación.

#### 7.1.1. C2AE

Antes de describir en detalle C2AE y su funcionamiento, conviene primero presentar dos arquitecturas sobre las cuales este se basa. Se trata de *Deep Canonical Correlation Analysis* (DCCA) [1] y la arquitectura de redes neuronales *autoencoder*.

#### DCCA

DCCA corresponde a una arquitectura de redes profundas utilizada para reunir información de dos perspectivas o vistas diferentes de un mismo fenómeno en una única representación la cual maximiza la correlación entre ambas fuentes, descartando las características no relacionadas al tratarlas como ruido y abriendo la posibilidad de reconstruir una vista en función de la otra.

Dado lo abstracto que el concepto de vista puede parecer, resulta conveniente introducir DCCA con un ejemplo. Para ello, supóngase el caso de un vídeo de una persona tocando batería, el cual se separa en un conjunto de imágenes y audios. Cada conjunto correspondería

a una vista diferente del mismo fenómeno, a saber, el de un baterista tocando su instrumento. Ahora bien, la aplicación de DCCA sobre dichas vistas permitiría reunir ambas informaciones en una única gran representación vectorial de correlaciones, de las cuales un ejemplo podría ser el hecho de que cada vez que una baqueta golpea la batería se emite un sonido de cierta frecuencia, y viceversa. Dicha representación podría ser utilizada para, posteriormente, obtener el sonido de otro video similar para el cual se cuente solamente con la imagen.

Más formalmente, DCCA recibe dos vistas  $x_1$  y  $x_2$ , y busca encontrar representaciones  $f_1(x_1)$  y  $f_2(x_2)$  tal que maximicen la Ecuación 7.1:

$$\text{corr}(f_1(x_1), f_2(x_2)) = \frac{\text{cov}(f_1(x_1), f_2(x_2))}{\sqrt{\text{var}(f_1(x_1)) \cdot \text{var}(f_2(x_2))}} \quad (7.1)$$

Para tal fin, DCCA se basa en *Canonical Correlation Analysis* (CCA), técnica con el mismo objetivo pero que busca representaciones lineales del tipo  $f_1(x_1) = w'x_1$  para  $w_1 \in R^{n_1}$ , lo cual sin embargo no tiene mucha utilidad por sí solo en problemas complejos pues las relaciones que logra rescatar son de muy bajo nivel.

En su lugar, DCCA propone anteponer CCA por el uso de funciones no lineales dictadas por redes profundas, las cuales permitan descubrir relaciones más abstractas y con mayor correlación. En específico, DCCA presenta el uso de dos perceptrones multicapa, uno por cada vista, los cuales son utilizados para encontrar representaciones profundas de las mismas, a fin de luego entregar dichas representaciones a CCA para reunirlas en un único espacio latente, como se ilustra en la Figura 7.1.

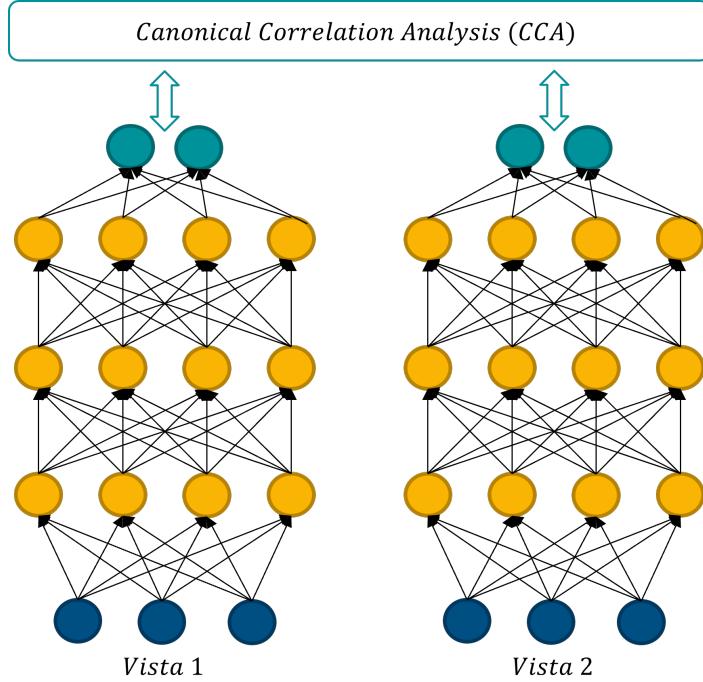


Figura 7.1: Arquitectura DCCA.

Un punto importante a mencionar sobre DCCA es la forma en que dicha arquitectura se entrena. Tal proceso se gesta en dos etapas, una primera donde el mapeo de cada vista se

entrena de manera individual (cada perceptrón en la Figura 7.1) y una segunda en la cual se ajustan todos los parámetros en conjunto para maximizar la correlación de las capas de salida (las entradas de CCA).

Desde luego, el uso de perceptrones multicapa es solo una de las posibles arquitecturas de DCCA, siendo la idea principal el uso de redes profundas con funciones de activación no lineales como antecede a CCA para dotar a este último de relaciones de más alto nivel.

### **Autoencoders**

Los *autoencoders* son un tipo de red neuronal frecuentemente utilizada para aprender representaciones (o codificaciones) eficientes de una cierta entrada a través de un entrenamiento enfocado en ignorar la información insignificante (o ruido) y resaltar aquellos aspectos relevantes de forma óptima. De forma general, un *autoencoder* se define por:

- Un espacio de entrada  $X$  y un espacio de codificación  $Z$  (este último también denominado espacio latente).
- Dos familias de funciones parametrizadas: aquellas encargadas de la codificación  $E_\phi$  tal que  $E_\phi : X \rightarrow Z$ , y las encargadas de la decodificación o reconstrucción,  $D_\theta : Z \rightarrow X$ .

Cada familia de funciones actúa como un componente con una tarea específica en el funcionamiento del *autoencoder*: mientras la familia de codificación tiene por misión el alcanzar una representación eficiente de la entrada en una dimensión por lo general más pequeña o “cuello de botella” (condición impuesta para limitar la capacidad de la red y así evitar una codificación ineficiente), la familia de decodificación tiene por objetivo el reconstruir la entrada original a partir de la codificación. La Figura 7.2 presenta un ejemplo de arquitectura *autoencoder*, en donde cada familia está compuesta por una única capa de neuronas *fully-connected*.

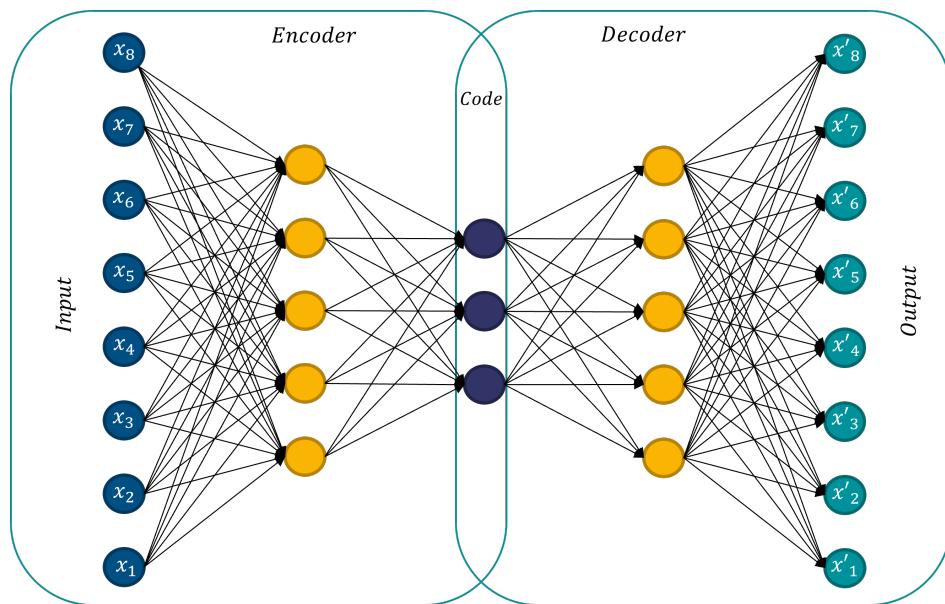


Figura 7.2: Ejemplo de arquitectura *autoencoder*.

Un aspecto clave en la definición de los *autoencoders* es la forma en que estos son entrenados para aprender una codificación apropiada. En este sentido, la función objetivo es aquella que, dada una cierta distribución de probabilidad  $\mu_{ref}$  sobre  $X$ , calcula la media de una función  $d : X \times X \rightarrow [0, \infty]$  tal que  $d(x, x')$  mide qué tanto difiere  $x'$  (la reconstrucción) de  $x$  (la entrada original), como se presenta en la Ecuación 7.2:

$$L(\theta, \phi) := \mathbb{E}_{x \sim \mu_{ref}} [d(x, D_\theta(E_\phi(x)))] \quad (7.2)$$

De esta manera el *autoencoder* óptimo viene dado por  $\text{argmin}_{\theta, \phi}(L(\theta, \phi))$ , es decir, por aquel con los parámetros  $\theta$  y  $\phi$  tal que la función de pérdida  $L$  alcanza su valor mínimo. Dicha configuración puede buscarse mediante diversas técnicas de optimización, tales como el descenso de gradiente.

Un escenario frecuente es aquel donde la distribución  $\mu$  sobre  $X$  corresponde a la distribución empírica del conjunto de datos y la función  $d$  a la función de pérdida L2:  $d(x, x') = \|x - x'\|_2^2$ . En tal caso, el problema se convierte en una optimización de mínimos cuadrados (Ecuación 7.3):

$$\min_{\theta, \phi}(L(\theta, \phi)) = \frac{1}{N} \sum_{i=1}^N \|x_i - D_\theta(E_\phi(x_i))\|_2^2 \quad (7.3)$$

Que es también el caso de la aplicación de *autoencoders* en C2AE, como se verá a continuación.

## C2AE

Una vez introducidas las arquitecturas de DCCA y *autoencoder*, la explicación de C2AE es directa. Acrónimo para *Canonical Correlated Autoencoder*, C2AE corresponde a un método de *Deep Embedding*, es decir, aquella propuesta de la familia del aprendizaje multietiqueta a través del *Deep Learning* en donde la idea principal es aprender una representación de los datos la cual sea capaz de codificar, en un espacio de dimensión limitada, todas las características importantes de la entrada relativas a su etiquetado.

Aún más, C2AE incorpora una innovación en este aspecto, y es que en lugar de buscar una representación solamente de las *features*, apunta a una representación conjunta (*joint embedding*) de *features* y etiquetas en un mismo espacio latente, lo cual le permite relacionar de mejor manera ambos dominios y así alcanzar mejores clasificaciones.

En particular, C2AE cuenta con las propiedades de:

- *Feature-aware label embedding*, es decir, de alcanzar una descripción de las etiquetas fuertemente relacionada con las *features* asociadas, lo cual logra a través del aprendizaje conjunto de un perceptrón multicapa para las *features* y el componente codificador de un *autoencoder* para las etiquetas, los cuales se reúnen y ajustan a través de DCCA (siendo entonces una aplicación de DCCA con un perceptrón multicapa para una vista y la primera etapa de un *autoencoder* para la otra).
- *Label-correlation aware prediction*, es decir, de explotar las correlaciones entre etiquetas al momento de predecir para una entrada desconocida. Tal propiedad (ampliamente discutida para los métodos tradicionales) es en este caso alcanzada mediante la introducción de una función de pérdida especialmente diseñada para la salida de C2AE, la cual se detallará más adelante.

Es así como, a través de DCCA, C2AE obtiene proyecciones de las entradas en un espacio latente conjunto, lo cual es a su vez ajustado con la retroalimentación recibida desde el componente de decodificación del *autoencoder* (encargado de recuperar las etiquetas originales a partir de la proyección). La Figura 7.3 presenta dicha arquitectura.

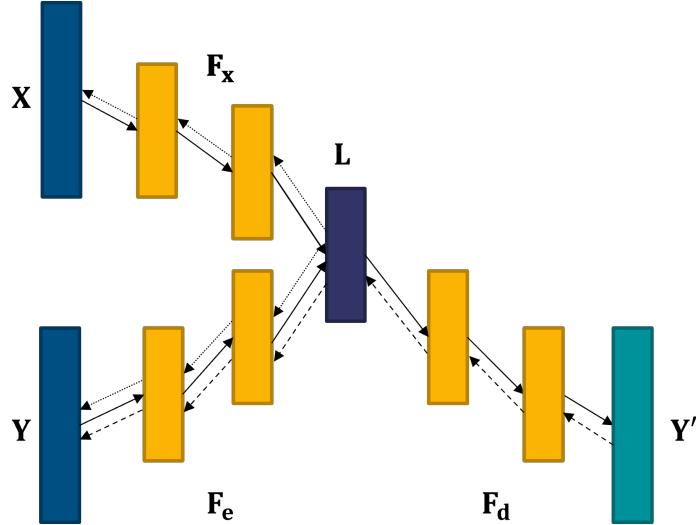


Figura 7.3: Arquitectura C2AE. **Observación:** las flechas a puntos y a líneas discontinuas representan la señal de error en el espacio latente ( $\Phi(\mathbf{F}_x, \mathbf{F}_e)$ ) y en la salida de C2AE ( $\Gamma(\mathbf{F}_e, \mathbf{F}_d)$ ), respectivamente.

Nuevamente, un punto crítico para el entendimiento de C2AE es cómo dicha arquitectura es entrenada y, en particular, como se propaga la señal de error entre los distintos componentes. Dicho aspecto se aborda a continuación.

Sean  $\mathbf{F}_x$ ,  $\mathbf{F}_e$  y  $\mathbf{F}_d$  las funciones de mapeo de *features*, de codificación y decodificación, respectivamente (tal cual se muestran en la Figura 7.3). Durante la fase de entrenamiento, la entrada de C2AE son las instancias  $\mathbf{X}$  y sus etiquetas  $\mathbf{Y}$ , mientras que la salida son las etiquetas de interés  $\mathbf{Y}'$  (las mismas de entrada). El componente DCCA de C2AE asocia  $\mathbf{X}$  e  $\mathbf{Y}$  para determinar el espacio latente  $\mathbf{L}$ , mientras que el componente *autoencoder* impone que la salida a partir de  $\mathbf{L}$  corresponda a  $\mathbf{Y}'$ , dando lugar a la siguiente función objetivo (Ecuación 7.4):

$$\Theta = \min_{\mathbf{F}_x, \mathbf{F}_e, \mathbf{F}_d} (\Phi(\mathbf{F}_x, \mathbf{F}_e) + \alpha \Gamma(\mathbf{F}_e, \mathbf{F}_d)) \quad (7.4)$$

Donde  $\Phi(\mathbf{F}_x, \mathbf{F}_e)$  y  $\Gamma(\mathbf{F}_e, \mathbf{F}_d)$  corresponden a las funciones de pérdida en el espacio latente y en la salida de C2AE, respectivamente. El parámetro  $\alpha$ , por otro lado, se incorpora para dar más peso a una u otra pérdida según resulte necesario según las especificaciones del problema.

Más en específico, la función  $\Phi(\mathbf{F}_x, \mathbf{F}_e)$  viene definida por la Ecuación 7.5:

$$\min_{\mathbf{F}_x, \mathbf{F}_e} (\|\mathbf{F}_x(\mathbf{X}) - \mathbf{F}_e(\mathbf{Y})\|_F^2) \quad (7.5)$$

Sujeto a las restricciones de codificación (Ecuación 7.6):

$$\mathbf{F}_x(\mathbf{X})\mathbf{F}_x(\mathbf{X})^T = \mathbf{F}_e(\mathbf{Y})\mathbf{F}_e(\mathbf{Y})^T = \mathbf{I} \quad (7.6)$$

Donde  $\mathbf{I}$  denota la función identidad. De esta manera,  $\Phi(\mathbf{F}_x, \mathbf{F}_e)$  calcula la pérdida entre el espacio latente generado por  $\mathbf{F}_x$  y  $\mathbf{F}_e$ .

La función  $\Gamma(\mathbf{F}_e, \mathbf{F}_d)$ , por otro lado, se detalla a continuación (Ecuación 7.7):

$$\Gamma(\mathbf{F}_e, \mathbf{F}_d) = \sum_{i=1}^N E_i \quad (7.7)$$

Sujeto a:

$$E_i = \frac{1}{|y_i^1||y_i^0|} \sum_{(p,q) \in y_i^1 \times y_i^0} \exp(\mathbf{F}_d(\mathbf{F}_e(\mathbf{y}_i))^q - \mathbf{F}_d(\mathbf{F}_e(\mathbf{y}_i))^p) \quad (7.8)$$

Donde  $\mathbf{y}_i^1$  denota el conjunto de etiquetas positivas en  $\mathbf{y}_i$  para la  $i$ -ésima instancia  $\mathbf{x}_i$ , e  $\mathbf{y}_i^0$  al conjunto de las etiquetas negativas. De esta forma, dada la entrada  $\mathbf{x}_i$ ,  $\mathbf{F}_d(\mathbf{F}_e(\mathbf{x}_i))^p$  retorna la  $p$ -ésima entrada de la salida de C2AE, lo cual se traduce en que minimizar la función de pérdida  $\Gamma$  corresponda a maximizar la predicción de las etiquetas reales y la no predicción de las etiquetas incorrectas para cada entrada, lo cual implícitamente fuerza la preservación de la información de co-ocurrencia entre etiquetas [63].

Especificado el proceso de entrenamiento de C2AE y las funciones de pérdida involucradas, la explicación del funcionamiento en predicción resulta mucho más simple. Sea  $\hat{x}$  una instancia de prueba, la cual C2AE recibe y transforma al espacio latente mediante  $\mathbf{F}_x$ . Dicha transformación es posteriormente pasada por  $\mathbf{F}_d$  para su decodificación en un conjunto de etiquetas  $\hat{y} \in L$  (es decir,  $\hat{y} = \mathbf{F}_d(\mathbf{F}_x(\hat{x}))$ ), las cuales se entregan como predicción de C2AE.

## Implementación, configuración e hiperparámetros

Una constante en los experimentos con el Enfoque XMLC será la utilización de *Pytorch* como *framework* para la implementación de los modelos, lo cual abre la posibilidad de realizar cálculos sobre GPU. Cabe mencionar que, si bien los experimentos del Enfoque Tradicional fueron llevados a cabo utilizando *scikit-learn* y por ende ocupando solo CPU para su procesamiento, en el Enfoque XMLC los métodos son mucho más costosos computacionalmente y por ende el uso de librerías habilitadas para el trabajo sobre GPU se vuelve necesario.

En particular, la implementación de C2AE sobre *Pytorch* se lleva a cabo de forma principalmente manual, mas utilizando como guía la implementación original en *Matlab* [62] y la versión en *TensorFlow 1.0* desarrollada por el usuario *znhub* en GitHub [75].

Ahora bien, en lo que respecta a la configuración de los distintos componentes a utilizar, se optará por lo siguiente:

- Para el perceptrón multicapa ( $\mathbf{F}_x$ ) se utilizarán 2 capas de 120 neuronas.
- Para los bloques codificador ( $\mathbf{F}_e$ ) y decodificador ( $\mathbf{F}_d$ ) del *autoencoder*, se utilizarán 2 capas *fully-connected* con una cantidad de neuronas igual a la mitad del número de etiquetas en estudio.
- Las funciones de activación para cada capa corresponderán a *Leaky-ReLU*, siguiendo la configuración experimental utilizada en la publicación original de C2AE.
- Por último, la dimensión del espacio latente se fijará a 70.

Dichos valores fueron encontrados de forma experimental usando el *fold 1* del conjunto de datos base y se ilustran en la Figura 7.4:

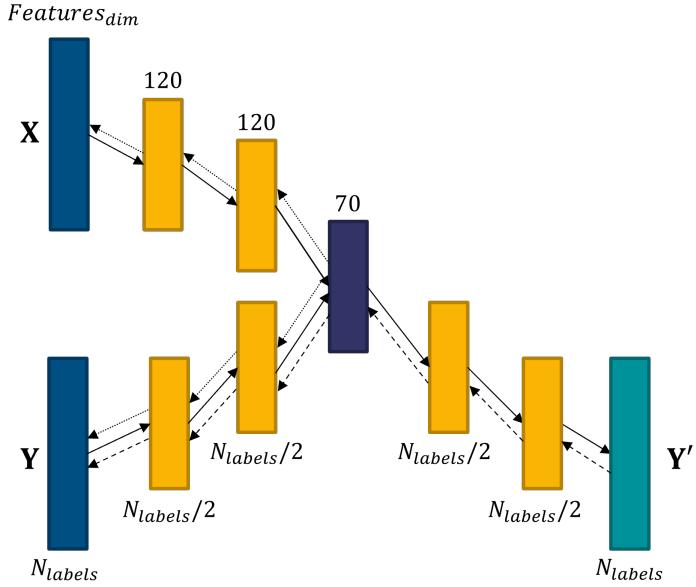


Figura 7.4: Configuración C2AE a utilizar. **Observación:** salvo por la entrada de la dimensión latente (en morado), las flechas hacia la derecha representan conexiones *fully-connected*.

Por último, el entrenamiento se llevará a cabo con los siguientes hiperparámetros:

- Función de pérdida: las especificadas por el modelo, descritas por la Ecuación 7.4.
- Número de épocas: 600
- *Batch size*: 128
- Optimizador: Adam
- *Learning-rate* inicial: 0,005
- Uso de *Early-Stopping* con paciencia de 100 épocas.
- *Scheduler: ReduceLROnPlateau*, técnica disponible en *Pytorch* la cual reduce el *learning-rate* en un cierto factor cuando una métrica objetivo deja de mejorar durante una cantidad dada de épocas (paciencia). En específico, se utilizará un factor de reducción de 0,75 con un *learning rate* mínimo de 0,0001, una paciencia de 40 épocas y con el micro F<sub>2</sub>-Score como métrica objetivo.

### 7.1.2. AlexNet TDNN

Antes de introducir la técnica de AlexNet TDNN, conviene primero revisar la familia de redes a la cual AlexNet pertenece: las redes neuronales convolucionales (CNN).

## CNN

Las CNN son un tipo de red neuronal las cuales amplían el funcionamiento clásico de recibir un vector de *features* para permitir que la entrada se presente como una matriz (que por lo general corresponde a una imagen). Este pequeño cambio permite que la red reciba no solamente descripciones locales, sino también información espacial (por ejemplo, qué localidad está junto a otra).

Para ello, las redes convolucionales incorporan un proceso de aprendizaje levemente distinto al de las redes neuronales clásicas, en cuanto si bien mantienen la existencia de pesos y el uso de una función de activación no lineal, dichos pesos ya no se asocian a una neurona como tal sino más bien a “filtros”: matrices pequeñas las cuales se utilizan para someter la entrada a un proceso de convolución, dando lugar a las denominadas “capas convolucionales”.

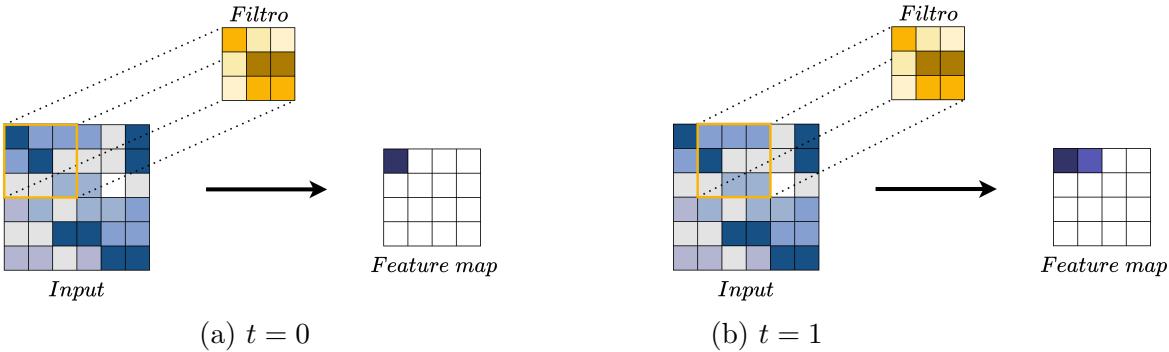


Figura 7.5: Ejemplo de convolución

A modo de apoyo, considérese la Figura 7.5a. En ella, el primer cuadrante de  $3 \times 3$  de la entrada (que podría corresponder, por ejemplo, al promedio de los píxeles de una imagen) es sometida a convolución con un filtro del mismo tamaño, lo cual da como resultado un valor que pasa a ser la primera entrada de la salida (denominada *feature map*). En la Figura 7.5b, por otro lado, el cuadrante en la matriz de entrada se desplaza una posición a la derecha para dar origen a la segunda entrada de la salida. El proceso se repite a lo largo y ancho de la entrada hasta completar el *feature map*, manteniendo el mismo filtro durante toda la operación.

Dicho procedimiento corresponde a la idea principal detrás de las redes convolucionales. Nótese que:

- Si bien en el ejemplo se utilizan entrada, filtro y salida en 2 dimensiones, estos podrían ser de cualquier dimensión.
- El procedimiento descrito corresponde a la aplicación de un único filtro (símil de operar una única neurona en las redes *fully-connected*). Una capa convolucional en general corresponde a  $N > 1$  filtros, lo cual da lugar a  $N$  *feature maps*.

Ahora bien, existen dos operaciones adicionales que suelen formar parte de las redes convolucionales. La primera de ellas corresponde al *pooling*, operación a través de la cual una matriz (que podría ser la entrada o la salida de alguna capa) es subsampleada a una matriz de menor

dimensión, a través de la convolución con una función de agregación como el máximo o el promedio. La Figura 7.6 presenta un ejemplo de *max pooling* para  $t = 0$ .

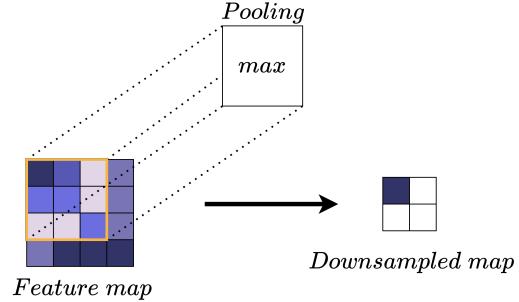


Figura 7.6: Ejemplo de *Max Pooling* en  $t = 0$

Por otro lado, la tercera operación clásica de las redes convolucionales corresponde al aplanamiento de un *feature map* para convertirlo en un vector 1D y poder alimentarlo a una red *fully-connected* clásica. La Figura 7.7 ilustra tal proceso.

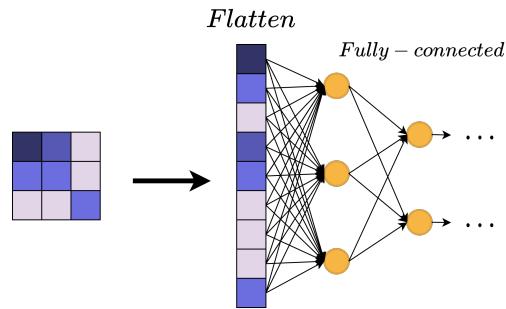


Figura 7.7: Ejemplo de conexión entre capas convolucionales y *fully-connected*.

Son estas tres operaciones las que, al ser operadas de forma conveniente, dotan a las redes convolucionales de la capacidad de aprender descripciones de alto nivel. Al respecto, una configuración frecuente es la de combinar capas de convolución y *pooling* en las primeras etapas para la generación de descriptores (por ejemplo, una única capa de convolución acostumbra captar bordes, dos capas captan contornos y tres capas ya comienzan a codificar objetos más complejos) y posteriormente aplicar una o más capas *fully-connected* para forzar la reducción de dimensionalidad y dar lugar a la clasificación.

Una característica importante de las CNN guarda relación con su capacidad de crecer en profundidad al agregar más capas y las consecuencias de dicho crecimiento. En este sentido, la intuición podría llevar a pensar que si el hecho de ser redes “más profundas” que los algoritmos tradicionales es lo que las dota de buenas propiedades, entonces el crecer más debería continuar entregando más valor a la red. Sin embargo, en la práctica tal supuesto no se cumple (o al menos no de inmediato) debido a una dificultad que surge en el entrenamiento, conocida como “desvanecimiento del gradiente”.

Dicha debilidad tiene relación con la forma en que las redes (tanto *fully-connected* como convolucionales) son entrenadas, donde una señal de error se obtiene en la salida y se propaga

hacia las capas más tempranas a través de un proceso de multiplicaciones sucesivas conocido como *back propagation*, lo cual permite ajustar los parámetros de cada capa. El problema de dicho proceder con el crecimiento de las redes se encuentra en que, dado que los parámetros a multiplicar viven en el rango  $[0, 1]^1$ , la multiplicación sucesiva entre muchas capas da lugar a valores cada vez más pequeños, resultando en que las primeras capas no perciban ajuste alguno y por ende en que la red comience a perder su capacidad de aprendizaje.

Tal fenómeno ha dado lugar a diversos intentos de solución en los últimos años. Entre ellos se cuentan las ya presentadas arquitecturas ResNet (Sección 6.2) las cuales buscan evadir el desvanecimiento mediante conexiones “residuales” que saltan bloques y por ende permiten que la señal de error llegue más atrás, las arquitecturas DenseNet en donde todas las capas se comunican con todas, el uso de funciones de activación especialmente diseñadas para que la derivada no resulte en valores excesivamente pequeños, etc.

## AlexNet TDNN

Una vez introducido el concepto de CNN, el siguiente paso es presentar la arquitectura particular a utilizar para los experimentos. Se trata de AlexNet, una red convolucional de 8 capas ampliamente conocida por tratarse de la primera propuesta en demostrar como las *features* abstractas aprendidas por redes neuronales pueden superar ampliamente al *feature engineering*. El funcionamiento de AlexNet se presenta a través de un diagrama de bloques en la Figura 7.8, mientras que la Tabla 7.1 incorpora el detalle de las dimensiones de cada operación relacionada. Cabe notar que la función de activación utilizada en cada bloque es ReLU, salvo por  $FC3$  que lleva una función *softmax*.

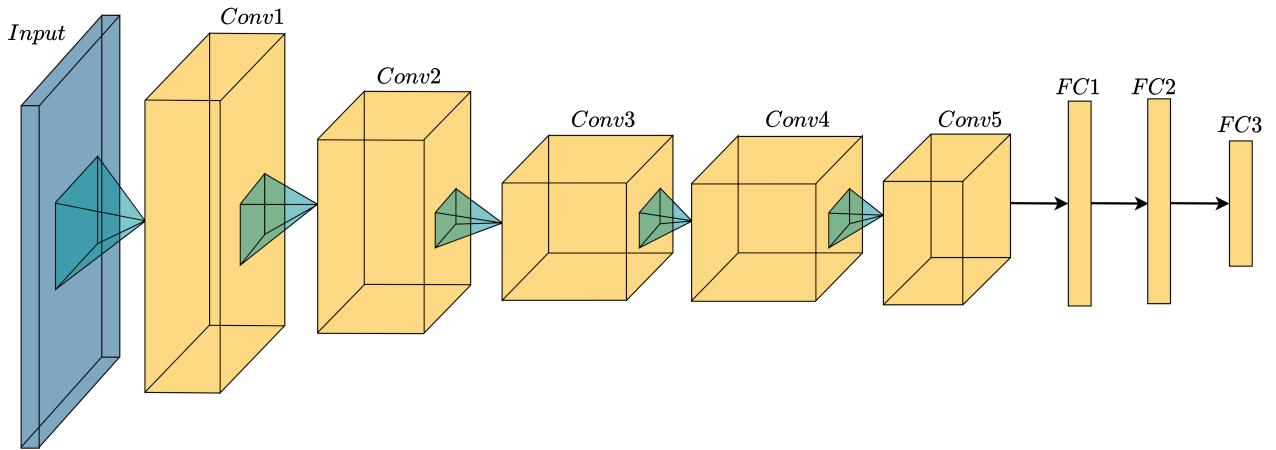


Figura 7.8: Diagrama de bloques 3D para AlexNet.

Dicha arquitectura será utilizada en este caso como *Threshold Dependent Neural Network* (TDNN), es decir, como una red neuronal a la cual se le cambiará el acto final a realizar sobre la salida para que en lugar de predecir una única clase prediga un conjunto de clases o etiquetas, lo cual se logrará a través de la definición de un umbral de asignación según probabilidad y el uso de una función sigmoide.

---

<sup>1</sup>Producto de la normalización. Cuando esto no se cumple, el problema se mantiene, pero los valores tienden a infinito en lugar de 0. Se habla entonces de *exploding gradient*.

Tabla 7.1: Detalle arquitectura AlexNet

Capa	Tipo	Dim. entrada	Dim. filtro	Stride	Padding	Dim. salida
Datos	Input	$227 \times 227 \times 3$	N/A	N/A	N/A	$227 \times 227 \times 3$
Conv1	Convolución	$227 \times 227 \times 3$	$11 \times 11$	4	0	$55 \times 55 \times 96$
Pool1	Max Pooling	$55 \times 55 \times 96$	$3 \times 3$	2	0	$27 \times 27 \times 96$
Conv2	Convolución	$27 \times 27 \times 96$	$5 \times 5$	1	2	$27 \times 27 \times 256$
Pool2	Max Pooling	$27 \times 27 \times 256$	$3 \times 3$	2	0	$13 \times 13 \times 256$
Conv3	Convolución	$13 \times 13 \times 256$	$3 \times 3$	1	1	$13 \times 13 \times 384$
Conv4	Convolución	$13 \times 13 \times 384$	$3 \times 3$	1	1	$13 \times 13 \times 384$
Conv5	Convolución	$13 \times 13 \times 384$	$3 \times 3$	1	1	$13 \times 13 \times 256$
Pool5	Max Pooling	$13 \times 13 \times 256$	$3 \times 3$	2	0	$6 \times 6 \times 256$
FC1	Fully Connected	$6 \times 6 \times 256$	$6 \times 6$	1	0	$4096 \times 1$
FC2	Fully Connected	$4096 \times 1$	$1 \times 1$	1	0	$4096 \times 1$
FC3	Fully Connected	$4096 \times 1$	$1 \times 1$	1	0	$N_{classes}$

AlexNet TDNN se plantea como una alternativa mucho más simple en términos de funcionamiento que C2AE, pero de mayor profundidad: mientras C2AE incorpora en su componente más profundo (*autoencoder*) un total de 5 capas *fully-connected*, AlexNet incorpora 8 capas de las cuales las 5 primeras son convolucionales y las últimas 3 *fully-connected*. Así mismo, mientras C2AE busca una representación conjunta de *features* y etiquetas, AlexNet solo apunta a una representación profunda de las primeras. Se espera que ello aporte tanto con ventajas como desventajas al problema en estudio, a saber:

- Por un lado, se considera que la flexibilidad que AlexNet ofrece para manipular hiperparámetros como la función de pérdida podría traer beneficios en cuanto abre la posibilidad de utilizar técnicas para tratar el desbalance de casos negativos y positivos, lo cual habría demostrado jugar un rol esencial en el Enfoque Tradicional y que, sin embargo, no sería posible con C2AE dado el uso de funciones de pérdida especialmente diseñadas para alcanzar el *joint embedding*.
- Similarmente, la literatura sugeriría que redes más profundas detectarían relaciones de más alto nivel [30], por lo cual AlexNet debería llevar la ventaja en este aspecto.
- El intercambio del *joint embedding* por un *embedding* exclusivamente de *features*, por otro lado, podría poner a AlexNet en desventaja frente a C2AE.
- Por último, la desventaja quizás más importante de AlexNet vendrá dada por la aplicación como TDNN, en cuanto la elección del ya mencionado umbral de asignación según probabilidad incorporará una tarea adicional que deberá ser resuelta de forma experimental, perdiendo así la propiedad de ser un método entrenable *End-to-End* como es el caso de C2AE.

## Implementación, configuración e hiperparámetros

AlexNet TDNN será puesta a prueba mediante su implementación en *Pytorch*, haciendo uso de que la arquitectura original forma parte de las redes incorporadas en la librería (por lo cual solo basta cambiar la última capa y manipular las salidas según el umbral de asignación). Aún más, *Pytorch* incorpora también la posibilidad de invocar la arquitectura con pesos preentrenados para la tarea multiclas de ImageNet, lo cual se utilizará como punto de partida para todos los experimentos.

La configuración específica a utilizar, por otro lado, será aquella de la Figura 7.8, bajo la salvedad de que el tamaño de la salida variará según la cantidad de etiquetas en estudio (cambiando  $N_{classes}$  en la Tabla 7.1 por  $N_{labels}$ ). Por ejemplo, para la tarea reducida a 26 etiquetas, la capa *FC3* tomará una dimensión de salida igual a 26, a fin de representar la probabilidad de cada etiqueta por una componente de la salida.

Los umbrales de asignación de etiquetas, por otro lado, dependerán del experimento (se especificarán según corresponda) mas mantendrán las siguientes simplificaciones:

- El umbral de asignación será el mismo para todas las etiquetas,
- El umbral de asignación será el mismo en entrenamiento, validación y test.

Finalmente, el entrenamiento se llevará a cabo utilizando los siguientes hiperparámetros:

- Función de pérdida: *BCEWithLogitsLoss*, función disponible en *Pytorch* y correspondiente a la combinación de una sigmoide seguida de la *Binary Cross Entropy Loss* tradicional.
- Número de épocas: 300
- *Batch size*: 128
- Optimizador: Adam
- *Learning-rate* inicial: 0,005
- *Weight-decay*: 0,001
- Uso de *Early-Stopping* con paciencia de 15 épocas
- *Scheduler*: *ReduceLRonPlateau* con un factor de reducción del 0,75, un *learning rate* mínimo de 0,0001, una paciencia de 10 épocas y con el micro F<sub>2</sub>-Score como objetivo.

### 7.1.3. CNN-RNN

Antes de describir CNN-RNN, resulta preciso introducir tres conceptos adicionales necesarios para el entendimiento de este. Se trata de las redes recurrentes (RNN), la red recurrente *Long-Short Term Memory* (LSTM) y el algoritmo de búsqueda *Beam Search*.

## RNN

Las RNN son un tipo particular de redes neuronales pensadas para el tratamiento de datos secuenciales y de tamaño arbitrario mediante la incorporación del concepto de “memoria” en las neuronas, lo cual permite modelar el comportamiento dinámico de las secuencias.

Para ello, las RNN se definen de manera recursiva: la entrada en un instante  $t$  da lugar a una salida que es, a su vez, la entrada del instante  $t + 1$ . Esto puede verse como la inclusión de ciclos en las neuronas, lo cual da lugar a múltiples “estados ocultos”  $s_i$  que corresponden a la salida de dichos ciclos en un instante  $t_i$ . La Figura 7.9 presenta un esquema de dicha configuración, donde  $x_i$ ,  $y_i$  y  $s_i$  representan a la entrada, salida y estado oculto en un instante  $i$ , mientras que  $O$  y  $R$  corresponden a funciones parametrizadas por  $\theta$  utilizadas para calcular la salida y el estado oculto siguiente, respectivamente. En particular, 7.9a presenta una abstracción de RNN mediante un único componente cíclico, mientras que 7.9b presenta la versión descomprimida a múltiples componentes idénticos puestos en secuencia.

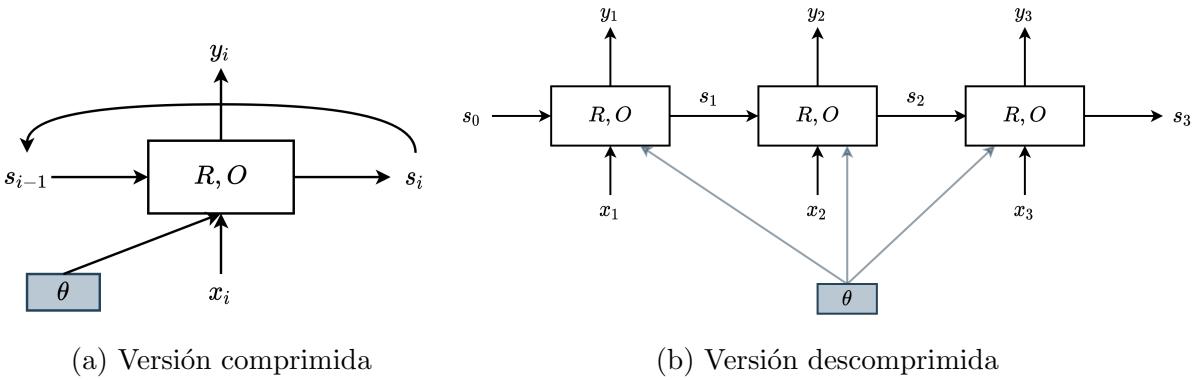


Figura 7.9: Representación abstracta de una RNN.

Son dichos estados ocultos los que, en teoría, permitirían codificar el historial de la secuencia y entregarlo a los nodos más tardíos, dotando así a la arquitectura de memoria a corto y largo plazo. En la práctica, sin embargo, dicha propiedad se ve limitada al corto plazo por efecto del desvanecimiento del gradiente en el entrenamiento, lo cual dificulta el paso de la señal de error a las iteraciones más tempranas. Dicha dificultad, presentada en la introducción de las CNN (Sección 7.1.2), se hace también presente para las RNN y ha sido igualmente abordado a través de distintos planteamientos, como se verá en la introducción de LSTM.

Un último punto a destacar sobre las RNN es su similitud con los modelos ocultos de Markov. En este sentido, las RNN pueden interpretarse como una extensión a los mismos, la cual incorpora funciones no lineales para la transición entre nodos y que no requiere del supuesto de Markov sobre la independencia de un estado con sus antecesores a partir de cierto nivel, tratándose entonces de una técnica mucho más poderosa.

## LSTM

El siguiente concepto a introducir es el de LSTM, una arquitectura particular de RNN la cual busca dar solución al problema de la memoria a corto plazo a través de la división de los estados ocultos en dos partes: una primera que mantiene el funcionamiento a corto plazo del estado oculto original y una segunda denominada “celda de memoria”, diseñada para preservar la memoria y los gradientes en el tiempo mediante un sistema de funciones suaves y diferenciables (y en consecuencia, entrenables) que imitan el comportamiento de compuertas lógicas y que deciden qué parte de la entrada pasa a memoria y qué parte es olvidada.

En particular, LSTM incorpora:

- Una compuerta de entrada  $i$ , que decide qué partes de la entrada se preservan en el nuevo estado,
- Una compuerta de olvido  $f$ , que decide qué partes de la memoria se olvidan en el nuevo estado,
- una compuerta de salida  $o$ , que decide qué partes de la entrada y memoria se entregan como salida (nuevo estado oculto).
- Un candidato de actualización  $z$ , el cual se computa como la combinación lineal de la entrada y el estado oculto previo, pasado por una función de activación no lineal.

De esta manera, para actualizar la celda de memoria  $c_i$  a  $c_{i+1}$ :

- Se olvidan los componentes de  $c_i$  que  $f$  dicte,
- Y se mantienen los componentes de  $z_i$  según lo que la función  $i$  indique.

Tras lo cual, para calcular la actualización del estado oculto  $s_{i+1}$ :

- Se calcula la tangente hiperbólica de la información en la celda actualizada, y se entrega como salida lo que la función  $o$  aplicada sobre dicho cálculo entregue.

Matemáticamente, dicho proceder corresponde a lo planteado en la Ecuación 7.9, donde  $x_j$  corresponde al componente  $j$  de la entrada,  $\odot$  al producto hadamardiano,  $\sigma$  a una función de activación y las distintas matrices  $W$  a parámetros ajustables:

$$\begin{aligned}
 s_j &= R(s_{j-1}, x_j) = [c_j; h_j] \\
 c_j &= f \odot c_{j-1} + i \odot z \\
 h_j &= o \odot \tanh(c_j) \\
 i &= \sigma(x_j W^{xi} + h_{j-1} W^{hi}) \\
 f &= \sigma(x_j W^{xf} + h_{j-1} W^{hf}) \\
 o &= \sigma(x_j W^{xo} + h_{j-1} W^{ho}) \\
 z &= \tanh(x_j W^{xz} + h_{j-1} W^{hz}) \\
 y_j &= O(s_j) = h_j
 \end{aligned} \tag{7.9}$$

Tal configuración permite que el gradiente relacionado a la celda de memoria se mantenga alto durante períodos prolongados de entrenamiento, lo cual dota a la RNN subyacente de la deseada memoria a largo plazo (manteniendo además la propiedad a corto plazo a través del estado oculto), convirtiéndose así en una herramienta sumamente poderosa para el tratamiento de datos secuenciales.

### **Beam Search**

El último concepto a introducir antes de entrar al detalle de CNN-RNN es el de *Beam Search*, un algoritmo heurístico para la búsqueda eficiente de secuencias más probables a través de la generación de todos los sucesores posibles en un cierto instante mas la mantenición de solo los  $w$  caminos más prometedores. La Figura 7.10 muestra un ejemplo de dicho algoritmo aplicado con un  $w = 2$ .

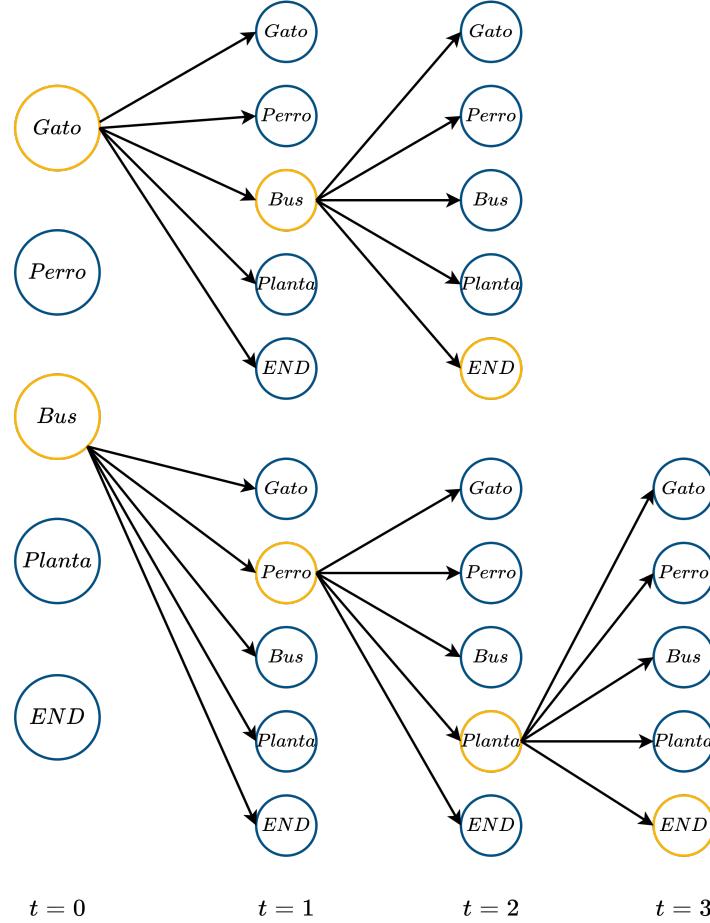


Figura 7.10: Ejemplo de *Beam Search* con  $w = 2$ . Nótese que en cada paso el camino candidato (conformado por los nodos en amarillo) es evaluado con todos los posibles sucesores para el nivel siguiente.

Más específicamente, *Beam Search* se basa en la búsqueda de los  $w$  caminos  $P_t$  de predicción más probables en cada instante  $t$  (cada nivel del árbol). Dichos caminos pasan a

denominarse camino intermedio, y dan lugar al conjunto  $S(t)$  (Ecuación 7.10):

$$S(t) = \{P_1(t), P_2(t), \dots, P_N(t)\} \quad (7.10)$$

De esta manera, dada una configuración para el instante  $t$ , el paso hacia el instante  $t + 1$  se lleva a cabo generando todos los posibles sucesores para cada cada camino  $P_i(t)$ , tras lo cual se lleva a cabo un proceso de selección en donde se mantienen solo los  $w$  caminos con la probabilidad más alta, los cuales pasan a conformar el nuevo conjunto  $S(t + 1)$ .

- En el ejemplo de la Figura 7.10, en  $t = 0$  los caminos intermedios serían *Gato* y *Bus* (en amarillo), mientras que en  $t = 1$  la comparación de todos los sucesores posibles para cada uno terminaría por entregar *Gato – Bus* y *Bus – Perro* como los nuevos caminos intermedios.
- Nótese que no hay restricciones que impidan que ambos caminos intermedios resulten en una misma rama: si en el ejemplo el camino *Gato – Perro* tuviese mayor probabilidad que *Bus – Perro*, ambos caminos intermedios quedarían en la rama superior del árbol de búsqueda.

Cuando un camino intermedio llega a la señal de término (*END* en el ejemplo) este sale de los caminos intermediarios para pasar a conformar el conjunto de caminos candidatos  $C$ .

- Tal sería el caso del camino *Gato – Bus – END* en la Figura 7.10 una vez realizado el segundo paso ( $t = 2$ ). Nótese que el otro camino intermedio (aquel que hasta ese instante corresponde a *Bus – Perro – Planta*) sigue aún pendiente de finalizar, por lo cual la búsqueda continua.

La búsqueda termina cuando todos los caminos intermediarios han llegado a una señal de término, es decir, cuando  $|C| = w$ , tras lo cual el algoritmo retorna el candidato de mayor probabilidad.

Un punto a destacar sobre *Beam Search* es su similitud con *Breadth-First Search* (BFS). En este sentido, *Beam Search* se basa en BFS para la construcción del árbol de búsqueda en cuanto en cada nivel genera todos los sucesores de los estados en el nivel actual, sin embargo, a diferencia de BFS, solo mantiene una cantidad predeterminada  $w$  (denominada *beam width*) de mejores estados a ser expandidos, tratándose entonces de una versión avara de BFS y de comportamiento idéntico cuando  $w \rightarrow \infty$ .

## CNN-RNN

Una vez introducidos los conceptos necesarios, finalmente es posible pasar al detalle de como dichas técnicas se reunirán en la arquitectura CNN-RNN para dar lugar a una predicción multietiqueta.

CNN-RNN es, al igual que C2AE, una técnica de *joint embedding*, pero más profunda y que cambia la forma en que las etiquetas se codifican y predicen. En particular, CNN-RNN divide la tarea en dos grandes componentes profundos: por un lado se encuentra una CNN, la cual es encomendada con la misión de extraer representaciones semánticas de las imágenes y,

por el otro, una RNN es asignada con la tarea de modelar las relaciones entre imagen-etiqueta y las dependencias entre etiquetas.

Para tal fin, en primer lugar se genera un *embedding* de etiquetas, el cual es utilizado para alimentar una celda recurrente. En paralelo, una CNN se encarga el modelado de información semántica en las imágenes, dando lugar al *embedding* de imágenes. Las salidas de ambos componentes son proyectadas al mismo espacio de baja dimensión del *embedding* de etiquetas, que pasa a denominarse “capa de proyección”. De esta manera, la descripción de la imagen, el *embedding* de las etiquetas y la salida de RNN para un instante  $t$  viven ahora en un mismo espacio, el cual es sometido a un proceso de predicción secuencial donde en cada paso la probabilidad de asignación de una etiqueta es calculada basándose en la distancia entre el *embedding* de la imagen y la salida de las neuronas recurrentes. El proceso y la arquitectura se ilustran de mejor manera en las Figuras 7.11a y 7.11b.

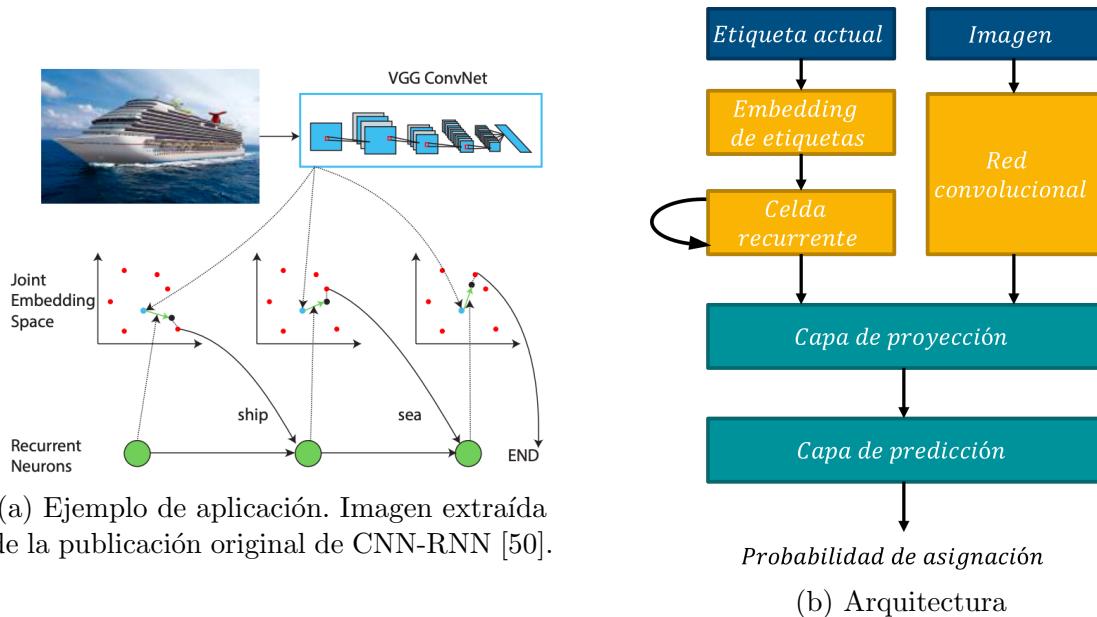


Figura 7.11: Esquemas CNN-RNN.

De esta manera, el modelo logra aprender un *joint embedding* de imagen y etiquetas de baja dimensión, lo cual permite modelar la relevancia semántica entre imágenes y etiquetas [50] entregando una probabilidad de asignación para cada etiqueta según las etiquetas previamente asignadas, siendo la señal de inicio (*START*) la etiqueta inicial para todas las imágenes.

Más en detalle, la probabilidad *a priori* de una etiqueta dada las etiquetas predichas en el camino actual se calcula como el producto punto entre estas últimas con la suma de los *embeddings* de la imagen y de la salida recurrente para la etiqueta de interés. La probabilidad de un camino de predicciones, por otro lado, se obtiene como el producto de las probabilidades *a priori* de cada etiqueta. El mejor camino de predicciones (aquel con la probabilidad global más alta) es buscado de forma aproximada en inferencia mediante la aplicación de *Beam Search*.

Cabe notar que, a diferencia de otros métodos, CNN-RNN no entrega un conjunto de etiquetas a asignar, sino más bien caminos de probabilidad entre todas las posibles deci-

siones. Dicho cambio de paradigma dota de importancia al orden en que las etiquetas sean presentadas, en cuanto el camino que toma *Perro* en  $t = 0$  y *Planta* en  $t = 1$  podría no tener la misma probabilidad del camino que toma las mismas decisiones en orden inverso, un fenómeno propio de la predicción secuencial (como ocurriese con *Classifier Chain* en el Enfoque Tradicional, Sección 6.1). Al respecto, los autores de CNN-RNN recomiendan un enfoque de “presentar las etiquetas más frecuentes primero”, bajo el supuesto de que estas son las más fáciles de asignar y cuya suma contribuye a identificar las etiquetas menos frecuentes (más complejas).

## Implementación y configuración a utilizar

La técnica de CNN-RNN será implementada en *Pytorch*. Dado que la publicación original no especifica un tipo específico de CNN a utilizar (aun cuando usa como ejemplo VGGNet), se optará por utilizar la misma arquitectura convolucional vista hasta el momento, correspondiente a AlexNet. Para el caso del componente RNN, por otro lado, los autores recomiendan la utilización de LSTM, que será entonces la arquitectura a utilizar. Nótese que tanto AlexNet como LSTM forman parte de los modelos ya incorporados en la librería.

Sobre las configuraciones específicas, por otro lado, se tiene:

- Para el *joint embedding* se utilizará una dimensión de  $N_{labels}/2$ , valor encontrado de forma experimental y que busca responder al hecho de que la dimensión de *embedding* debe ser menor a la cantidad de etiquetas para forzar la explotación de relaciones.
- Para AlexNet se mantendrá su configuración original, disponible en la Tabla 7.1. La dimensión de la salida corresponderá a la del *joint embedding*.
- Para LSTM se utilizará 1 capa oculta de 256 neuronas, valor encontrado de forma experimental utilizando el *fold 0*.

Por último, en lo que respecta a los hiperparámetros del entrenamiento, se utilizará:

- Función de pérdida: *BCEWithLogitsLoss*.
- Número de épocas: 200
- *Batch size*: 128
- Optimizador: Adam
- *Learning-rate* inicial: 0,001
- *Weight-decay*: 0,00002
- Uso de *Early-Stopping* con paciencia de 15 épocas
- *Scheduler*: *ReduceLROnPlateau* con un factor de reducción del 0,75, un *learning rate* mínimo de 0,0001, una paciencia de 10 épocas y con el micro F<sub>2</sub>-Score como objetivo.

## 7.2. Descriptores

A diferencia del Enfoque Tradicional, en el Enfoque XMLC el uso de descriptores pre-entrenados no resulta necesario pues los métodos se plantean como capaces de aprender descripciones apropiadas por sí mismos (siendo esta una de las principales ventajas del Enfoque XMLC). Por esta razón, las entradas de los métodos corresponderán simplemente a las imágenes de los patrones redimensionadas a los tamaños esperados por los métodos y/o que permitan un tiempo de cómputo razonable, a saber:

- Para el caso de AlexNet TDNN y CNN-RNN se utilizarán las imágenes redimensionadas a  $257 \times 257$  pixeles, formato impuesto por AlexNet.
- Para el caso de C2AE, por otro lado, se utilizarán las imágenes redimensionadas a  $40 \times 40$  pixeles, tamaño escogido buscando un nivel de detalle amplio pero que no significase requerir más de 3 horas de procesamiento para evaluar el método frente a todo el espectro de etiquetas en el conjunto *base*.
- Adicionalmente, para C2AE se generará también un conjunto de descriptores obtenidos en base a AlexNet preentrenada sobre ImageNet (sin reentrenamiento) a fin de replicar la configuración utilizada por sus autores en la publicación original [63].

## 7.3. Experimentos

### 7.3.1. E4: C2AE, variando descriptores y conjunto de datos

El primero de los experimentos del Enfoque XMLC corresponde al uso de C2AE como técnica para abordar el problema. En específico, se busca identificar la mejor configuración según técnica de D.A. y origen de descriptores para, basándose en ella, comentar qué tan bien responde el algoritmo al problema en cuestión.

Al corresponder C2AE a una arquitectura relativamente pequeña, el entrenamiento en GPU resulta rápido y por ende cruzar múltiples variables en un mismo experimento se vuelve factible. En específico, para esta ocasión se cruzaría la evolución al variar la cantidad de etiquetas, el origen de los descriptores y el conjunto de datos a utilizar, todo lo cual tomaría alrededor de 40 horas de procesamiento.

Los resultados se presentan a través de curvas de desempeño, como ya es costumbre (el detalle se encuentra disponible en el Apéndice D). Se comienza con las Figuras 7.12a y 7.12b, las cuales incorporan el *Micro F2-Score* de los distintos métodos al utilizar descriptores AlexNet y *flatten*, respectivamente.

Al estudiar dichas curvas destaca de inmediato la fuerte diferencia en el comportamiento de las mismas con respecto a las vistas hasta el momento en el Enfoque Tradicional: mientras estas últimas tenían un crecimiento más bien monótono en todo el recorrido, las curvas de C2AE muestran un comportamiento inicialmente errático que converge a medida que la cantidad de etiquetas aumenta, fenómeno presente en ambos conjuntos de descriptores pero especialmente notorio en el caso de 7.12a.

Aún más, al observar la correspondencia entre curvas y conjuntos, se observa que tal inicio errático se acentúa a medida que se utilizan más datos sintéticos, siendo los conjuntos 13 y 14 los que presentarían saltos más grandes en un inicio, con ambos descriptores. El conjunto base, por otro lado, sobresale por presentar una curva de decrecimiento monótono prácticamente en todo su recorrido, la cual sin embargo termina por corresponder a los resultados más bajos en el problema con 281 etiquetas.

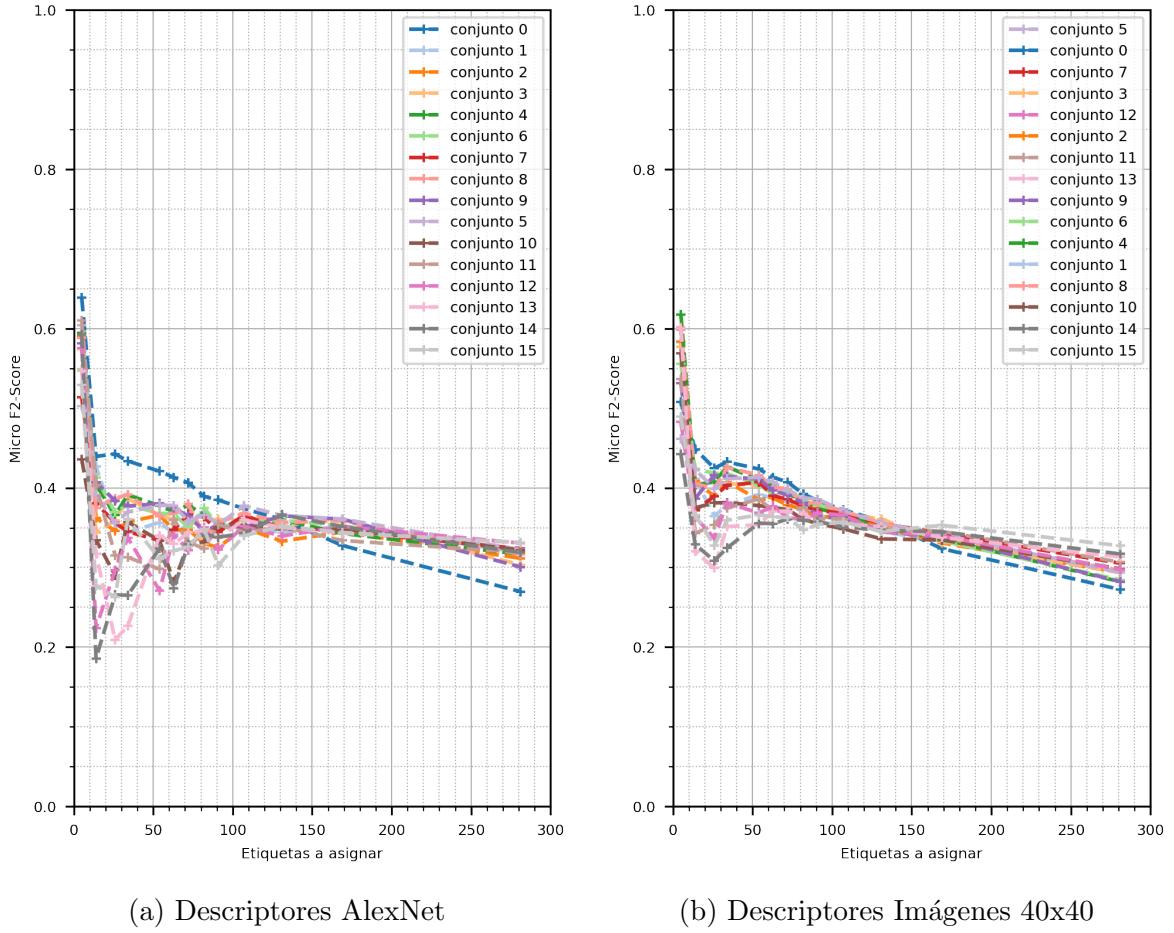


Figura 7.12: *Micro F<sub>2</sub>-Score C2AE*, variando conjunto de descriptores.

Dicho fenómeno dotaría de complejidad la elección de una configuración específica como la más apropiada pues, mientras el uso de descriptores obtenidos desde AlexNet daría lugar a curvas más erráticas en los conjuntos grandes, también daría lugar los mejores resultados en pocas etiquetas a través del conjunto 0. En cantidades elevadas, por otro lado, el comportamiento de las mejores curvas sería similar, presentando un empate en el mejor caso para 281 etiquetas con un *Micro F<sub>2</sub>-Score* de 0.3310 a través del conjunto 15.

Ahora bien, un aspecto evidente pese a la complejidad de los resultados es la inferioridad que estos presentan con respecto a los del Enfoque Tradicional, lo cual justificaría que la elección de una mejor configuración perdiera relevancia. Sin embargo, entender el motivo detrás de dicha complejidad permanecería como una cautivadora interrogante. Al respecto:

- Se especularía que la razón detrás del comportamiento de los conjuntos más grandes (que comienzan entregando los peores resultados para terminar siendo los mejores)

podría deberse a la disparidad de información entre el *embedding* de imágenes (que en el caso de los conjuntos más grandes estaría alcanzando un aprendizaje más amplio) versus la información disponible en el *embedding* de etiquetas, que en un inicio sería muy baja y posiblemente insuficiente para explicar toda la variabilidad en las imágenes.

- Lo anterior justificaría tanto el caso de los conjuntos más grandes como el del conjunto base (conjunto 0), en cuanto bajo el mismo razonamiento se podría esperar que con pocas imágenes el componente a cargo de dicho *embedding* (perceptrón multicapa) captase menos información relativa a los patrones geométricos, coincidiendo así con la información disponible en el *embedding* de etiquetas en cantidades bajas pero siendo insuficiente para cantidades más altas (explicando así su fuerte decaimiento).

Desde luego, lo anterior requeriría de experimentos particulares para su validación, mas en vista de que los resultados hasta el momento estarían muy lejos de competir con la línea base entregada por el Enfoque Tradicional (con métodos que mantendrían valores de *Micro F<sub>2</sub>-Score* sobre 0,45 en todo el espectro de etiquetas), se considera que una exploración más profunda no responde a las necesidades del problema, aun cuando podría representar una interesante línea de investigación.

### 7.3.2. E5: AlexNet TDNN, variando tratamiento y conjunto

El siguiente experimento corresponde a la evaluación de AlexNet como TDNN, lo cual se realiza en primera instancia variando el tratamiento al desbalance de datos (a fin de medir cómo este aspecto impacta en TDNN) y el conjunto de datos a utilizar (a fin de reconocer las mejores técnicas de D.A. para este caso). Al respecto:

- Resulta necesario destacar el alto costo computacional que someter una red convolucional profunda a evaluación ante múltiples variables al mismo tiempo significa. En particular, evaluar el comportamiento de AlexNet TDNN sobre 16 conjuntos de datos con 2 opciones para el tratamiento al desbalance significa entrenar un total de 32 modelos distintos, cantidad la cual se ve multiplicada por 4 al considerar la aplicación de *4-Fold Cross Validation*.

Es por esta razón que, a fin de mantener los experimentos dentro del alcance que las limitaciones de tiempo y *hardware* imponen, se opta por cambiar la metodología de comparar configuraciones en todo el espectro de etiquetas para hacerlo ahora en una cantidad fija de las mismas, asumiendo que la configuración que arroje los mejores resultados es la mejor también en todo el recorrido. Al respecto:

- Esto implica introducir un supuesto fuerte, especialmente teniendo en cuenta lo visto en el caso de C2AE (donde a distintas cantidades de etiquetas se tenían distintas mejores configuraciones).
- Sin embargo, es también un supuesto necesario: de lo contrario la cantidad de entrenamientos aumentaría a  $32 \text{ modelos} \times 13 \text{ cantidades} = 416$ , lo cual al incorporar los 4 *folds* alcanzaría un total de 1560 modelos a entrenar.

- Los dos puntos anteriores encontrarán su equilibrio en el experimento siguiente, que utilizará la salida de este para estudiar la evolución de la mejor configuración encontrada, permitiendo así observar si el comportamiento es más bien errático (como en el caso de C2AE, demostrando entonces que el supuesto era demasiado grande) o monótono (como en el caso del Enfoque Tradicional).

Bajo tal razonamiento, se opta entonces por comparar los distintos conjuntos y opciones de tratamiento al desbalance en el desafío simplificado a 107 etiquetas, cantidad escogida por tratarse de una tarea compleja, pero para la cual los métodos del Enfoque Tradicional habrían logrado buenos resultados (y por ende cabría esperar que TDNN también lo lograse).

Los resultados según *Micro F<sub>2</sub>-Score* se presentan en la Figura 7.13. El detalle para las demás métricas de interés se encuentra disponible en el Anexo E.

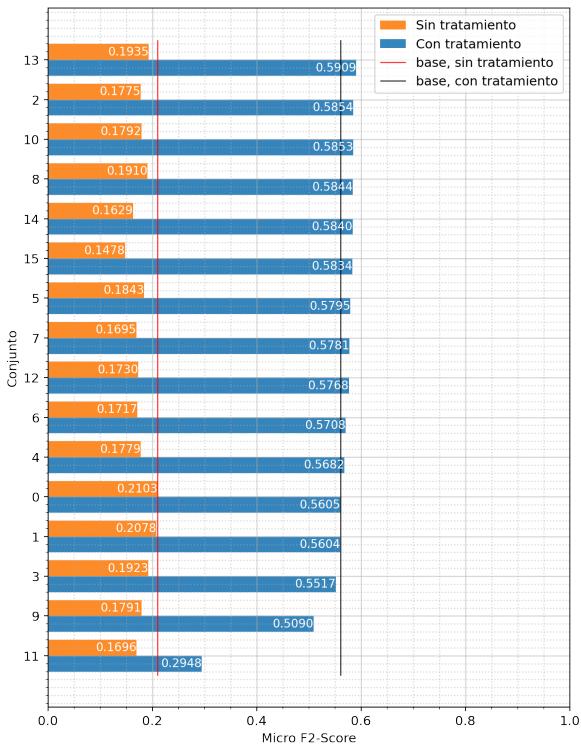


Figura 7.13: *Micro F<sub>2</sub>-Score* en 107 etiquetas, variando tratamiento al desbalance, según conjunto de datos. **Observación:** se incluyen líneas verticales a modo de guía para la comparación de los resultados de cada conjunto con los del conjunto 0 (base).

Distintas observaciones se desprenden del estudio de la Figura anterior:

- En primer lugar, se observa una evidente mejora de todos los conjuntos al incorporar pesos en el entrenamiento, alcanzando valores de F<sub>2</sub>-Score que incluso triplican el resultado sin pesos (como es el caso del conjunto 7).
- Así mismo, al utilizar las guías proporcionadas se observa que, para todos los conjuntos, el utilizar datos sintéticos sin incorporar pesos da lugar a peores resultados. Esto tiene sentido pues, en general, los datos sintéticos contribuyen al desbalance al aportar  $q$  casos positivos y  $p$  casos negativos con  $q \ll p$ .

- Por otro lado, al incorporar pesos en el entrenamiento el escenario se invierte, con 11 de 15 conjuntos superando el punaje obtenido por el conjunto 0 (base).
- Por último, el experimento habría arrojado que el mejor conjunto sería el número 13 (*rotación + reflexión + lluvia + elástico + avg. blur*), con un *Micro F<sub>2</sub>-Score* de 0,5909 en el desafío de 131 etiquetas. Destaca, sin embargo, el hecho de que el segundo lugar se lo lleva un conjunto de mucha menor cardinalidad, correspondiente al número 2 (*reflexión*).

Lo anterior permitiría dilucidar no solo una candidata a mejor configuración para TDNN, sino que también contribuiría a resolver la interrogante sobre si existe una técnica de D.A. que sea la mejor para todos los métodos al evidenciar que, bajo el supuesto principal de este experimento, tal no es el caso: mientras en el Enfoque Tradicional se viese que el conjunto más conveniente era el 14 (seguido de los conjuntos 15, 12 y 13), en este caso el mejor conjunto resultaría ser el 13, seguido por el conjunto 2.

Por último, vale notar que los resultados obtenidos esta vez sí son bastante prometedores: mientras en el Enfoque Tradicional la mejor configuración lograse un *Micro F<sub>2</sub>-Score* de 0,5436 en 131 etiquetas, en el caso de AlexNet TDNN se tiene que 14 de 16 configuraciones logran superar dicho valor y, en particular, la mejor configuración lo supera con una diferencia de más de 4 puntos porcentuales.

### 7.3.3. E6: AlexNet TDNN, evolución en mejor conjunto

Tras haber identificado el Conjunto 13 y el uso de pesos para el balance como la configuración más prometedora para TDNN, el siguiente objetivo sería evaluar dicha configuración en todo el espectro de etiquetas, tarea a abordar en este experimento.

Aún más, el experimento en cuestión tendría también un segundo objetivo. Se trata de la verificación de hipótesis que permitan justificar el supuesto realizado en el Experimento 5, relacionado con la generalización del mejor conjunto en la tarea simplificada a 107 etiquetas al desafío sobre todo el espectro. Al respecto:

- Una primera condición se encontraría en el que la evolución del *F<sub>2</sub>-Score* presente una curva de crecimiento relativamente monótono, en cuanto una curva con muchos ascensos y descensos (como las vistas en la Sección 7.3.1 para C2AE) tendría mayor probabilidad de dar lugar a intersecciones y que por ende la mejor configuración dependa del umbral de frecuencia.
- Una segunda condición se encontraría en el que, al escoger un método con pocos ejemplos (como el conjunto 0, *base*) o uno con una cantidad intermedia de ejemplos (como el conjunto 8, *avg. blur*), la curva del conjunto 13 se posicione siempre por encima.
- Desde luego, ambas condiciones podrían considerarse como necesarias pero no son, en ningún caso, suficientes: la verificación total del supuesto requeriría realizar el experimento para todos los conjuntos, lo cual no es factible por razones ya mencionadas (Sección 7.3.2).

Lo anterior conduciría a realizar el experimento considerando no solo el conjunto 13, sino también los conjuntos 0 y 8 a modo de complemento. Los resultados se presentan en la Figura 7.14 (detalle disponible en el Apéndice F).

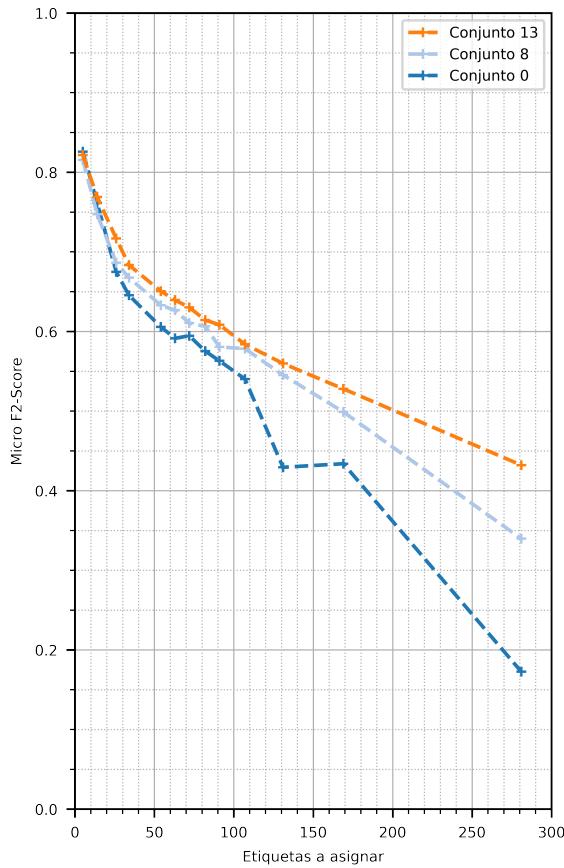


Figura 7.14: *Micro F<sub>2</sub>-Score*, múltiples conjuntos

Los resultados permiten, en primer lugar, comprobar que las condiciones se cumplen:

- Si bien las curvas de  $F_2$ -Score presentan pendientes mucho más pronunciadas que las vistas en el Enfoque Tradicional, mantienen un decrecimiento monótono en prácticamente todo su recorrido.
- De la misma forma, aun cuando la curva del conjunto 8 se aproxima bastante en un inicio a la del conjunto 13, esta última mantiene la superioridad en todo el espectro.

Ahora bien, al centrar el análisis en la curva del conjunto 13, se observa que esta mantiene valores de  $F_2$ -Score sobre 0,6 hasta las 91 etiquetas y que, aún en el caso más complejo (281 etiquetas) se mantiene sobre 0,4. Al respecto:

- Si bien la comparación de Enfoques se realizará en el Análisis de Resultados (Capítulo 8), resulta necesario mencionar que tal comportamiento posiciona de inmediato a AlexNet TDNN con la configuración escogida como una técnica competitiva, alcanzando un  $F_2$ -Score promedio de 0,6337 y un AUC de 230,0954 (versus un 0,6340 y un 244,0482 respectivos para el mejor método del Enfoque Tradicional).

Ante la identificación de AlexNet TDNN como una técnica competitiva según  $F_2$ -Score, surge entonces la necesidad de entender más a fondo sus resultados. Para ello se analizan sus curvas de *Micro Recall* y *Micro Precision*, disponibles en la Figura 7.15.

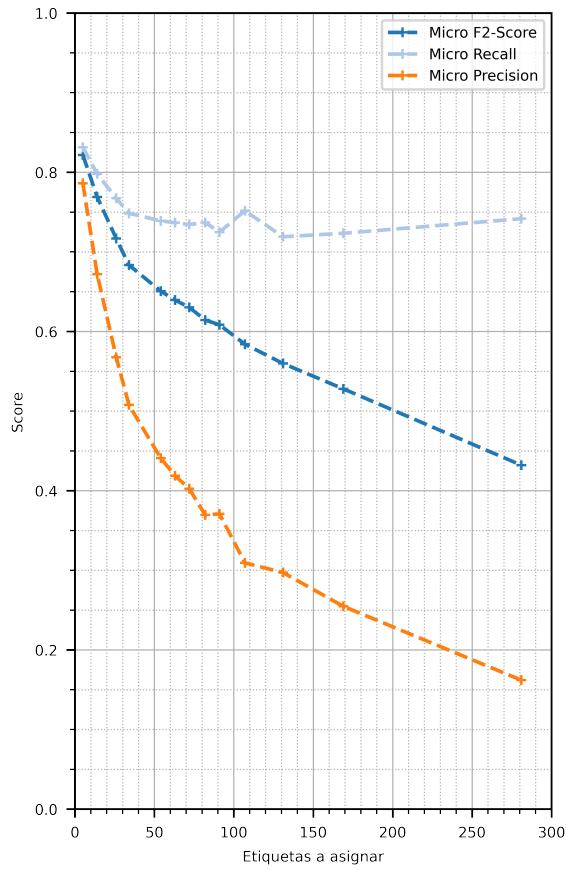


Figura 7.15: Múltiples métricas, conjunto 13

Al respecto:

- Se observa una curva de *Recall* que se mantiene sobre 0,7 y que, si bien decrece inicialmente, adquiere una pendiente positiva a partir de 131 etiquetas.
- La curva de *Precision*, por otro lado, comienza casi tan alta como el *Recall*, pero decae fuertemente - explicando así el comportamiento del  $F_2$ -Score -.

#### 7.3.4. E7: AlexNet TDNN, variando umbral de asignación

El último experimento con AlexNet TDNN corresponde al estudio del impacto de cambiar el umbral de asignación de etiquetas, es decir, la cantidad a partir de la cual una probabilidad  $p_i$  para la etiqueta  $i$  se convierte en una asignación negativa o positiva. Al respecto:

- Se opta por experimentar con valores de umbral  $t \in [0,25, 0,50, 0,75]$ .
- Para cada valor de umbral, se entrena AlexNet TDNN sobre el conjunto 13.

- En cada configuración, el valor de umbral se mantiene fijo tanto para entrenamiento como para validación y test, siendo además el mismo valor para todas las etiquetas.

Los resultados se presentan en las Figuras 7.16 (*Micro F<sub>2</sub>-Score*), 7.17a (*Micro Recall*) y 7.17b (*Micro Precision*). El detalle de las curvas se encuentra disponible en el Apéndice G.

Tabla 7.2: Valores *Micro F<sub>2</sub>-Score* Experimento 7

Umbral	5 etiquetas	54 etiquetas	281 etiquetas
$t=0.25$	0,8319	0,6464	0,4106
$t=0.50$	0,8217	0,6507	0,4321
$t=0.75$	0,7978	0,6380	0,4711

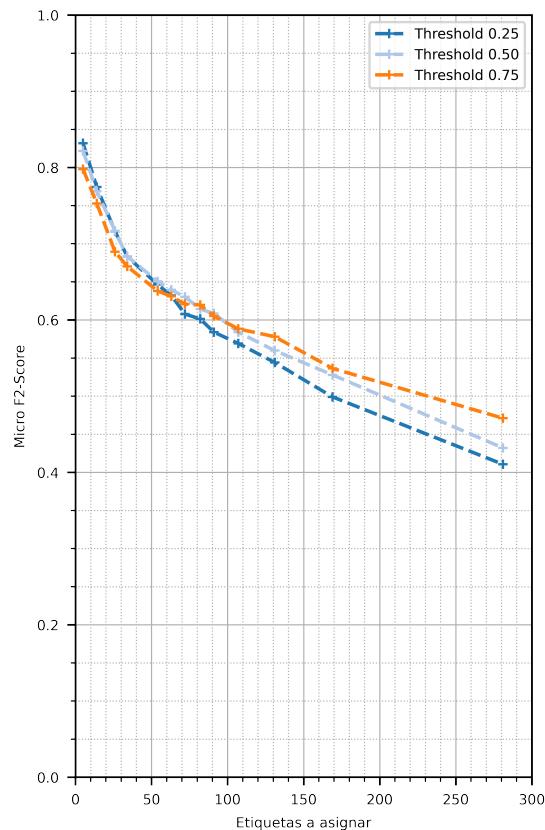
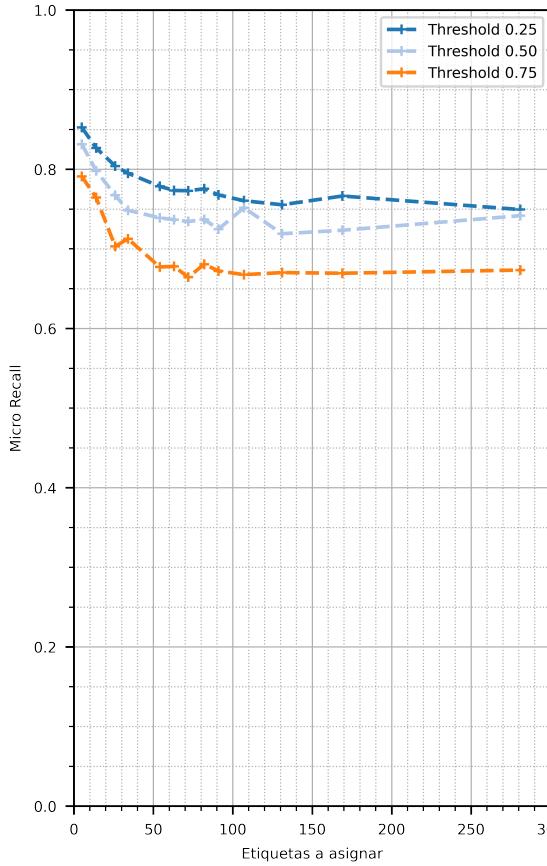
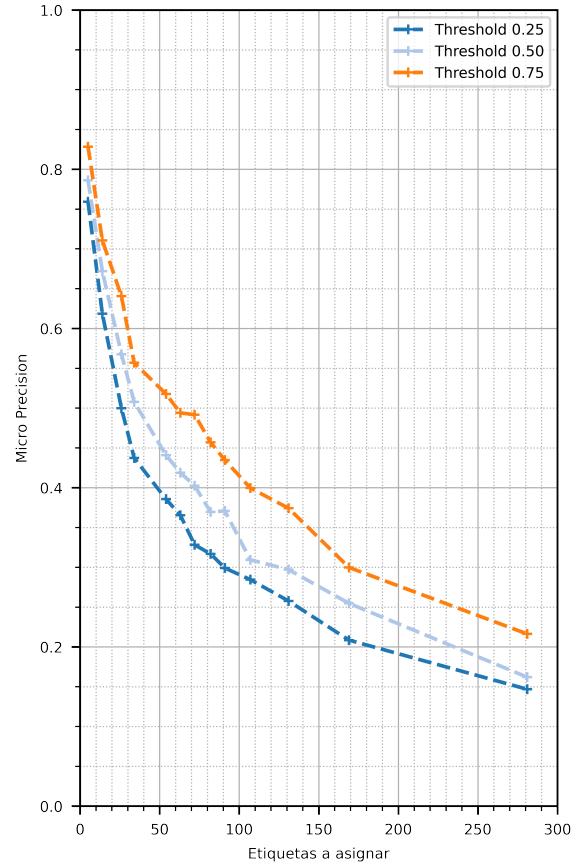


Figura 7.16: *Micro F<sub>2</sub>-Score* conjunto 13 variando umbral de asignación.



(a) *Micro Recall*



(b) *Micro Precision*

Figura 7.17: *Micro Precision* y *Micro Recall* conjunto 13 variando umbral de asignación.

Al respecto:

- Las curvas de  $F_2$ -Score evidencian un fenómeno particular: mientras en cantidades de etiquetas menores a 100 las tres configuraciones muestran comportamientos muy similares (con curvas muy cercanas para los casos  $t \in [0,25, 0,50]$  y quedando levemente por debajo la configuración con  $t = 0,75$ ), a medida que la cantidad de etiquetas aumenta las curvas comienzan a distanciarse, lo cual concluye con la curva de  $t = 0,75$  superando ampliamente al resto (véase Tabla 7.2).
- Al estudiar las curvas de *Recall* y *Precision*, por otro lado, se observa como el cambio en el umbral impacta en el *trade-off* entre ambas métricas: mientras aumentar dicho valor implica una decisión “más segura” que conlleva a una mejor *Precision* (lo cual se refleja en que la curva que maximiza esta métrica sea aquella de  $t = 0,75$ ), esta exigencia también tiene por efecto la disminución del *Recall* (que se ve maximizado por la configuración con  $t = 0,25$ ).
- Aún más, al estudiar la magnitud del cambio en *Recall* y *Precision* se explica también el fenómeno descrito en el primer punto: mientras el *Recall* comienza alrededor de 0,80 y disminuye hasta 0,65 en el peor caso, la *Precision* pasa de valores cercanos a 0,80 a resultados bajo 0,25 en el desafío con 281 etiquetas. Es decir, la pérdida en

*Precision* llega a ser mayor que el doble de la pérdida en *Recall*, lo cual determina que la configuración ganadora en el largo plazo sea aquella que minimice la pérdida en *Precision*.

- Este fenómeno se ve además potenciado por el hecho de que mientras el *Recall* se “estanca” rápidamente (acercándose a un valor final a partir de las 100 etiquetas y manteniéndose alrededor de dicho valor en el resto del espectro), la *Precision* continúa disminuyendo en todo el recorrido.

El experimento cumpliría entonces con su objetivo, en cuanto permitiría evaluar cómo el cambio en el umbral impacta en los resultados (y por ende, en el eventual sistema recomendador). La conclusión en este sentido sería que valores de umbral más elevado permiten alcanzar un mayor nivel de precisión en las sugerencias, mas al costo de un leve sacrificio en completitud.

Ahora bien, al comparar las distintas configuraciones bajo la métrica principal y en especial al considerar la ventaja que el caso con  $t = 0,75$  presenta en los escenarios más complejos (que son también los más interesantes), una consecuencia natural del experimento es la de plantear dicha configuración de TDNN como la mejor configuración encontrada. Esto se ve además complementado por el AUC (Tabla 7.3), donde la curva del caso  $t = 0,75$  se posiciona como la con mejores resultados.

Tabla 7.3: AUC *Micro F<sub>2</sub>-Score* Experimento 7

Umbral	AUC
$t=0.25$	222,8762
$t=0.50$	230,0954
$t=0.75$	234,6024

- Desde luego, la variable del umbral tiene aún mucho por explorar y representa una importante línea de investigación tanto en lo que corresponde a evaluar el impacto de usar distintos valores para entrenamiento e inferencia como en cuanto a la posibilidad de utilizar distintos valores para cada etiqueta (por ejemplo, siendo más permisivos con las etiquetas más difíciles de predecir, y más exigentes con las más fáciles). Este aspecto se visitará nuevamente en el Trabajo Futuro (Sección 9.1).

### 7.3.5. E8: CNN-RNN, variando conjunto en tarea simplificada

El último modelo a poner a prueba corresponde a CNN-RNN. Para ello se opta por mantener la metodología utilizada en TDNN, a saber, evaluar en primera instancia las distintas técnicas de D.A. en 107 etiquetas para, posteriormente, estudiar el comportamiento de las mejores configuraciones en el espectro completo. Este experimento aborda la primera parte.

Cabe mencionar que, dado que CNN-RNN incorpora en su modelo el uso de AlexNet y que en el estudio de TDNN se demostrase la conveniencia de utilizar pesos en el entrenamiento de dicha red (Sección 7.3.2), se optará por asumir que dicho fenómeno se mantiene en este caso y en consecuencia mantener el uso de pesos en el entrenamiento.

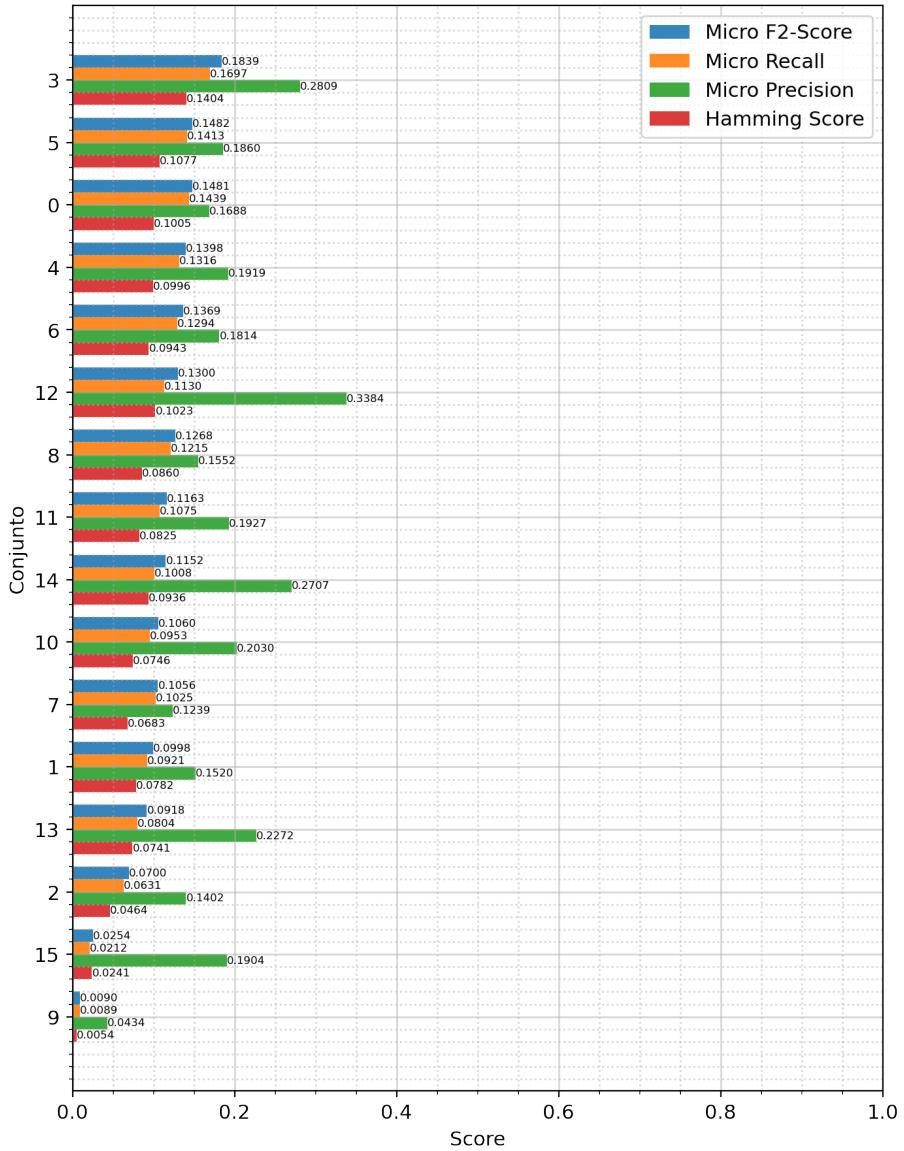


Figura 7.18: *Micro F<sub>2</sub>-Score, Recall, Precision y Hamming Score* por conjunto en 131 etiquetas, Experimento 8. **Observación:** los conjuntos están ordenados de forma descendente según *F<sub>2</sub>-Score*.

La Figura 7.18 presenta las principales métricas obtenidas por cada configuración. Al respecto:

- En primer lugar, destaca el hecho de que los valores de *F<sub>2</sub>-Score* son en general bajos, observándose un máximo de 0,1839 en el conjunto 3 y un mínimo de 0,0090 en el conjunto 9.
- Se observa también que, para cada método, el orden descendiente de las métricas es en general *Precision > F<sub>2</sub> > Recall > Hamming Score*, con valores de *Precision* que sobresalen ampliamente por sobre el resto y alcanzando máximos de hasta 0,3384 (conjunto 12).

- No se observa, por otro lado, una relación clara entre la calidad de las métricas y el tamaño de los conjuntos, pues si bien los 5 conjuntos con mayor  $F_2$ -Score corresponden al uso de una única técnica de D.A., otros conjuntos que comparten esta característica (como los conjuntos 1, 2 y 7) se posicionan entre los con peores resultados.

Los resultados sugerirían que CNN-RNN no estaría logrando el aprendizaje esperado, sino más bien un comportamiento ingenuo bajo el cual priorizaría la asignación de unas pocas etiquetas (las más frecuentes) a la mayoría de los patrones, alcanzando así un buen valor de *Precision* pero un bajo *Recall*. A fin de validar esta hipótesis, se estudian las primeras 35 matrices de confusión para el conjunto 3 (Figura 7.19):

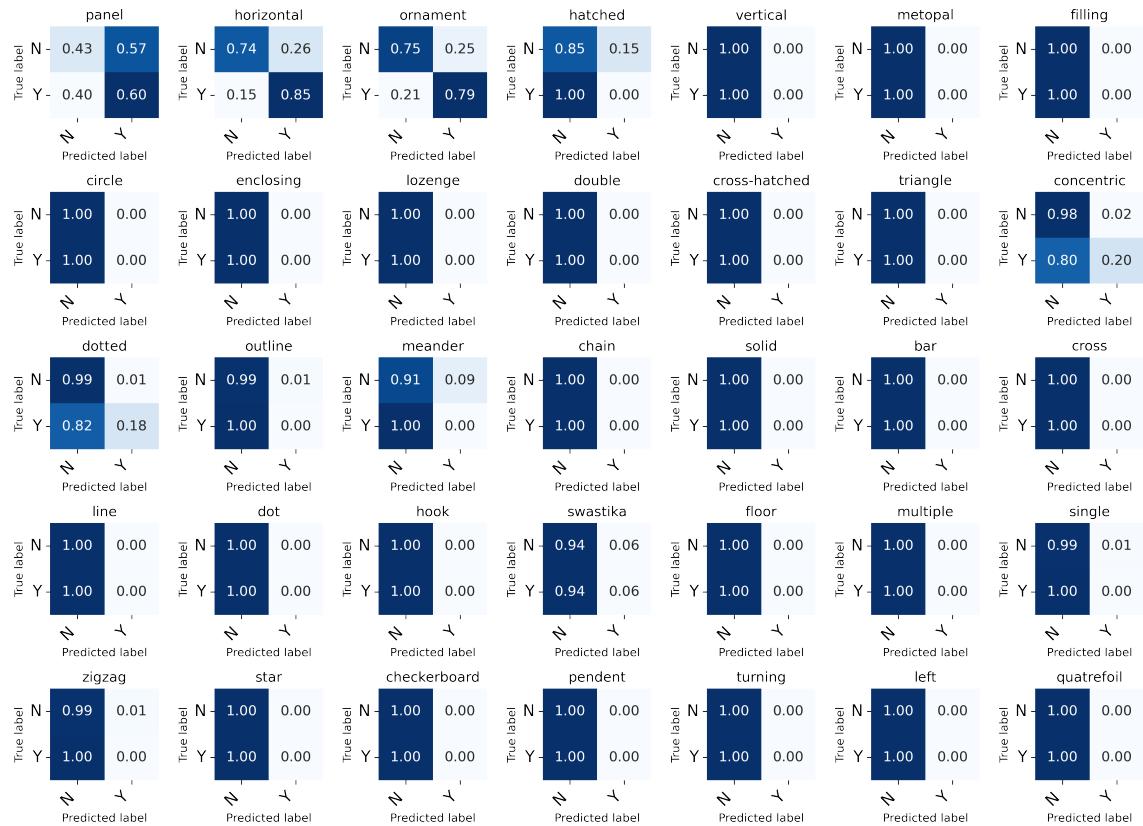


Figura 7.19: Primeras 35 matrices de confusión en 107 etiquetas, obtenidas a partir del *fold* 1, Experimento 8

Las matrices de confusión mostrarían que la hipótesis sería correcta, en cuanto CNN-RNN estaría optando por la no asignación en la mayoría de los casos, siendo las etiquetas *panel*, *horizontal* y *ornament* las únicas excepciones.

De esta forma, CNN-RNN se posicionaría como un modelo no competitivo para el problema en 107 etiquetas, lo cual podría llevar a considerar que su exploración en mayor profundidad no fuese de mayor interés para la investigación. Sin embargo, la posibilidad de que el método mejore en otras cantidades de etiquetas o las posibles razones detrás de los resultados son aún incógnitas interesantes de explorar, razón por la cual se opta por utilizar la salida de este experimento solamente como un indicador de la mejor configuración (objetivo original del experimento) y propagar dicha información al experimento

ponder tales interrogantes. En particular, se escoge el caso con el conjunto 3 como la mejor opción (decisión basada en la métrica principal, *Micro F<sub>2</sub>-Score*).

### 7.3.6. E9: CNN-RNN, evolución mejores conjuntos

El último experimento a conducir corresponde al estudio de la evolución de CNN-RNN sobre el mejor conjunto indicado por el Experimento 8 (Sección 7.3.5) en todo el espectro de etiquetas. Al respecto:

- Dicho conjunto corresponde al número 3 (*recorte*).
- Por completitud, se complementará con la evolución del conjunto 0 (*base*) y el conjunto 15 (todas las técnicas), de manera similar a como se hizo con C2AE (Sección 7.3.2).
- Esto permitirá tanto reconocer si la superioridad del conjunto 3 se mantiene en todo el espectro o es particular al caso con 107 etiquetas, como el responder a si la cantidad de datos sintéticos influye de manera distinta en diferentes niveles de dificultad.

El detalle de los resultados según las diferentes métricas se presenta en el Anexo H. A continuación se presentan las curvas de *F<sub>2</sub>-Score* de cada conjunto (Figura 7.20).

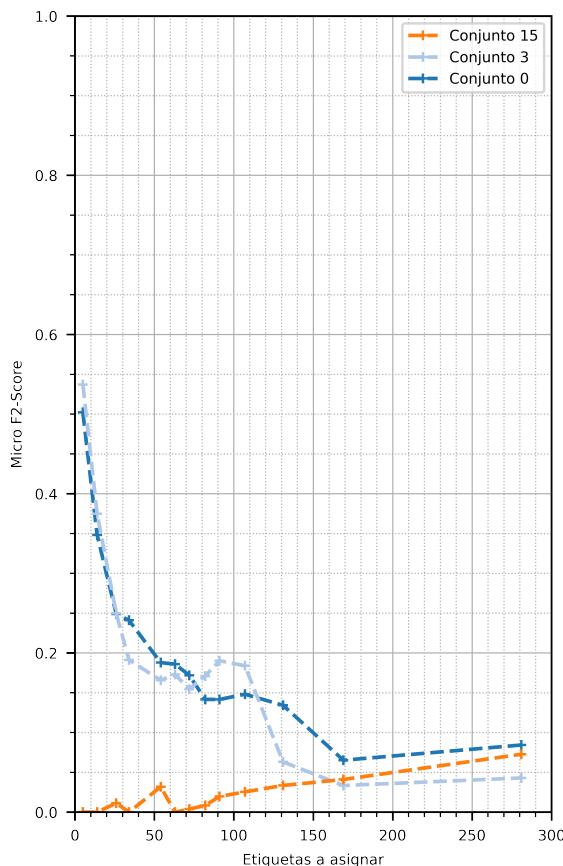


Figura 7.20: *Micro F<sub>2</sub>-Score*, múltiples conjuntos

Al respecto:

- Se observa que, salvo por las cantidades de etiquetas menores a 50, las tres configuraciones evaluadas se mantienen con un  $F_2$ -Score menor a 0,2, ampliamente inferior a lo visto incluso con C2AE.
- Se observa también que la superioridad del conjunto 3 correspondería efectivamente a un fenómeno particular del caso con 107 etiquetas, lo cual no generaliza a todo el espectro.
- Aún más, se observa que salvo por las cantidades de etiquetas 91 y 107, la configuración con el conjunto *base* es la que mejores puntajes alcanza.
- Destaca también el hecho de que si bien los conjuntos 0 y 3 comienzan con puntajes relativamente altos y decaen a medida que la dificultad del problema aumenta, el conjunto 15 comienza con puntaje prácticamente nulo y asciende hasta posicionarse próximo al mejor puntaje en el caso con 281 etiquetas (con un valor de 0,0727 contra un 0,0843 del conjunto 0).
- Por último, destaca el cambio de signo que la pendiente de las configuraciones respectivas al conjunto 0 y 3 experimentan en 169 etiquetas, pasando a presentar una mejora en el caso con 281 etiquetas (de 0,0652 a 0,0843 en el conjunto 0 y de 0,0334 a 0,0429 en el conjunto 3).

Los fenómenos anteriores permitirían confirmar que CNN-RNN no formaría parte de los métodos competitivos para el problema y refutarían también la hipótesis de que el conjunto 3 dominaría en todo el espectro, además de evidenciar una clara diferencia en el comportamiento de las configuraciones según la cantidad de datos sintéticos. Sobre esto último, se teoriza que las razones para tal efecto podrían corresponder con las planteadas para el mismo fenómeno en el caso de C2AE (Sección 7.3.2), relativas a una asimetría de información disponible en los distintos *embeddings* cuando hay muchos patrones y pocas etiquetas.

Desde luego, tal teoría requeriría una exploración más profunda para su verificación. Sin embargo, dada la no competitividad del método para el problema en estudio, se opta por dejar dicha exploración como propuesta de trabajo futuro.

# Capítulo 8

## Análisis de Resultados

Este capítulo aborda el análisis global de los resultados obtenidos y el cómo ello se relaciona el objetivo principal de la investigación, a saber, construir un sistema de apoyo al etiquetado. Para tal fin, primero se comienza estudiando en profundidad los mejores resultados obtenidos por cada enfoque, con énfasis en la comparación entre ambos (Sección 8.1). Luego se aborda la elección de un método en particular para formular la herramienta de apoyo buscada (8.2).

### 8.1. Comparación de enfoques

El análisis particular a cada experimento habría posicionado a los siguientes métodos como la mejor alternativa para cada enfoque según el espectro de cantidades de etiquetas posibles:

1. *Binary Relevance* sobre *Logistic Regression* (BR LR) con descriptores obtenidos a partir de ResNet18, utilizando pesos para el balance de casos y las técnicas de D.A. correspondientes al conjunto 14 (*rotación + reflexión + lluvia + distorsión elástica + avg. blur + recorte*) como mejor alternativa del Enfoque Tradicional.
2. AlexNet como *Threshold Dependent Neural Network* (TDNN) utilizando pesos para el balance de casos y un umbral de asignación  $t = 0,75$  sobre el conjunto 13 (*rotación + reflexión + lluvia + distorsión elástica + avg. blur*) como mejor alternativa del Enfoque XMLC.

Tales decisiones se habrían realizado considerando el *Micro F<sub>2</sub>-Score* como métrica principal y utilizando también *Micro Recall* y *Micro Precision* para comprender el comportamiento del  $F_2$ . Esto aportaría una buena noción de qué tan bien se estaría cubriendo el desafío desde un punto de vista de métricas basadas en etiquetas, sin embargo, aún restaría realizar dicho análisis desde una perspectiva de comparación entre enfoques y evaluar también cómo dicha información se traduce a métricas por ejemplos. Dichas tareas se abordan a continuación.

### 8.1.1. Comparación según métricas basadas en etiquetas

La primera interrogante a resolver corresponde a qué método (o, en este punto, qué enfoque) se posicionaría como el mejor en cuanto a la métrica principal,  $F_2$ -Score. Para ello se construye nuevamente una visualización de curvas de desempeño, pero esta vez con las mejores configuraciones de cada enfoque como protagonistas (Figura 8.1, a continuación).

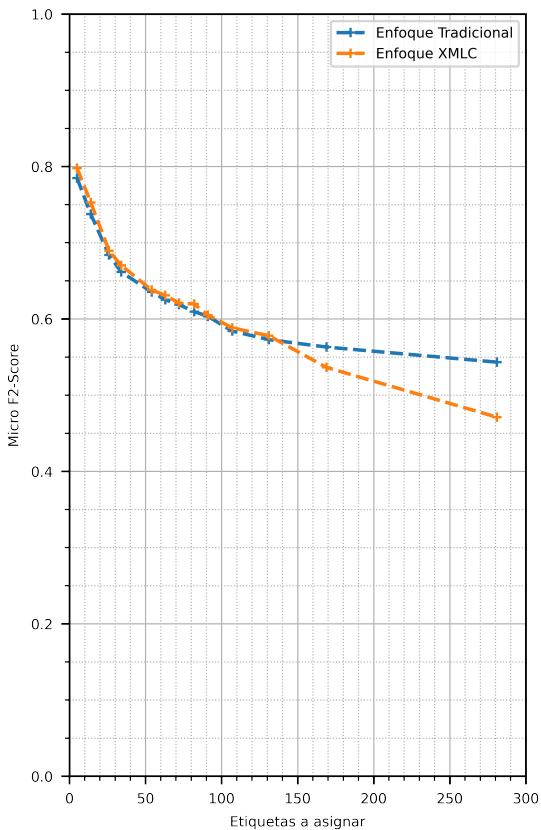


Figura 8.1: *Micro F<sub>2</sub>-Score*

De la Figura 8.1 se observa que, si bien las curvas correspondientes a cada enfoque comparten un comportamiento prácticamente idéntico en gran parte del espectro de cantidades de etiquetas, comienzan a distanciarse una vez que la cantidad asciende sobre 131, momento a partir del cual el Enfoque XMLC comienza a quedarse ampliamente por detrás hasta llegar a una diferencia de 0,0722 puntos en el caso con 281 etiquetas. Esto conlleva también a una menor área bajo la curva para dicho enfoque, como se presenta en la Tabla 8.1.

Tabla 8.1: AUC *Micro F<sub>2</sub>-Score* mejor método de cada enfoque

Enfoque	AUC
Tradicional	244,0482
XMLC	234,6024

Lo anterior podría causar sorpresa, en cuanto indica que al menos según *Micro F<sub>2</sub>-Score*

los métodos tradicionales lograrían mantenerse por sobre los métodos más modernos (y, de hecho, lo haría un método “ingenuo” como lo es *Binary Relevance*). Ante la necesidad entonces de comprender en mayor detalle este fenómeno, se comparan también las curvas de *Micro Recall* (8.2a) y *Micro Precision* (8.2b).

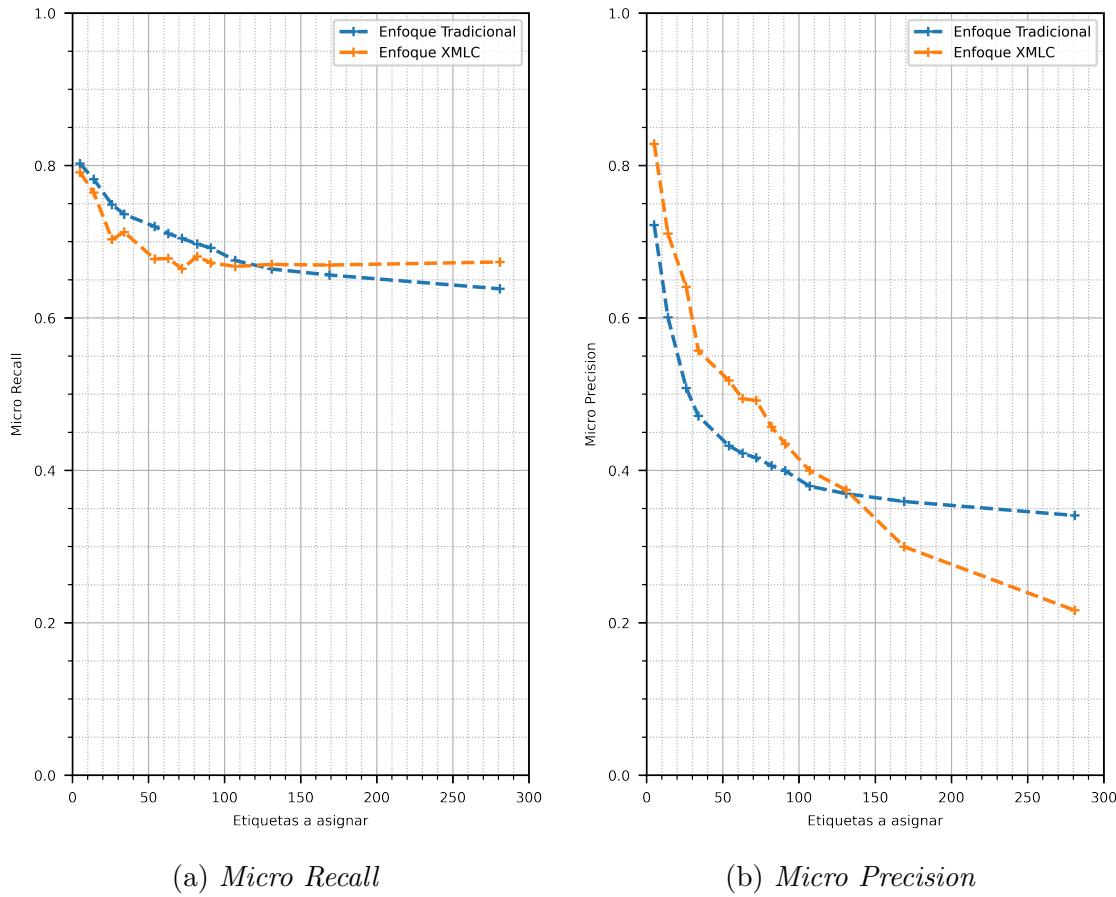


Figura 8.2: Comparación entre enfoques según métricas basadas en etiquetas

Para las curvas de *Recall*, por un lado, se observan resultados similares, con valores que si bien disminuyen ante el aumento de la dificultad, no lo hacen de manera abrupta. Destaca el hecho de que las curvas se intersectan alrededor de las 107 etiquetas, dando lugar a una jerarquía de métodos dependiente de la cantidad de etiquetas en estudio, siendo el método propuesto por el Enfoque Tradicional el mejor en desafíos que consideren hasta 107 etiquetas y tomando el Enfoque XMLC la delantera desde dicho punto en adelante, mas siempre a una diferencia de a lo más 0,0426 (caso de 54 etiquetas).

Distinto es el caso de la *Precision*, donde se observan pendientes mucho más pronunciadas y, si bien la intersección de curvas se mantiene, esta se da alrededor de las 131 etiquetas y en el orden inverso: en cantidades menores el método dominante es el del Enfoque XMLC y a partir de la intersección pasa a ser el Enfoque Tradicional. Destaca también que en este caso la diferencia entre ambas curvas es mucho más pronunciada, llegando incluso a una ventaja de 0,1245 puntos a favor del Enfoque Tradicional en el caso con 281 etiquetas.

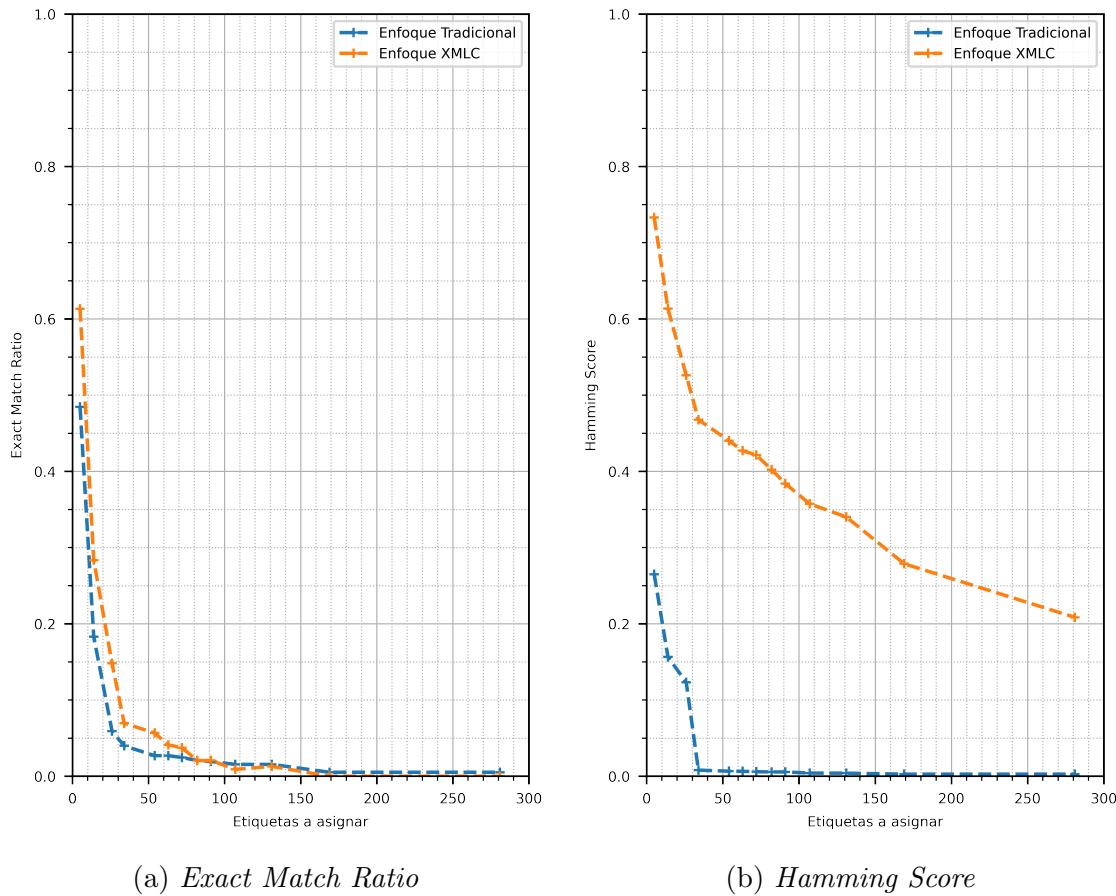
Lo anterior permitiría comprender el porqué del cambio en el *Micro F<sub>2</sub>-Score* en cantidades elevadas, en cuanto si bien el método del Enfoque XMLC se comportaría de manera

similar al Enfoque Tradicional (logrando incluso ventaja en ciertos puntos), sufriría una pérdida de *Precision* en los desafíos más complejos, lo cual el *F<sub>2</sub>-Score* estaría castigando.

De esta manera el método propuesto por el Enfoque Tradicional se plantearía como el mejor según las métricas basadas en etiquetas, con la razón de dicha jerarquía dada principalmente por una diferencia en *Precision*, métrica que si bien no responde a lo “más importante” según los supuestos realizados (pues se considera que para el usuario será más útil cubrir una cantidad alta de etiquetas correctas que el recibir solo etiquetas correctas con la posibilidad de omitir varias de ellas), juega un papel de todas maneras valioso (pues de lo contrario bastaría construir un sistema que optimice el *Recall* entregando siempre todas las etiquetas, lo cual no aportaría valor alguno).

### 8.1.2. Comparación según métricas basadas en ejemplos

Una vez realizado el análisis según métricas basadas en etiquetas, resta aún estudiar como dichos resultados se ven complementados por las métricas basadas en ejemplos. Para ello, en primera instancia se estudian *Exact Match Ratio* y *Hamming Score* (Figuras 8.3a y 8.3b).



(a) *Exact Match Ratio*

(b) *Hamming Score*

Figura 8.3: Comparación entre enfoques, métricas por ejemplo

En cuanto al *Exact Match Ratio*, los resultados coinciden con lo indicado en la presentación de las métricas (Sección 4.6): al tratarse de la fracción de patrones con todas sus etiquetas

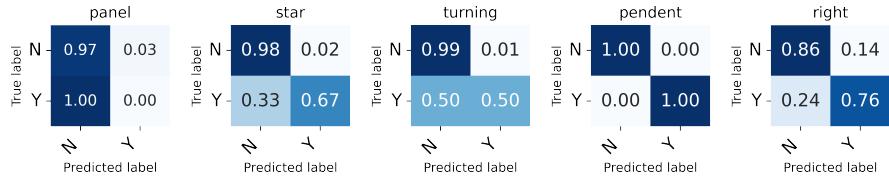
correctas y ninguna incorrecta, dicha métrica resulta demasiado estricta y en consecuencia no se adapta bien al problema, presentando para ambos enfoques curvas que decaen rápidamente a valores prácticamente nulos. En este sentido el resultado es poco informativo, sin embargo, su estudio permite validar la necesidad de utilizar otras métricas (tales como *Hamming Score* y *K-Match Ratio*).

Tabla 8.2: AUC *Hamming Score* mejor método de cada enfoque

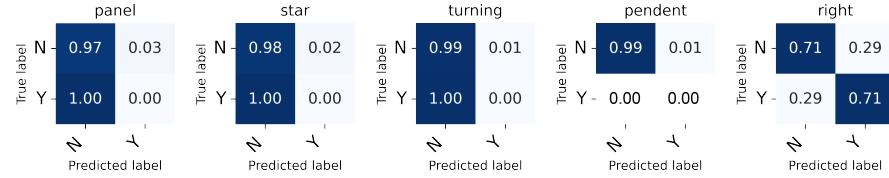
Enfoque	AUC
Tradicional	7,0063
XMLC	138,3504

En lo que respecta al *Hamming Score*, por otro lado, se observa una diferencia sustancial: mientras el método candidato del Enfoque XMLEC comienza en 0,7332 y decae hasta un mínimo de 0,2085 en 281 etiquetas, el candidato del Enfoque Tradicional comienza en 0,2649 y alcanza un valor de 0,0080 en el caso con 34 etiquetas, punto a partir del cual su puntaje se mantiene prácticamente nulo, dando lugar a un AUC a más de 100 puntos del obtenido por el Enfoque XMLEC (Véase Tabla 8.2). Esto podría resultar sorpresivo, en cuanto denota una diferencia importante entre lo visto por las métricas basadas en etiquetas - donde los métodos tenían comportamientos similares y el candidato Tradicional se posicionaba como el mejor - y lo arrojado por las métricas basadas en ejemplos - donde dicho candidato se vería ampliamente superado por el del Enfoque XMLEC -. Al respecto:

- Podría parecer razonable pensar que si el *Hamming Score* castiga el error por patrón tanto en el sentido de no entregar las etiquetas reales como en el de entregar etiquetas incorrectas, entonces su puntaje debería guardar cierta proporcionalidad con el *F<sub>2</sub>-Score*, en cuanto este último evalúa también ambos aspectos pero desde la perspectiva de las etiquetas.
- Sin embargo, dicha idea ignoraría un aspecto central de la forma en que se calculan las métricas basadas en etiquetas y que explica la diferencia observada con el *Hamming Score*. Se trata del cálculo de forma *Micro*, en el cual las etiquetas menos frecuentes reciben menos relevancia y por ende un fallo en ellas no afecta sustancialmente al resultado, *versus* el cálculo por ejemplo, en donde cada etiqueta recibe la misma importancia y, en particular para el *Hamming Score*, un falso positivo se castiga más que un falso negativo.
- En consecuencia, lo que el estudio del *Hamming Score* agregaría a la información entregada por el *Micro F<sub>2</sub>-Score* es que, mientras en etiquetas frecuentes el método propuesto por el Enfoque Tradicional estaría obteniendo los mejores resultados al predecir correctamente las etiquetas de frecuencia alta e intermedia (lo cual le posicionaría como el mejor según *F<sub>2</sub>*), existirían etiquetas poco frecuentes para las cuales este no estaría dando abasto, y, en cambio, el Enfoque XMLEC sí (y de ahí que este se posicione como el mejor en *Hamming Score*). Tal fenómeno se ilustra de buena forma a través del estudio de las matrices de confusión de etiquetas de alta y baja frecuencia, para lo cual las Figuras 8.4 y 8.5 presentan una selección de casos notables.

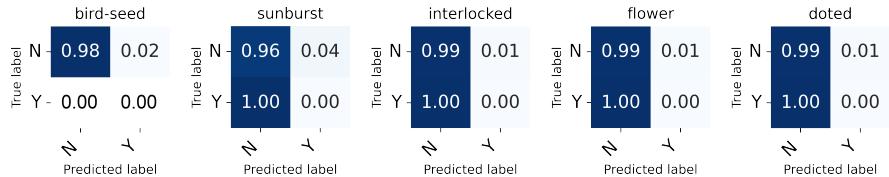


(a) *Enfoque Tradicional*

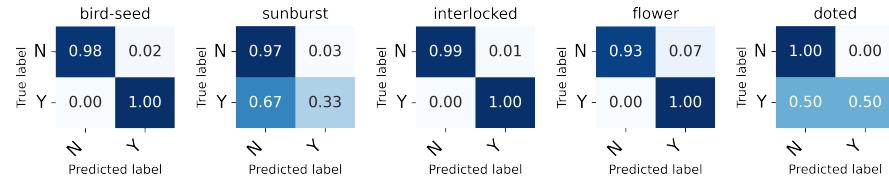


(b) *Enfoque XMLC*

Figura 8.4: Selección de matrices de confusión para etiquetas con al menos 15 eventos, según enfoque.



(a) *Enfoque Tradicional*



(b) *Enfoque XMLC*

Figura 8.5: Selección de matrices de confusión para etiquetas con a lo más 2 eventos, según enfoque.

Ahora bien, aun cuando el *Hamming Score* estaría permitiendo identificar ventajas del Enfoque XMLC que las métricas basadas en etiquetas no habrían recogido, aún quedaría pendiente la traducción de toda la información recolectada a una o más aseveraciones simples sobre el comportamiento del eventual sistema, a fin de formular una información más clara y directa para los eventuales usuarios (considerando que estos no tendrían por qué contar con un trasfondo que les permita comprender las métricas estudiadas hasta ahora). Para tal fin se estudia la métrica *K-Match Ratio* con  $K \in [1, 3, 5]$ , lo cual se presenta en las Figuras 8.6, 8.7a y 8.7b respectivamente.

Comenzando con el *1-Match Ratio*, los resultados serían sumamente positivos: los candidatos de ambos enfoques estarían logrando predecir al menos una etiqueta correcta para el 95 % de los patrones en prácticamente todo el espectro de etiquetas. Esto revelaría que el mínimo impuesto por el supuesto S2 se estaría cumpliendo, en cuanto este indicaría que el aportar al menos una única etiqueta correcta para cada patrón ya representaría valor agregado para el área.

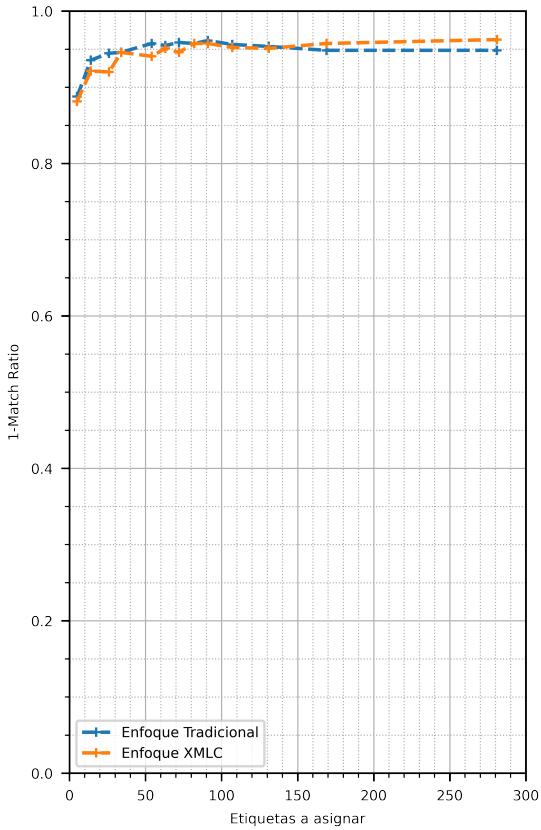


Figura 8.6: Comparación entre enfoques, 1-Match Ratio.

Ahora bien, al pasar al caso del *3-Match Ratio*, se observa que ambas curvas permanecen sobre 0,60 en todo el espectro de cantidades de etiquetas, con una leve ventaja en términos de AUC para el Enfoque XMLC (véase Tabla 8.3). Esto permitiría afirmar que, sea cual sea el método escogido y el tamaño del diccionario de etiquetas a utilizar, las recomendaciones resultantes incorporarían al menos 3 etiquetas correctas para por lo menos el 60% de los patrones, considerando desde luego patrones provenientes del libro o muy similares a estos (en particular, con un origen similar y con la misma taxonomía).

Lo anterior correspondería a garantías interesantes, sin embargo, estarían lejos aún del ideal, en cuanto previamente se observó que con el procesamiento de datos realizado el promedio de etiquetas por patrón habría ascendido a 5 en lugar de 3. Es ahí donde el *5-Match Ratio* cobra relevancia, aumentando la dificultad a predecir al menos 5 etiquetas correctas.

En este sentido, la Figura 8.7b permite observar que si bien el *5-Match Ratio* comienza con valores cercanos a 0,80 para ambos enfoques, decae rápidamente a medida que la cantidad de etiquetas aumenta. Esto resulta coherente con la formulación de la métrica, en cuanto para casos con pocas etiquetas existirían muchos patrones con menos de 5 etiquetas reales y para los cuales bastaría predecir dicho conjunto para ser considerado dentro de la suma (tratándose entonces de una forma vana de lograr un *K-Match Ratio* alto). Cuando la cantidad de etiquetas aumenta, por otro lado, los patrones comienzan a adquirir al menos 5 etiquetas reales y la métrica alcanza su complejidad (lo cual la Figura sugeriría ocurre alrededor de las 150 etiquetas).

Tabla 8.3: AUC  $K$ -Match Ratio mejor método de cada enfoque

$K$	Enfoque	AUC
1	Tradicional	392,9123
1	XMLC	393,4002
3	Tradicional	279,3995
3	XMLC	283,6788
5	Tradicional	153,8686
5	XMLC	167,2539

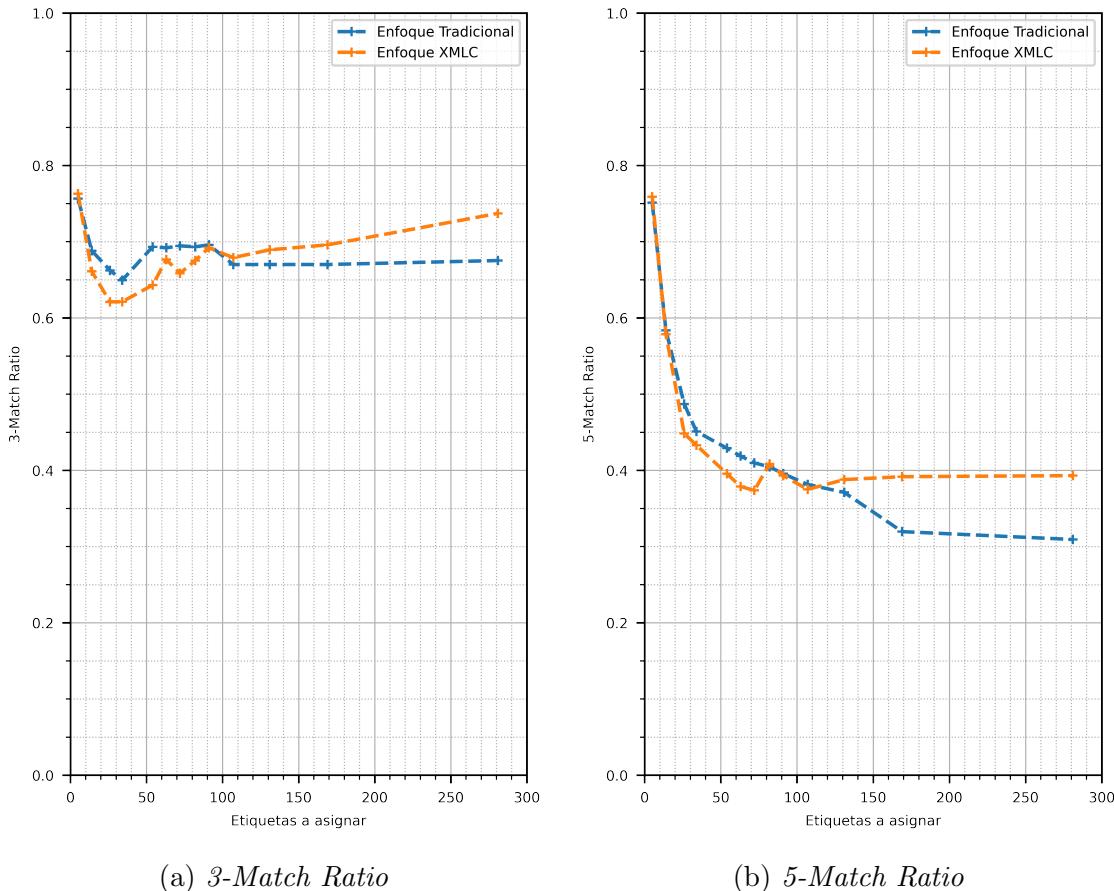


Figura 8.7: Comparación entre enfoques,  $K$ -Match Ratio para  $K \in [3, 5]$ .

Ahora bien, en lo que respecta a la construcción de afirmaciones, se observa que ambos métodos se mantienen por sobre 0,30 en todo el espectro, sin embargo, la curva del método perteneciente al Enfoque XMLC se mantiene mucho más próxima al 0,40 (y así se refleja también en los valores de AUC, disponibles en la Tabla 8.3). En consecuencia, de escoger trabajar con el método del Enfoque XMLC la afirmación resultante sería que, sin importar la cantidad de etiquetas con la que se desee trabajar, las recomendaciones incorporarían al menos 5 etiquetas correctas (o su totalidad, en caso de que no existan 5 etiquetas correctas) para el 37,88 % de los patrones. Para el método del Enfoque Tradicional, por otro lado, dicho

valor se vería reducido a un 30,93 %.

Desde luego, el estudio del *K-Match Ratio* no entrega información sobre la cantidad de etiquetas incorrectas que el sistema recomendaría, lo cual podría ser también una información interesante para los posibles usuarios. En este sentido, el desarrollo de una métrica que ofrezca tal información de forma directa pasa a formar parte del Trabajo Futuro (Sección 9.1).

## 8.2. Formulación de la herramienta de apoyo

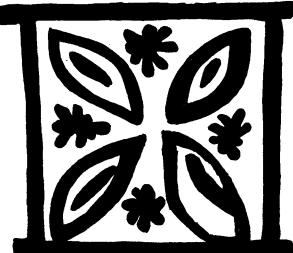
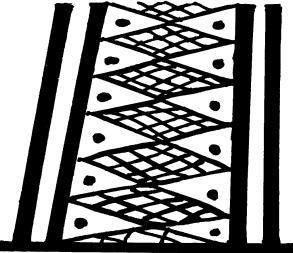
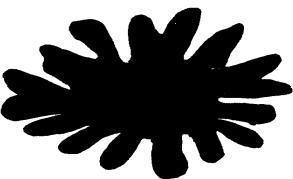
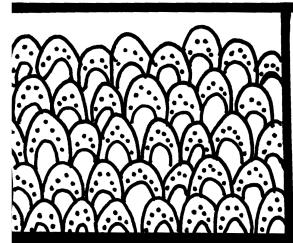
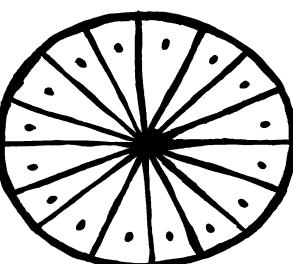
Una vez concluido el análisis de los mejores candidatos de cada enfoque según las distintas métricas de interés, sigue la elección de un método a proponer como herramienta de apoyo al etiquetado. Al respecto:

- Si bien el *Micro F<sub>2</sub>-Score* habría permitido jerarquizar de manera simple las distintas configuraciones al interior de un mismo método, el análisis complementario con métricas basadas en ejemplos habría demostrado que existirían fenómenos los cuales el *Micro F<sub>2</sub>-Score* no estaría rescatando, y que serían de todas maneras importantes.
- En vista de lo anterior, se opta por tomar una decisión considerando no solo dicha métrica, sino también los descubrimientos y afirmaciones resultantes de las métricas basadas en ejemplos.
- Aún más, la elección a realizar vendría fuertemente condicionada por el supuesto S3, que guardase relación con la relevancia de las etiquetas menos frecuentes (y por ende con la importancia de predecirlas correctamente).

Dado lo anterior, el método seleccionado para conformar la herramienta de apoyo al etiquetado es el propuesto por el Enfoque XMLC, a saber, el uso de AlexNet como *Threshold Dependent Neural Network* (TDNN) con un umbral de asignación 0,75, utilizando pesos para el balance de casos y trabajando con las técnicas de D.A. relativas al conjunto 13.

A continuación se presentan algunos ejemplos de patrones de prueba y las etiquetas que el sistema propuesto les recomendaría al utilizar un universo con 107 etiquetas posibles, junto a las etiquetas reales (Tabla 8.4).

Tabla 8.4: Ejemplos de sugerencias de la herramienta con un universo de las 107 etiquetas más frecuentes. **Observación:** las etiquetas subrayadas forman parte de las etiquetas reales mas no del universo en estudio. Los patrones presentados corresponden a la reproducción manual de bocetos realizados por Norbert Kunisch en su obra *Ornamente Geometrischen Vasen: Ein Kompendium* [28].

Ref.	Patrón	Etiquetas reales	Etiquetas predichas
16h (pág. 38)		line, simple, metopal, double, single, quatre- foil, outline	line, simple, metopal, hatched, double, single, quatrefoil, tree
40b (pág. 101)		panel, vertical, cross- hatched, dot, lozenge, chain	panel, vertical, hat- ched, small, multiple, outline, lozenge, chain, row, tongue
71f (pág. 184)		filling, ornament, star, <u>sixteen-pointed</u>	filling, ornament, star, swastika
74e (pág. 192)		horizontal, panel, pat- tern, double, dotted, <u>scale</u>	horizontal, panel, ver- tical, hatched, wavy, band, multiple, net, zigzag, scribble
77j (pág. 200)		enclosing, line, orna- ment, star, dot, circle, floor	enclosing, ornament, star, circle, octofoil, swastika, floor

# Capítulo 9

## Conclusiones

Habiendo sido alcanzado el objetivo principal de la investigación (a saber, la formulación de una herramienta de apoyo al etiquetado de patrones geométricos), resulta apropiado concluir la misma con una revisión de las preguntas que la habrían motivado (Sección 4.3) y reconocer así también el nivel de cumplimiento de los objetivos específicos. A continuación se aborda cada pregunta junto a las respuestas que la investigación habría otorgado.

La primera pregunta (PG1) guardaría relación con la posibilidad de formular una herramienta de apoyo al etiquetado mediante aprendizaje multietiqueta en virtud de los datos disponibles. En este sentido, la investigación habría demostrado que efectivamente tal formulación habría sido posible, y en particular lo habría sido tanto desde un Enfoque Tradicional como desde un Enfoque XMLC.

La segunda pregunta (PG2) se referiría a la cantidad máxima de etiquetas sobre la cual el sistema sería capaz de aportar valor, dada la cantidad reducida de ejemplos con los que se cuenta para cada una. En este sentido, se buscaron y encontraron métodos capaces de cumplir con los mínimos impuestos por el supuesto S2 (el cual dijese que entregar al menos una etiqueta correcta ya significaría valor agregado para el experto) al utilizar desde 5 hasta 281 etiquetas distintas, de las 330 disponibles en el conjunto de datos, llegando entonces hasta etiquetas con una única aparición en el conjunto de datos.

- Nótese que no se experimentó con 330 etiquetas por tratarse de un caso con umbral de frecuencia 1 (al igual que el caso con 281 etiquetas), lo cual hacía que para dejar el único ejemplo en todos los conjuntos de prueba resultase necesario dotar a estos conjuntos de demasiadas entradas (dejando poco material para entrenamiento). El caso con 281 etiquetas fue resultado de particiones aleatorias y se asume presenta un comportamiento similar al caso completo.

La tercera pregunta (PG3) tendría que ver con el uso de técnicas de *Data Augmentation* y como ello influiría en los resultados. Al respecto, los experimentos habrían puesto en evidencia los beneficios que incorporar ejemplos sintéticos en el entrenamiento entregaría, observándose mejoras tanto en los métodos competitivos del Enfoque XMLC como en los del Enfoque Tradicional.

- Aún más, en ambos casos los conjuntos con los mejores resultados habrían correspondido a aquellos aumentados con las técnicas de *reflexión*, *rotación*, *deformación elástica*

*y avg. blur*, además de *recorte* en el caso del Enfoque Tradicional.

Continuando con la cuarta pregunta (PG4), esta se relacionaría con el diagnóstico realizado en un inicio a través del *Label Density*, bajo el cual se identificaría un importante desbalance entre casos positivos y negativos para cada etiqueta, siendo entonces la pregunta el qué técnicas podrían utilizarse para lidiar con este problema y cómo ello influiría en los resultados.

- Al respecto, en la investigación se experimentó exclusivamente con la aplicación de pesos en el entrenamiento como herramienta para el tratamiento al desbalance, tanto en el Enfoque Tradicional (incorporando pesos en los clasificadores base) como en el Enfoque XMLC (incorporando pesos en las funciones de pérdida de los modelos). En ambos casos, su aplicación resultó esencial para alcanzar resultados competitivos.
- En particular, el método seleccionado como herramienta de apoyo habría alcanzado un *Micro F<sub>2</sub>-Score* de hasta tres veces su valor original al incorporar pesos en el entrenamiento.

Por último, la quinta pregunta (PG5) tendría relación con la confianza con la cual el usuario final podría interpretar las sugerencias entregadas por la herramienta, y si podría eventualmente delegar toda la responsabilidad del proceso en ella sin necesidad de realizar una revisión manual. Al respecto:

- Los resultados indican que un usuario que buscarse trabajar con los mismos patrones utilizados para evaluar los métodos podría contar con la garantía de que al trabajar con cualquier cantidad de etiquetas entre 5 y 281, el sistema recomendaría al menos 3 etiquetas correctas para al menos el 66,11 % de los patrones, y que similarmente el 37,88 % de ellos recibiría 5 etiquetas correctas.
- Ahora bien, el sistema incorpora también una tasa promedio de etiquetas incorrectamente recomendadas por patrón la cual aún no ha sido estudiada, y forma parte de los lineamientos para el Trabajo Futuro. Dicha tasa debería venir relacionada con la precisión calculada como métrica por etiqueta, lo cual desciende hasta un 20 % de las etiquetas asignadas siendo correctas en el caso con 281 etiquetas.
- A base de lo anterior, la respuesta a la última pregunta es directa: la herramienta propuesta no es aún capaz de realizar la tarea del etiquetado de patrones geométricos por sí sola, en cuanto requiere de un experto que sea capaz de identificar las sugerencias apropiadas y descartar el resto.

De esta manera, la investigación concluiría con éxito, en cuanto habría logrado dar respuesta a todas las preguntas que la habrían motivado (cubriendo así todos los objetivos específicos). Sin embargo, es necesario hacer énfasis en que los resultados obtenidos vienen desde la experimentación en un escenario sumamente protegido (patrones de un único libro, con un origen común y una misma taxonomía) y que no habría recibido aún retroalimentación de potenciales usuarios. Dichos aspectos se abordan a continuación, en la Sección destinada al Trabajo Futuro (9.1).

## 9.1. Trabajo Futuro

La investigación habría dado lugar a múltiples lineamientos para futuras iteraciones, tanto en el aspecto técnico como metodológico.

Desde un punto de vista técnico se habrían identificado distintas formas de tratar el problema, a saber:

- La posibilidad de abordar el desafío como un problema de MLC con múltiples instancias, es decir, que en lugar de asociar una imagen a múltiples etiquetas, sea la imagen descompuesta en distintos componentes para posteriormente asociar una etiqueta a cada componente. Este enfoque podría permitir aumentar la cantidad de entradas para cada etiqueta (al tener, por ejemplo, múltiples ejemplos de *circle* en un patrón con dicha etiqueta) pero requeriría de un trabajo adicional de construcción de un “lenguaje visual” bajo el cual se asocie una forma a cada etiqueta, lo cual necesitaría a su vez de conocimiento experto y podría además presentar dificultades para las etiquetas contextuales (es decir, aquellas que corresponderían a más de una instancia al mismo tiempo, o al patrón completo).
- El *deploy* de la herramienta obtenida en esta investigación y su posterior estudio como un caso de MLC online y con supervisión limitada, en cuanto el despliegue de la plataforma permitiría comenzar a recibir retroalimentación usuaria y con ello abrir la posibilidad a obtener una mejor taxonomía a partir de la extraída desde el conjunto de datos original.
- La posibilidad de continuar explorando los mejores métodos encontrados hasta el momento, con miras a concebir un modelo que, siendo entrenable de principio a fin, sea capaz de explotar las buenas propiedades que tanto el uso de *Binary Relevance* para etiquetas frecuentes como el de redes profundas para etiquetas específicas entregaría. En este sentido, se considera que un camino con mucho potencial es el de explorar el ajuste por etiqueta del umbral de asignación en la TDNN desarrollada.
- Existe también la opción de incorporar información adicional disponible en *Ornamente Geometrischen Vasen: Ein Kompendium* relativa a la ubicación en museos de los patrones. Esto podría permitir alcanzar mejores predicciones en el conjunto de datos utilizado, sin embargo, tal idea debe ser tratada con precaución, en cuanto podría dar lugar a resultados poco generalizables al requerir que, para cualquier patrón nuevo, se conozca su ubicación en museos (lo cual probablemente no será posible en la práctica, pues el etiquetado podría formar parte precisamente del proceso de musealización).
- Por último, una interesante línea de trabajo se encuentra en el comprender en mayor profundidad las razones por las cuales los métodos del Enfoque XMLC relacionados con la construcción de un *joint embedding* no habrían logrado posicionarse como métodos competitivos. Al respecto, se sugiere la hipótesis de que la cantidad de datos disponibles podría haber sido demasiado baja para el entrenamiento de métodos tan complejos, aún considerando las múltiples técnicas de *Data Augmentation* puestas en práctica.

Desde un punto de vista metodológico, por otro lado, los lineamientos tendrían que ver con la necesidad no cubierta de evaluar la generalización de la herramienta propuesta a conjuntos

de datos distintos al de la obra de Kunisch y cómo ello se traduciría en valor agregado para los posibles usuarios. Al respecto:

- Una primera tarea es la de conseguir fuentes de datos adicionales, encontrar la intersección entre las semánticas de etiquetado y estudiar qué tan bien se adapta la herramienta a estos nuevos datos. En este sentido, la exploración de nuevas fuentes debería también permitir evaluar la factibilidad de replicar el trabajo en objetos provenientes de culturas locales, obteniendo así una noción del impacto que la herramienta podría tener en la arqueología chilena y andina.
- De la mano con lo anterior, una línea de investigación se encuentra en el desarrollo de alguna medida que permita evaluar qué tan bien se adapta la taxonomía utilizada en la investigación a las taxonomías de otros conjuntos de datos.
- Otra importante tarea guarda relación con la generación de un proceso que, ante patrones etiquetados con semánticas similares pero cuyo formato diste ampliamente del utilizado para entrenar la herramienta (por ejemplo, imágenes a color o figuras 3D), realice la transformación a un formato más cercano.
- Otra necesidad urgente en este punto es la de recibir retroalimentación de usuarios reales sobre la utilidad actual de la herramienta, sobre la magnitud de los supuestos realizados y sobre qué tan dispuestos estarían a incorporar la herramienta de apoyo como parte de su trabajo cotidiano.
- Por último, sería de gran utilidad discutir con expertos acerca del *trade-off* que se genera entre el aumento de la cantidad de etiquetas a asignar (es decir, el tamaño del universo de etiquetas posibles) y la calidad de los resultados, en especial en cuanto a la intromisión de ruido en las recomendaciones se refiere. En este sentido, llegar a un acuerdo del estilo “basta predecir bien las 150 etiquetas más frecuentes” permitiría enfocar los esfuerzos en probar configuraciones especialmente diseñadas para tal caso, reduciendo también los costos computacionales asociados a la experimentación.

De esta manera, se concluye comentando que si bien la automatización total del proceso de etiquetado de patrones geométricos en objetos de herencia cultural se ve aún como una meta lejana, se considera que el trabajo realizado habría demostrado que tal objetivo no sería imposible, mas requeriría de la puesta en práctica de múltiples esfuerzos interdisciplinarios a fin de explorar las distintas líneas de investigación posibles.

# Bibliografía

- [1] Galen Andrew, Raman Arora, Jeff A. Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 1247–1255. JMLR.org, 2013.
- [2] F. Benites and E. Sapozhnikova. Haram: A hierarchical aram neural network for large-scale text classification. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 847–854, Nov 2015.
- [3] Flávia Cristina Bernardini, Rodrigo Barbosa da Silva, Rodrigo Magalhaes Rodovalho, and Edwin B. Mitacc Meza. Cardinality and density measures and their influence to multi-label learning methods. *Learning and Nonlinear Models*, 12:53–71, 2014.
- [4] Kush Bhatia, Himanshu Jain, Purushottam Kar, Manik Varma, and Prateek Jain. Sparse local embeddings for extreme multi-label classification. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 730–738, 2015.
- [5] Silvia Biasotti, Elia Moscoso Thompson, Loïc Barthe, Stefano Berretti, Andrea Giachetti, Thibault Lejemble, Nicolas Mellado, Konstantinos Moustakas, Iason Manolas, Dimitrios Dimou, Claudio Tortorici, Santiago Velasco-Forero, Naoufel Werghi, Martina Polig, Giusi Sorrentino, and Sorin Hermon. Recognition of geometric patterns over 3d models. In Alexandru C. Telea, Theoharis Theoharis, and Remco C. Veltkamp, editors, *11th Eurographics Workshop on 3D Object Retrieval, 3DOR@Eurographics 2018, Delft, The Netherlands, April 16, 2018*, pages 71–77. Eurographics Association, 2018.
- [6] Alberto Blanco, Arantza Casillas, Alicia Pérez, and Arantza Díaz de Ilarrazá. Multi-label clinical document classification: Impact of label-density. *Expert Syst. Appl.*, 138, 2019.
- [7] Feren Castillo Luján. Tipología y seriación de la cerámica proveniente del cementerio chimú de huaca de la luna, perú. *Boletín del Museo Chileno de Arte Precolombino*, 23:27 – 58, 00 2018.
- [8] Tianshui Chen, Muxin Xu, Xiaolu Hui, Hefeng Wu, and Liang Lin. Learning semantic-specific graph representation for multi-label image recognition. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 522–531. IEEE, 2019.

- [9] Wei-Jie Chen, Yuan-Hai Shao, Chun-Na Li, and Nai-Yang Deng. Mltsvm: a novel twin support vector machine to multi-label learning. *Pattern Recognition*, 52:61–74, 2016.
- [10] Weiwei Cheng and Eyke Hüllermeier. Combining instance-based learning and logistic regression for multilabel classification. *Mach. Learn.*, 76(2-3):211–225, 2009.
- [11] Hong-Min Chu, Kuan-Hao Huang, and Hsuan-Tien Lin. Dynamic principal projection for cost-sensitive online multi-label classification. *Mach. Learn.*, 108(8-9):1193–1230, 2019.
- [12] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*, pages 3008–3017. Computer Vision Foundation / IEEE, 2020.
- [13] Rachana Buch Dhatri Ganda. A survey on multi label classification. *Recent Trends in Programming languages*, 5(1):19–23, 2018.
- [14] S. Diplaris, G. Tsoumakas, P. Mitkas, and I. Vlahavas. Protein classification with multiple algorithms. In *Proc. 10th Panhellenic Conference on Informatics, Volos, Greece, PCI05*, pages 448–456, 2005.
- [15] Meng Joo Er, Rajasekar Venkatesan, and Ning Wang. An online universal classifier for binary, multi-class and multi-label classification. In *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016, Budapest, Hungary, October 9-12, 2016*, pages 3701–3706. IEEE, 2016.
- [16] Johannes Fürnkranz, Eyke Hüllermeier, Eneldo Loza Mencía, and Klaus Brinker. Multilabel classification via calibrated label ranking. *Mach. Learn.*, 73(2):133–153, 2008.
- [17] Nadia Ghamrawi and Andrew McCallum. Collective multi-label classification. In Otthein Herzog, Hans-Jörg Schek, Norbert Fuhr, Abdur Chowdhury, and Wilfried Teijken, editors, *Proceedings of the 2005 ACM CIKM International Conference on Information and Knowledge Management, Bremen, Germany, October 31 - November 5, 2005*, pages 195–200. ACM, 2005.
- [18] Shantanu Godbole and Sunita Sarawagi. Discriminative methods for multi-labeled classification. In Honghua Dai, Ramakrishnan Srikant, and Chengqi Zhang, editors, *Advances in Knowledge Discovery and Data Mining, 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26-28, 2004, Proceedings*, volume 3056 of *Lecture Notes in Computer Science*, pages 22–30. Springer, 2004.
- [19] Xiuwen Gong, Dong Yuan, and Wei Bao. Online metric learning for multi-label classification. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 4012–4019. AAAI Press, 2020.

- [20] Xiuwen Gong, Dong Yuan, and Wei Bao. Online metric learning for multi-label classification. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 4012–4019. AAAI Press, 2020.
- [21] Paula Gonzalez. *Arte y Cultura Diaguita Chilena: Simetría, simbolismo e identidad*. Ucayali Editores, Santiago, Chile, 2013.
- [22] Yufei Han, Guolei Sun, Yun Shen, and Xiangliang Zhang. Multi-label learning with highly incomplete data via collaborative embedding. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1494–1503. ACM, 2018.
- [23] Huihui He and Rui Xia. Joint binary neural network for multi-label learning with applications to emotion classification. In Min Zhang, Vincent Ng, Dongyan Zhao, Sujian Li, and Hongying Zan, editors, *Natural Language Processing and Chinese Computing - 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26-30, 2018, Proceedings, Part I*, volume 11108 of *Lecture Notes in Computer Science*, pages 250–259. Springer, 2018.
- [24] Yi Huang, Buse Giledereli, Abdullatif Köksal, Arzucan Özgür, and Elif Ozkirimli. Balancing methods for multi-label text classification with long-tailed class distribution. *CoRR*, abs/2109.04712, 2021.
- [25] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.
- [26] Alexander B. Jung. imgaug. <https://github.com/aleju/imgaug>, 2018. [Online; accessed 30-Oct-2018].
- [27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [28] Norbert Kunisch. *Ornamente Geometrischer Vasen: Ein Kompendium*. Böhlau Verlag, Cologne-Weimar, 1998.
- [29] Jieun Lee, Ji Hyun Yi, and Seungjun Kim. Cultural heritage design element labeling system with gamification. *IEEE Access*, 8:127700–127708, 2020.
- [30] Shen X. Liu W., Wang H. and Tsang I. The Emerging Trends of Multi-Label Learning. To appear in IEEE Transactions on Pattern Analysis & Machine Intelligence.
- [31] William Logan. *Closing Pandora’s Box: Human Rights Conundrums in Cultural Heritage Protection*, pages 33–52. Springer Science+Business Media, 01 2007.

- [32] Edward Loper and Steven Bird. Nltk: The natural language toolkit. *CoRR*, cs.CL/0205028, 2002.
- [33] Mathworks. Multilabel image classification using Deep Learning (r2022b), 2022. Retrieved October 25, 2022 from [www.mathworks.com/help/deeplearning/ug/multilabel-image-classification-using-deep-learning.html](http://www.mathworks.com/help/deeplearning/ug/multilabel-image-classification-using-deep-learning.html).
- [34] NHCN. Heritage objects: The national heritage council of namibia, 2005.
- [35] la Ciencia y la Cultura Organización de las Naciones Unidas para la Educación. Co-operation and coordination between the unesco conventions concerning heritage: the yamato declaration on integrated approaches for safeguarding tangible and intangible cultural heritage. In *World Heritage Committee, 7th extraordinary session, Paris, 2004.*, 2004.
- [36] la Ciencia y la Cultura Organización de las Naciones Unidas para la Educación. Text of the convention for the safeguarding of the intangible cultural heritage, Jan 2006.
- [37] Sunho Park and Seungjin Choi. Online multi-label learning with accelerated nonsmooth stochastic gradient descent. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, pages 3322–3326. IEEE, 2013.
- [38] Yashoteja Prabhu and Manik Varma. Fastxml: a fast, accurate and stable tree-classifier for extreme multi-label learning. In Sofus A. Macskassy, Claudia Perlich, Jure Leskovec, Wei Wang, and Rayid Ghani, editors, *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 263–272. ACM, 2014.
- [39] Jesse Read, Bernhard Pfahringer, Geoffrey Holmes, and Eibe Frank. Classifier chains for multi-label classification. In Wray L. Buntine, Marko Grobelnik, Dunja Mladenic, and John Shawe-Taylor, editors, *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II*, volume 5782 of *Lecture Notes in Computer Science*, pages 254–269. Springer, 2009.
- [40] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA.
- [41] Si Si, Huan Zhang, S. Sathiya Keerthi, Dhruv Mahajan, Inderjit S. Dhillon, and Cho-Jui Hsieh. Gradient boosted decision trees for high dimensional sparse output. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 3182–3190. PMLR, 2017.
- [42] Ivan Sipiran, Patrick Lazo, Cristian Lopez, Milagritos Jimenez, Nihar Bagewadi, Benjamin Bustos, Hieu Dao, Shankar Gangisetty, Martin Hanik, Ngoc-Phuong Ho-Thi, Mike Holenderski, Dmitri Jarnikov, Arniel Labrada, Stefan Lengauer, Roxane Licandro, Dinh-Huan Nguyen, Thang-Long Nguyen-Ho, Luis A. Pérez Rey, Bang-Dang Pham, Reinhold

Preiner, Tobias Schreck, Quoc-Huy Trinh, Loek Tonnaer, Christoph von Tycowicz, and The-Anh Vu-Le. SHREC 2021: Retrieval of cultural heritage objects. *Comput. Graph.*, 100:1–20, 2021.

- [43] Yu-Yin Sun, Yin Zhang, and Zhi-Hua Zhou. Multi-label learning with weak label. In Maria Fox and David Poole, editors, *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. AAAI Press, 2010.
- [44] P. Szymanski and T. Kajdanowicz. A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*, February 2017.
- [45] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *Int. J. Data Warehous. Min.*, 3(3):1–13, 2007.
- [46] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. I.: A review of multi-label classification methods. In *In: Proceedings of the 2nd ADBIS Workshop on Data Mining and Knowledge Discovery (ADMKD)*, pages 99–109, 2006.
- [47] Grigorios Tsoumakas and Ioannis P. Vlahavas. Random  $k$ -labelsets: An ensemble method for multilabel classification. In Joost N. Kok, Jacek Koronacki, Ramón López de Mántaras, Stan Matwin, Dunja Mladenic, and Andrzej Skowron, editors, *Machine Learning: ECML 2007, 18th European Conference on Machine Learning, Warsaw, Poland, September 17-21, 2007, Proceedings*, volume 4701 of *Lecture Notes in Computer Science*, pages 406–417. Springer, 2007.
- [48] Rajasekar Venkatesan, Meng Joo Er, Mihika Dave, Mahardhika Pratama, and Shiqian Wu. A novel online multi-label classifier for high-speed streaming data applications. *Evol. Syst.*, 8(4):303–315, 2017.
- [49] Haobo Wang, Weiwei Liu, Yang Zhao, Chen Zhang, Tianlei Hu, and Gang Chen. Discriminative and correlative partial multi-label learning. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 3691–3697. ijcai.org, 2019.
- [50] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. CNN-RNN: A unified framework for multi-label image classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2285–2294. IEEE Computer Society, 2016.
- [51] Kaixiang Wang. Robust embedding framework with dynamic hypergraph fusion for multi-label classification. In *IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, July 8-12, 2019*, pages 982–987. IEEE, 2019.
- [52] Kaixiang Wang, Ming Yang, Wanqi Yang, and Yilong Yin. Deep correlation structure preserved label space embedding for multi-label classification. In Jun Zhu and Ichiro Takeuchi, editors, *Proceedings of The 10th Asian Conference on Machine Learning, ACML 2018, Beijing, China, November 14-16, 2018*, volume 95 of *Proceedings of Machine Learning Research*, pages 1–16. PMLR, 2018.

- [53] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3):63:1–63:34, 2020.
- [54] Baoyuan Wu, Fan Jia, Wei Liu, Bernard Ghanem, and Siwei Lyu. Multi-label learning with missing labels using mixed dependency graphs. *Int. J. Comput. Vis.*, 126(8):875–896, 2018.
- [55] Zhaodong Wu, Jun Zhang, and Shengliang Hu. Review on classification algorithm and evaluation system of machine learning. In *2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pages 214–218, 2020.
- [56] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning with noisy label identification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(7):3676–3687, 2022.
- [57] Eleftherios Spyromitros Xioufis, Grigoris Tsoumakas, and Ioannis P. Vlahavas. An empirical study of lazy multilabel classification algorithms. In John Darzentas, George A. Vouros, Spyros Vosinakis, and Argyris Arnellos, editors, *Artificial Intelligence: Theories, Models and Applications, 5th Hellenic Conference on AI, SETN 2008, Syros, Greece, October 2-4, 2008. Proceedings*, volume 5138 of *Lecture Notes in Computer Science*, pages 401–406. Springer, 2008.
- [58] Miao Xu, Rong Jin, and Zhi-Hua Zhou. Speedup matrix completion with side information: Application to multi-label learning. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2301–2309, 2013.
- [59] Miao Xu, Gang Niu, Bo Han, Ivor W. Tsang, Zhi-Hua Zhou, and Masashi Sugiyama. Matrix co-completion for multi-label classification with missing features and labels. *CoRR*, abs/1805.09156, 2018.
- [60] Hao Yang, Joey Tianyi Zhou, Jianfei Cai, and Yew-Soon Ong. MIML-FCN+: multi-instance multi-label learning via fully convolutional networks with privileged information. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5996–6004. IEEE Computer Society, 2017.
- [61] Menglin Yang, Min Zhou, Zhihao Li, Jiahong Liu, Lujia Pan, Hui Xiong, and Irwin King. Hyperbolic graph neural networks: A review of methods and applications. *CoRR*, abs/2202.13852, 2022.
- [62] Chih-Kuan Yeh. C2AE - Implementación en Matlab, 2018.
- [63] Chih-Kuan Yeh, Wei-Chieh Wu, Wei-Jen Ko, and Yu-Chiang Frank Wang. Learning deep latent space for multi-label classification. In Satinder Singh and Shaul Markovitch, editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 2838–2844. AAAI Press, 2017.

- [64] Ian En-Hsu Yen, Xiangru Huang, Pradeep Ravikumar, Kai Zhong, and Inderjit S. Dhillon. Pd-sparse : A primal and dual sparse approach to extreme multiclass and multilabel classification. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 3069–3077. JMLR.org, 2016.
- [65] Hsiang-Fu Yu, Prateek Jain, Purushottam Kar, and Inderjit S. Dhillon. Large-scale multi-label learning with missing labels. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 593–601. JMLR.org, 2014.
- [66] Min-Ling Zhang and Jun-Peng Fang. Partial multi-label learning via credible label elicitation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(10):3587–3599, 2021.
- [67] Min-Ling Zhang, Yu-Kun Li, Xu-Ying Liu, and Xin Geng. Binary relevance for multi-label learning: an overview. *Frontiers Comput. Sci.*, 12(2):191–202, 2018.
- [68] Min-Ling Zhang and Kun Zhang. Multi-label learning by exploiting label dependency. In Bharat Rao, Balaji Krishnapuram, Andrew Tomkins, and Qiang Yang, editors, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 999–1008. ACM, 2010.
- [69] Min-Ling Zhang and Zhi-Hua Zhou. Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Trans. Knowl. Data Eng.*, 18(10):1338–1351, 2006.
- [70] Min-Ling Zhang and Zhi-Hua Zhou. Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Trans. Knowl. Data Eng.*, 18(10):1338–1351, 2006.
- [71] Min-Ling Zhang and Zhi-Hua Zhou. Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Trans. Knowl. Data Eng.*, 18(10):1338–1351, 2006.
- [72] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.*, 26(8):1819–1837, 2014.
- [73] Yue Zhu, James T. Kwok, and Zhi-Hua Zhou. Multi-label learning with global and local label correlation. *IEEE Trans. Knowl. Data Eng.*, 30(6):1081–1094, 2018.
- [74] G.K. Zipf. Human behavior and the principle of least effort. *Addison-Wesley*, 1949.
- [75] znhub. C2AE - Implementación en Tensorflow 1.0, 2017.

# ANEXOS

## Anexo A

### Experimento 1

Tabla A.1: Métricas Experimento 1 - Sin tratamiento al desbalance.

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR LR	5	0,2789	0,7282	0,7143	0,7899	0,5090	0,6508
BR SVC	5	0,2726	0,5158	0,4806	0,7332	0,3389	0,4368
BR DT	5	0,2249	0,6295	0,6390	0,5954	0,2668	0,5657
BR GNB	5	0,1917	0,5600	0,6432	0,3693	0,0851	0,6224
LP LR	5	0,3007	0,7146	0,7027	0,7670	0,5464	0,6766
LP SVC	5	0,2659	0,5458	0,5195	0,6854	0,4098	0,5245
LP DT	5	0,2352	0,5937	0,5909	0,6091	0,3814	0,5954
LP GNB	5	0,1932	0,5651	0,6152	0,4271	0,1585	0,5464
CC LR	5	0,2948	0,7272	0,7109	0,8009	0,5567	0,6740
CC SVC	5	0,2684	0,5010	0,4642	0,7359	0,3466	0,4407
CC DT	5	0,2388	0,6079	0,6123	0,5939	0,3273	0,5876
CC GNB	5	0,1917	0,5564	0,6366	0,3703	0,0889	0,6224
RakelD LR	5	0,2947	0,7353	0,7224	0,7920	0,5451	0,6727
RakelD SVC	5	0,2508	0,5516	0,5269	0,6858	0,3441	0,4845
RakelD DT	5	0,2460	0,6022	0,6013	0,6066	0,3183	0,5541
RakelD GNB	5	0,1947	0,5694	0,6416	0,3947	0,1057	0,5683
BRkNN	5	0,2537	0,6200	0,6133	0,6484	0,4059	0,6031

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLkNN	5	0,2583	0,6167	0,6087	0,6558	0,3905	0,5760
MLTSVM	5	0,3901	0,4203	0,3862	0,6551	0,2668	0,3801
BR LR	14	0,1655	0,6337	0,6079	0,7638	0,2500	0,3724
BR SVC	14	0,1603	0,4272	0,3863	0,7452	0,1688	0,2114
BR DT	14	0,1558	0,5516	0,5606	0,5193	0,0773	0,3080
BR GNB	14	0,1480	0,5503	0,7207	0,2833	0,0245	0,5129
LP LR	14	0,1652	0,6444	0,6380	0,6722	0,2874	0,4304
LP SVC	14	0,1529	0,5037	0,4914	0,5597	0,1868	0,2861
LP DT	14	0,1510	0,5116	0,5105	0,5165	0,1546	0,3157
LP GNB	14	0,1490	0,5259	0,5305	0,5085	0,1366	0,3312
CC LR	14	0,1625	0,6458	0,6230	0,7576	0,2887	0,4033
CC SVC	14	0,1562	0,4172	0,3761	0,7458	0,1752	0,2139
CC DT	14	0,1430	0,5559	0,5634	0,5284	0,1224	0,3505
CC GNB	14	0,1480	0,5353	0,6755	0,2930	0,0245	0,4214
RakelD LR	14	0,1594	0,6387	0,6181	0,7388	0,2307	0,3789
RakelD SVC	14	0,1550	0,4052	0,3655	0,7244	0,1624	0,2062
RakelD DT	14	0,1504	0,5248	0,5352	0,4873	0,0825	0,2938
RakelD GNB	14	0,1480	0,5489	0,6641	0,3251	0,0180	0,4214
BRkNN	14	0,1590	0,5518	0,5509	0,5555	0,1894	0,3531
MLkNN	14	0,1560	0,5245	0,5050	0,6214	0,1727	0,2951
MLTSVM	14	0,2494	0,3553	0,3440	0,4097	0,0528	0,1830
BR LR	26	0,1283	0,5457	0,5119	0,7421	0,1379	0,2191
BR SVC	26	0,1177	0,3530	0,3120	0,7483	0,0773	0,0954
BR DT	26	0,1235	0,4907	0,5026	0,4489	0,0348	0,1868
BR GNB	26	0,1222	0,4788	0,7265	0,2032	0,0000	0,4678
LP LR	26	0,1284	0,5596	0,5481	0,6108	0,1778	0,2822
LP SVC	26	0,1212	0,4016	0,3812	0,5116	0,1121	0,1572
LP DT	26	0,1258	0,4377	0,4359	0,4452	0,0928	0,1907
LP GNB	26	0,1209	0,4318	0,4187	0,4939	0,0966	0,1959
CC LR	26	0,1253	0,5615	0,5302	0,7355	0,1649	0,2410
CC SVC	26	0,1148	0,3453	0,3044	0,7497	0,0850	0,0992
CC DT	26	0,1190	0,4829	0,4922	0,4500	0,0425	0,2139
CC GNB	26	0,1222	0,4512	0,5486	0,2643	0,0039	0,2565
RakelD LR	26	0,1270	0,5708	0,5442	0,7106	0,1327	0,2358
RakelD SVC	26	0,1177	0,3518	0,3114	0,7350	0,0760	0,0979

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
RakelD DT	26	0,1230	0,4732	0,4815	0,4429	0,0219	0,1714
RakelD GNB	26	0,1222	0,4732	0,6283	0,2387	0,0013	0,2861
BRkNN	26	0,1266	0,4947	0,4932	0,5006	0,1083	0,2397
MLkNN	26	0,1248	0,4524	0,4271	0,5940	0,0967	0,1662
MLTSVM	26	0,1772	0,3212	0,3348	0,2772	0,0065	0,1134
BR LR	34	0,0170	0,5122	0,4760	0,7368	0,1031	0,1573
BR SVC	34	0,0183	0,3263	0,2862	0,7483	0,0593	0,0735
BR DT	34	0,0093	0,4644	0,4750	0,4268	0,0219	0,1521
BR GNB	34	0,0067	0,4475	0,7293	0,1760	0,0000	0,4588
LP LR	34	0,0089	0,5201	0,5115	0,5575	0,1327	0,2384
LP SVC	34	0,0119	0,3472	0,3193	0,5388	0,0773	0,1096
LP DT	34	0,0067	0,4023	0,4027	0,4011	0,0618	0,1572
LP GNB	34	0,0089	0,3821	0,3645	0,4740	0,0683	0,1379
CC LR	34	0,0128	0,5283	0,4935	0,7369	0,1186	0,1753
CC SVC	34	0,0205	0,3188	0,2789	0,7510	0,0657	0,0760
CC DT	34	0,0080	0,4679	0,4775	0,4338	0,0348	0,1791
CC GNB	34	0,0067	0,4165	0,4919	0,2582	0,0000	0,1688
RakelD LR	34	0,0145	0,5286	0,4994	0,6899	0,0851	0,1688
RakelD SVC	34	0,0209	0,3296	0,2906	0,7164	0,0592	0,0760
RakelD DT	34	0,0093	0,4420	0,4496	0,4142	0,0245	0,1379
RakelD GNB	34	0,0067	0,4552	0,6413	0,2109	0,0000	0,2938
BRkNN	34	0,0119	0,4712	0,4700	0,4762	0,0941	0,2049
MLkNN	34	0,0149	0,4223	0,3956	0,5814	0,0748	0,1276
MLTSVM	34	0,1669	0,3199	0,3548	0,2302	0,0078	0,0966
BR LR	54	0,0170	0,4668	0,4286	0,7268	0,0761	0,1237
BR SVC	54	0,0182	0,2906	0,2523	0,7483	0,0451	0,0567
BR DT	54	0,0079	0,4277	0,4412	0,3814	0,0154	0,1314
BR GNB	54	0,0066	0,3928	0,7335	0,1375	0,0000	0,4600
LP LR	54	0,0127	0,4680	0,4599	0,5030	0,1096	0,2023
LP SVC	54	0,0168	0,2249	0,2012	0,4364	0,0605	0,0915
LP DT	54	0,0079	0,3672	0,3645	0,3788	0,0502	0,1301
LP GNB	54	0,0120	0,3252	0,3034	0,4572	0,0606	0,1044
CC LR	54	0,0126	0,4852	0,4481	0,7275	0,0889	0,1379
CC SVC	54	0,0204	0,2838	0,2458	0,7516	0,0490	0,0567
CC DT	54	0,0066	0,4178	0,4314	0,3718	0,0155	0,1443

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
CC GNB	54	0,0066	0,3694	0,4242	0,2437	0,0000	0,1160
RakelD LR	54	0,0156	0,4842	0,4508	0,6900	0,0606	0,1301
RakelD SVC	54	0,0195	0,2971	0,2584	0,7446	0,0477	0,0631
RakelD DT	54	0,0079	0,4065	0,4149	0,3766	0,0090	0,1082
RakelD GNB	54	0,0066	0,4071	0,6250	0,1702	0,0000	0,2603
BRkNN	54	0,0118	0,4364	0,4351	0,4416	0,0760	0,1727
MLkNN	54	0,0135	0,3823	0,3538	0,5664	0,0541	0,0979
MLTSVM	54	0,1314	0,2884	0,3512	0,1684	0,0026	0,0812
BR LR	63	0,0141	0,4562	0,4176	0,7252	0,0670	0,1070
BR SVC	63	0,0115	0,2828	0,2448	0,7483	0,0373	0,0463
BR DT	63	0,0090	0,4116	0,4243	0,3682	0,0103	0,1134
BR GNB	63	0,0064	0,3766	0,7278	0,1287	0,0000	0,4459
LP LR	63	0,0085	0,4575	0,4522	0,4799	0,1031	0,1894
LP SVC	63	0,0077	0,2469	0,2285	0,3707	0,0593	0,0928
LP DT	63	0,0064	0,3521	0,3508	0,3578	0,0489	0,1160
LP GNB	63	0,0081	0,3146	0,2941	0,4362	0,0593	0,0953
CC LR	63	0,0098	0,4747	0,4371	0,7250	0,0786	0,1211
CC SVC	63	0,0137	0,2762	0,2386	0,7516	0,0412	0,0463
CC DT	63	0,0064	0,4106	0,4219	0,3718	0,0142	0,1315
CC GNB	63	0,0064	0,3593	0,4094	0,2412	0,0000	0,0979
RakelD LR	63	0,0115	0,4752	0,4410	0,6910	0,0606	0,1173
RakelD SVC	63	0,0111	0,2777	0,2404	0,7466	0,0386	0,0515
RakelD DT	63	0,0064	0,3976	0,4076	0,3619	0,0129	0,0992
RakelD GNB	63	0,0064	0,3724	0,5910	0,1504	0,0000	0,2088
BRkNN	63	0,0090	0,4266	0,4254	0,4318	0,0734	0,1611
MLkNN	63	0,0094	0,3724	0,3435	0,5632	0,0490	0,0863
MLTSVM	63	0,1206	0,2766	0,3490	0,1515	0,0013	0,0760
BR LR	72	0,0136	0,4488	0,4100	0,7236	0,0618	0,1018
BR SVC	72	0,0111	0,2758	0,2383	0,7483	0,0373	0,0463
BR DT	72	0,0072	0,4066	0,4199	0,3610	0,0116	0,1018
BR GNB	72	0,0059	0,3622	0,7228	0,1210	0,0000	0,4420
LP LR	72	0,0106	0,4550	0,4504	0,4743	0,1018	0,1778
LP SVC	72	0,0072	0,2337	0,2154	0,3614	0,0593	0,0902
LP DT	72	0,0085	0,3488	0,3478	0,3532	0,0503	0,1185

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
LP GNB	72	0,0076	0,2994	0,2796	0,4194	0,0580	0,0928
CC LR	72	0,0080	0,4655	0,4277	0,7210	0,0722	0,1186
CC SVC	72	0,0132	0,2693	0,2322	0,7522	0,0412	0,0463
CC DT	72	0,0059	0,4019	0,4129	0,3641	0,0168	0,1237
CC GNB	72	0,0059	0,3525	0,3990	0,2404	0,0000	0,0902
RakelD LR	72	0,0124	0,4717	0,4371	0,6916	0,0618	0,1263
RakelD SVC	72	0,0111	0,2789	0,2415	0,7432	0,0348	0,0476
RakelD DT	72	0,0059	0,3837	0,3918	0,3545	0,0103	0,0889
RakelD GNB	72	0,0059	0,3797	0,6232	0,1485	0,0000	0,2564
BRkNN	72	0,0085	0,4198	0,4187	0,4244	0,0683	0,1546
MLkNN	72	0,0089	0,3652	0,3360	0,5626	0,0425	0,0799
MLTSVM	72	0,1126	0,2666	0,3475	0,1384	0,0013	0,0734
BR LR	82	0,0134	0,4421	0,4032	0,7214	0,0580	0,1005
BR SVC	82	0,0108	0,2692	0,2322	0,7483	0,0361	0,0438
BR DT	82	0,0069	0,3916	0,4055	0,3445	0,0116	0,1005
BR GNB	82	0,0056	0,3518	0,7189	0,1158	0,0000	0,4459
LP LR	82	0,0078	0,4484	0,4446	0,4646	0,0992	0,1804
LP SVC	82	0,0069	0,2104	0,1937	0,3245	0,0580	0,0786
LP DT	82	0,0056	0,3328	0,3314	0,3385	0,0464	0,1018
LP GNB	82	0,0078	0,2820	0,2622	0,4043	0,0567	0,0799
CC LR	82	0,0078	0,4578	0,4200	0,7168	0,0618	0,1057
CC SVC	82	0,0129	0,2629	0,2262	0,7522	0,0399	0,0438
CC DT	82	0,0056	0,3945	0,4079	0,3497	0,0090	0,1276
CC GNB	82	0,0056	0,3411	0,3803	0,2418	0,0000	0,0773
RakelD LR	82	0,0121	0,4685	0,4344	0,6842	0,0554	0,1070
RakelD SVC	82	0,0099	0,2692	0,2324	0,7441	0,0373	0,0464
RakelD DT	82	0,0056	0,3938	0,4036	0,3594	0,0103	0,0915
RakelD GNB	82	0,0056	0,3554	0,5983	0,1358	0,0000	0,2229
BRkNN	82	0,0082	0,4146	0,4137	0,4186	0,0683	0,1533
MLkNN	82	0,0073	0,3585	0,3290	0,5599	0,0387	0,0760
MLTSVM	82	0,1078	0,2596	0,3459	0,1302	0,0013	0,0709
BR LR	91	0,0121	0,4356	0,3964	0,7210	0,0554	0,0953
BR SVC	91	0,0108	0,2646	0,2280	0,7483	0,0335	0,0399
BR DT	91	0,0056	0,3934	0,4066	0,3486	0,0090	0,0928

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR GNB	91	0,0056	0,3423	0,7141	0,1112	0,0000	0,4356
LP LR	91	0,0082	0,4452	0,4427	0,4556	0,0928	0,1791
LP SVC	91	0,0056	0,1956	0,1802	0,3025	0,0567	0,0786
LP DT	91	0,0069	0,3305	0,3324	0,3234	0,0477	0,1095
LP GNB	91	0,0065	0,2707	0,2531	0,3751	0,0567	0,0799
CC LR	91	0,0078	0,4530	0,4146	0,7200	0,0645	0,1031
CC SVC	91	0,0116	0,2584	0,2221	0,7522	0,0360	0,0399
CC DT	91	0,0056	0,3897	0,4030	0,3450	0,0065	0,1134
CC GNB	91	0,0056	0,3356	0,3726	0,2403	0,0000	0,0683
RakelD LR	91	0,0108	0,4561	0,4215	0,6797	0,0567	0,1005
RakelD SVC	91	0,0108	0,2616	0,2256	0,7375	0,0335	0,0399
RakelD DT	91	0,0069	0,3859	0,3944	0,3555	0,0052	0,0902
RakelD GNB	91	0,0056	0,3432	0,5902	0,1284	0,0000	0,2410
BRkNN	91	0,0082	0,4096	0,4090	0,4125	0,0683	0,1469
MLkNN	91	0,0073	0,3531	0,3234	0,5597	0,0361	0,0709
MLTSVM	91	0,1034	0,2524	0,3440	0,1226	0,0013	0,0683
BR LR	107	0,0067	0,4196	0,3796	0,7240	0,0412	0,0774
BR SVC	107	0,0041	0,2558	0,2196	0,7578	0,0155	0,0207
BR DT	107	0,0056	0,3873	0,4023	0,3374	0,0065	0,0799
BR GNB	107	0,0056	0,3330	0,7041	0,1073	0,0000	0,4111
LP LR	107	0,0056	0,4276	0,4250	0,4387	0,0928	0,1559
LP SVC	107	0,0056	0,2299	0,2229	0,2656	0,0490	0,0670
LP DT	107	0,0056	0,3210	0,3209	0,3219	0,0438	0,0915
LP GNB	107	0,0056	0,2474	0,2412	0,2804	0,0464	0,0631
CC LR	107	0,0032	0,4326	0,3937	0,7163	0,0541	0,0876
CC SVC	107	0,0067	0,2495	0,2136	0,7617	0,0207	0,0207
CC DT	107	0,0056	0,3706	0,3833	0,3281	0,0090	0,0876
CC GNB	107	0,0056	0,3289	0,3632	0,2388	0,0000	0,0515
RakelD LR	107	0,0069	0,4430	0,4078	0,6782	0,0387	0,0851
RakelD SVC	107	0,0056	0,2548	0,2192	0,7375	0,0194	0,0296
RakelD DT	107	0,0056	0,3837	0,3958	0,3420	0,0013	0,0683
RakelD GNB	107	0,0056	0,3376	0,5982	0,1232	0,0000	0,2319
BRkNN	107	0,0056	0,4014	0,4008	0,4044	0,0683	0,1327
MLkNN	107	0,0048	0,3453	0,3154	0,5577	0,0309	0,0554

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLTSVM	107	0,0963	0,2420	0,3420	0,1118	0,0000	0,0644
BR LR	131	0,0040	0,4080	0,3679	0,7221	0,0309	0,0670
BR SVC	131	0,0040	0,2482	0,2126	0,7578	0,0129	0,0129
BR DT	131	0,0056	0,3719	0,3859	0,3250	0,0039	0,0670
BR GNB	131	0,0056	0,3198	0,6876	0,1020	0,0000	0,3827
LP LR	131	0,0056	0,4202	0,4186	0,4273	0,0889	0,1546
LP SVC	131	0,0056	0,2239	0,2159	0,2660	0,0490	0,0619
LP DT	131	0,0056	0,3164	0,3166	0,3158	0,0387	0,0812
LP GNB	131	0,0056	0,2453	0,2389	0,2795	0,0464	0,0580
CC LR	131	0,0032	0,4236	0,3842	0,7173	0,0438	0,0773
CC SVC	131	0,0040	0,2420	0,2068	0,7617	0,0129	0,0129
CC DT	131	0,0056	0,3658	0,3793	0,3218	0,0065	0,0863
CC GNB	131	0,0056	0,3213	0,3523	0,2378	0,0000	0,0438
RakelD LR	131	0,0056	0,4374	0,4022	0,6741	0,0374	0,0812
RakelD SVC	131	0,0056	0,2513	0,2158	0,7446	0,0194	0,0206
RakelD DT	131	0,0056	0,3657	0,3782	0,3233	0,0039	0,0631
RakelD GNB	131	0,0056	0,3367	0,6100	0,1208	0,0000	0,2487
BRkNN	131	0,0056	0,3911	0,3908	0,3925	0,0657	0,1263
MLkNN	131	0,0047	0,3363	0,3062	0,5550	0,0309	0,0489
MLTSVM	131	0,0875	0,2295	0,3408	0,0997	0,0000	0,0618
BR LR	169	0,0027	0,4127	0,3715	0,7423	0,0155	0,0567
BR SVC	169	0,0027	0,2307	0,1961	0,7851	0,0103	0,0103
BR DT	169	0,0052	0,3608	0,3748	0,3142	0,0052	0,0670
BR GNB	169	0,0052	0,3135	0,6699	0,1003	0,0000	0,3634
LP LR	169	0,0052	0,4141	0,4126	0,4206	0,0876	0,1520
LP SVC	169	0,0052	0,2139	0,2054	0,2597	0,0477	0,0592
LP DT	169	0,0052	0,2966	0,2968	0,2959	0,0360	0,0683
LP GNB	169	0,0052	0,2202	0,2130	0,2586	0,0438	0,0541
CC LR	169	0,0027	0,4111	0,3715	0,7171	0,0206	0,0619
CC SVC	169	0,0027	0,2272	0,1930	0,7824	0,0103	0,0103
CC DT	169	0,0052	0,3621	0,3769	0,3136	0,0065	0,0876
CC GNB	169	0,0052	0,3094	0,3369	0,2334	0,0000	0,0374
RakelD LR	169	0,0052	0,4254	0,3887	0,6842	0,0309	0,0774
RakelD SVC	169	0,0052	0,2419	0,2070	0,7522	0,0168	0,0193
RakelD DT	169	0,0052	0,3572	0,3683	0,3187	0,0039	0,0657

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
RakelD GNB	169	0,0052	0,3227	0,6110	0,1118	0,0000	0,2654
BRkNN	169	0,0052	0,3818	0,3813	0,3842	0,0644	0,1250
MLkNN	169	0,0046	0,3284	0,2982	0,5547	0,0296	0,0451
MLTSVM	169	0,0697	0,1885	0,2921	0,0782	0,0000	0,0374
BR LR	281	0,0027	0,3989	0,3579	0,7377	0,0155	0,0515
BR SVC	281	0,0027	0,2227	0,1889	0,7851	0,0103	0,0103
BR DT	281	0,0050	0,3511	0,3673	0,2992	0,0026	0,0657
BR GNB	281	0,0050	0,3080	0,6435	0,0999	0,0000	0,3183
LP LR	281	0,0050	0,3900	0,3872	0,4018	0,0837	0,1456
LP SVC	281	0,0050	0,2038	0,1945	0,2552	0,0477	0,0554
LP DT	281	0,0050	0,2798	0,2797	0,2800	0,0399	0,0683
LP GNB	281	0,0050	0,2041	0,1961	0,2485	0,0438	0,0515
CC LR	281	0,0027	0,4014	0,3618	0,7137	0,0206	0,0567
CC SVC	281	0,0027	0,2193	0,1859	0,7824	0,0103	0,0103
CC DT	281	0,0050	0,3422	0,3577	0,2925	0,0039	0,0747
CC GNB	281	0,0050	0,2987	0,3224	0,2306	0,0000	0,0284
RakelD LR	281	0,0047	0,4021	0,3649	0,6795	0,0223	0,0567
RakelD SVC	281	0,0043	0,2326	0,1986	0,7464	0,0154	0,0181
RakelD DT	281	0,0050	0,3428	0,3562	0,2984	0,0026	0,0593
RakelD GNB	281	0,0050	0,3103	0,5904	0,1072	0,0013	0,2268
BRkNN	281	0,0050	0,3675	0,3665	0,3717	0,0593	0,1134
MLkNN	281	0,0044	0,3168	0,2862	0,5547	0,0296	0,0412
MLTSVM	281	0,0601	0,1697	0,2846	0,0651	0,0000	0,0374

Tabla A.2: Métricas Experimento 1 - Con tratamiento al desbalance.

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR LR	5	0,2378	0,7753	0,7943	0,7078	0,4665	0,7448
BR SVC	5	0,1955	0,6683	0,6955	0,5781	0,2977	0,6431
BR DT	5	0,2254	0,6201	0,6219	0,6150	0,2900	0,5618
BR GNB	5	0,1917	0,5600	0,6432	0,3693	0,0851	0,6224
LP LR	5	0,2484	0,7331	0,7559	0,6552	0,4382	0,7332
LP SVC	5	0,1910	0,5984	0,6362	0,4843	0,2126	0,6186
LP DT	5	0,2365	0,5797	0,5785	0,5857	0,3441	0,5644

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
LP GNB	5	0,1932	0,5651	0,6152	0,4271	0,1585	0,5464
CC LR	5	0,2572	0,7763	0,7917	0,7206	0,5219	0,7680
CC SVC	5	0,1966	0,6711	0,6955	0,5887	0,3118	0,6495
CC DT	5	0,2322	0,5927	0,5886	0,6104	0,3544	0,5735
CC GNB	5	0,1917	0,5564	0,6366	0,3703	0,0889	0,6224
RakelD LR	5	0,2409	0,7679	0,7946	0,6774	0,4588	0,7577
RakelD SVC	5	0,1994	0,6424	0,6749	0,5426	0,2745	0,6353
RakelD DT	5	0,2370	0,5843	0,5782	0,6112	0,3338	0,5554
RakelD GNB	5	0,1904	0,6096	0,7161	0,3826	0,0915	0,6585
BRkNN	5	0,2537	0,6200	0,6133	0,6484	0,4059	0,6031
MLkNN	5	0,2583	0,6167	0,6087	0,6558	0,3905	0,5760
MLTSVM	5	0,3901	0,4203	0,3862	0,6551	0,2668	0,3801
BR LR	14	0,1488	0,7206	0,7704	0,5734	0,1469	0,5644
BR SVC	14	0,1473	0,6362	0,7217	0,4322	0,0554	0,5077
BR DT	14	0,1484	0,5432	0,5427	0,5466	0,0915	0,2977
BR GNB	14	0,1480	0,5503	0,7207	0,2833	0,0245	0,5129
LP LR	14	0,1505	0,5836	0,5975	0,5351	0,1520	0,4046
LP SVC	14	0,1480	0,4223	0,4375	0,3723	0,0400	0,2345
LP DT	14	0,1417	0,4822	0,4814	0,4868	0,1224	0,3093
LP GNB	14	0,1490	0,5259	0,5305	0,5085	0,1366	0,3312
CC LR	14	0,1505	0,7181	0,7602	0,5889	0,1817	0,5773
CC SVC	14	0,1480	0,6402	0,7212	0,4423	0,0631	0,5090
CC DT	14	0,1480	0,5455	0,5436	0,5537	0,1418	0,3454
CC GNB	14	0,1480	0,5353	0,6755	0,2930	0,0245	0,4214
RakelD LR	14	0,1492	0,6790	0,7280	0,5357	0,1056	0,5000
RakelD SVC	14	0,1476	0,5752	0,6686	0,3702	0,0258	0,4484
RakelD DT	14	0,1450	0,5237	0,5242	0,5232	0,0902	0,2745
RakelD GNB	14	0,1480	0,5562	0,6739	0,3281	0,0090	0,4253
BRkNN	14	0,1590	0,5518	0,5509	0,5555	0,1894	0,3531
MLkNN	14	0,1560	0,5245	0,5050	0,6214	0,1727	0,2951
MLTSVM	14	0,2494	0,3553	0,3440	0,4097	0,0528	0,1830
BR LR	26	0,1222	0,6597	0,7335	0,4706	0,0284	0,4665
BR SVC	26	0,1216	0,5666	0,7176	0,3081	0,0039	0,4459
BR DT	26	0,1206	0,4838	0,4835	0,4852	0,0322	0,1752

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR GNB	26	0,1222	0,4788	0,7265	0,2032	0,0000	0,4678
LP LR	26	0,1248	0,5284	0,5385	0,4917	0,0992	0,2603
LP SVC	26	0,1192	0,4110	0,4103	0,4199	0,0374	0,1379
LP DT	26	0,1213	0,4049	0,4078	0,3939	0,0464	0,1649
LP GNB	26	0,1209	0,4318	0,4187	0,4939	0,0966	0,1959
CC LR	26	0,1222	0,6528	0,7092	0,4956	0,0541	0,4381
CC SVC	26	0,1222	0,5703	0,7150	0,3157	0,0064	0,4433
CC DT	26	0,1208	0,4738	0,4713	0,4842	0,0593	0,1959
CC GNB	26	0,1222	0,4512	0,5486	0,2643	0,0039	0,2565
RakelD LR	26	0,1222	0,6141	0,6943	0,4204	0,0245	0,3982
RakelD SVC	26	0,1222	0,5252	0,6715	0,2815	0,0064	0,3428
RakelD DT	26	0,1223	0,4608	0,4618	0,4570	0,0296	0,1546
RakelD GNB	26	0,1222	0,4901	0,6652	0,2399	0,0013	0,3672
BRkNN	26	0,1266	0,4947	0,4932	0,5006	0,1083	0,2397
MLkNN	26	0,1248	0,4524	0,4271	0,5940	0,0967	0,1662
MLTSVM	26	0,1772	0,3212	0,3348	0,2772	0,0065	0,1134
BR LR	34	0,0067	0,6296	0,7134	0,4284	0,0193	0,4163
BR SVC	34	0,0067	0,5394	0,7159	0,2716	0,0000	0,4226
BR DT	34	0,0067	0,4625	0,4621	0,4642	0,0232	0,1430
BR GNB	34	0,0067	0,4475	0,7293	0,1760	0,0000	0,4588
LP LR	34	0,0093	0,5203	0,5274	0,4940	0,1160	0,2526
LP SVC	34	0,0067	0,3667	0,3638	0,3843	0,0374	0,1276
LP DT	34	0,0080	0,4086	0,4110	0,3995	0,0490	0,1546
LP GNB	34	0,0089	0,3821	0,3645	0,4740	0,0683	0,1379
CC LR	34	0,0067	0,6254	0,6869	0,4606	0,0361	0,3931
CC SVC	34	0,0067	0,5434	0,7114	0,2796	0,0026	0,4188
CC DT	34	0,0080	0,4406	0,4381	0,4514	0,0387	0,1572
CC GNB	34	0,0067	0,4165	0,4919	0,2582	0,0000	0,1688
RakelD LR	34	0,0067	0,5866	0,6750	0,3849	0,0232	0,3467
RakelD SVC	34	0,0067	0,4933	0,6538	0,2490	0,0000	0,3131
RakelD DT	34	0,0067	0,4399	0,4436	0,4258	0,0155	0,1263
RakelD GNB	34	0,0067	0,4473	0,6213	0,2110	0,0000	0,2706
BRkNN	34	0,0119	0,4712	0,4700	0,4762	0,0941	0,2049
MLkNN	34	0,0149	0,4223	0,3956	0,5814	0,0748	0,1276

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLTSVM	34	0,1669	0,3199	0,3548	0,2302	0,0078	0,0966
BR LR	54	0,0066	0,5887	0,6839	0,3784	0,0078	0,3750
BR SVC	54	0,0066	0,4789	0,7091	0,2085	0,0000	0,4085
BR DT	54	0,0066	0,4161	0,4149	0,4219	0,0129	0,1057
BR GNB	54	0,0066	0,3928	0,7335	0,1375	0,0000	0,4600
LP LR	54	0,0092	0,4797	0,4838	0,4642	0,0915	0,2126
LP SVC	54	0,0079	0,3412	0,3350	0,3778	0,0387	0,1095
LP DT	54	0,0092	0,3580	0,3594	0,3526	0,0451	0,1108
LP GNB	54	0,0120	0,3252	0,3034	0,4572	0,0606	0,1044
CC LR	54	0,0066	0,5774	0,6425	0,4110	0,0206	0,3351
CC SVC	54	0,0066	0,4872	0,6993	0,2204	0,0000	0,3969
CC DT	54	0,0079	0,4130	0,4122	0,4168	0,0258	0,1301
CC GNB	54	0,0066	0,3694	0,4242	0,2437	0,0000	0,1160
RakelD LR	54	0,0066	0,5546	0,6570	0,3416	0,0090	0,3312
RakelD SVC	54	0,0066	0,4452	0,6436	0,1997	0,0000	0,2990
RakelD DT	54	0,0062	0,4176	0,4190	0,4128	0,0090	0,1108
RakelD GNB	54	0,0066	0,3858	0,5874	0,1627	0,0000	0,2307
BRkNN	54	0,0118	0,4364	0,4351	0,4416	0,0760	0,1727
MLkNN	54	0,0135	0,3823	0,3538	0,5664	0,0541	0,0979
MLTSVM	54	0,1314	0,2884	0,3512	0,1684	0,0026	0,0812
BR LR	63	0,0064	0,5786	0,6764	0,3668	0,0078	0,3634
BR SVC	63	0,0064	0,4623	0,7053	0,1947	0,0000	0,4033
BR DT	63	0,0064	0,4035	0,4034	0,4038	0,0103	0,0902
BR GNB	63	0,0064	0,3766	0,7278	0,1287	0,0000	0,4459
LP LR	63	0,0076	0,4678	0,4710	0,4556	0,0786	0,1843
LP SVC	63	0,0056	0,3237	0,3234	0,3423	0,0309	0,0980
LP DT	63	0,0064	0,3382	0,3389	0,3352	0,0386	0,0940
LP GNB	63	0,0081	0,3146	0,2941	0,4362	0,0593	0,0953
CC LR	63	0,0064	0,5678	0,6334	0,4015	0,0206	0,3170
CC SVC	63	0,0064	0,4724	0,6943	0,2076	0,0000	0,3905
CC DT	63	0,0064	0,3930	0,3923	0,3967	0,0194	0,1018
CC GNB	63	0,0064	0,3593	0,4094	0,2412	0,0000	0,0979
RakelD LR	63	0,0064	0,5486	0,6545	0,3331	0,0078	0,3325
RakelD SVC	63	0,0064	0,4289	0,6370	0,1862	0,0000	0,2912
RakelD DT	63	0,0064	0,4014	0,4048	0,3883	0,0129	0,1057

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
RakelD GNB	63	0,0064	0,3726	0,5817	0,1532	0,0013	0,2062
BRkNN	63	0,0090	0,4266	0,4254	0,4318	0,0734	0,1611
MLkNN	63	0,0094	0,3724	0,3435	0,5632	0,0490	0,0863
MLTSVM	63	0,1206	0,2766	0,3490	0,1515	0,0013	0,0760
BR LR	72	0,0059	0,5697	0,6693	0,3574	0,0078	0,3531
BR SVC	72	0,0059	0,4482	0,7024	0,1833	0,0000	0,3995
BR DT	72	0,0059	0,3982	0,3984	0,3983	0,0103	0,0902
BR GNB	72	0,0059	0,3622	0,7228	0,1210	0,0000	0,4420
LP LR	72	0,0072	0,4661	0,4694	0,4538	0,0838	0,1920
LP SVC	72	0,0072	0,3320	0,3322	0,3392	0,0322	0,1044
LP DT	72	0,0055	0,3312	0,3307	0,3330	0,0438	0,0979
LP GNB	72	0,0076	0,2994	0,2796	0,4194	0,0580	0,0928
CC LR	72	0,0059	0,5594	0,6266	0,3914	0,0168	0,3119
CC SVC	72	0,0059	0,4589	0,6880	0,1971	0,0000	0,3828
CC DT	72	0,0059	0,3924	0,3926	0,3922	0,0129	0,0876
CC GNB	72	0,0059	0,3525	0,3990	0,2404	0,0000	0,0902
RakelD LR	72	0,0059	0,5385	0,6458	0,3236	0,0064	0,3119
RakelD SVC	72	0,0059	0,4192	0,6393	0,1764	0,0000	0,3015
RakelD DT	72	0,0072	0,3830	0,3845	0,3775	0,0065	0,0786
RakelD GNB	72	0,0059	0,3548	0,5699	0,1414	0,0000	0,2216
BRkNN	72	0,0085	0,4198	0,4187	0,4244	0,0683	0,1546
MLkNN	72	0,0089	0,3652	0,3360	0,5626	0,0425	0,0799
MLTSVM	72	0,1126	0,2666	0,3475	0,1384	0,0013	0,0734
BR LR	82	0,0056	0,5601	0,6627	0,3461	0,0064	0,3518
BR SVC	82	0,0056	0,4364	0,7004	0,1742	0,0000	0,4021
BR DT	82	0,0056	0,3953	0,3960	0,3926	0,0078	0,0838
BR GNB	82	0,0056	0,3518	0,7189	0,1158	0,0000	0,4459
LP LR	82	0,0082	0,4604	0,4642	0,4467	0,0825	0,1881
LP SVC	82	0,0056	0,3339	0,3298	0,3655	0,0310	0,0825
LP DT	82	0,0082	0,3144	0,3142	0,3151	0,0335	0,0850
LP GNB	82	0,0078	0,2820	0,2622	0,4043	0,0567	0,0799
CC LR	82	0,0056	0,5540	0,6234	0,3834	0,0168	0,3118
CC SVC	82	0,0056	0,4473	0,6846	0,1876	0,0000	0,3815
CC DT	82	0,0056	0,3846	0,3852	0,3823	0,0180	0,0838

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
CC GNB	82	0,0056	0,3411	0,3803	0,2418	0,0000	0,0773
RakelD LR	82	0,0056	0,5277	0,6366	0,3135	0,0065	0,3080
RakelD SVC	82	0,0056	0,4090	0,6376	0,1682	0,0000	0,2899
RakelD DT	82	0,0069	0,3797	0,3818	0,3720	0,0052	0,0709
RakelD GNB	82	0,0056	0,3462	0,5716	0,1346	0,0000	0,2255
BRkNN	82	0,0082	0,4146	0,4137	0,4186	0,0683	0,1533
MLkNN	82	0,0073	0,3585	0,3290	0,5599	0,0387	0,0760
MLTSVM	82	0,1078	0,2596	0,3459	0,1302	0,0013	0,0709
BR LR	91	0,0056	0,5512	0,6555	0,3372	0,0052	0,3376
BR SVC	91	0,0056	0,4220	0,6976	0,1638	0,0000	0,3956
BR DT	91	0,0056	0,3901	0,3902	0,3900	0,0129	0,0786
BR GNB	91	0,0056	0,3423	0,7141	0,1112	0,0000	0,4356
LP LR	91	0,0082	0,4548	0,4580	0,4430	0,0799	0,1765
LP SVC	91	0,0056	0,3236	0,3224	0,3297	0,0284	0,0851
LP DT	91	0,0095	0,3189	0,3186	0,3204	0,0361	0,0915
LP GNB	91	0,0065	0,2707	0,2531	0,3751	0,0567	0,0799
CC LR	91	0,0056	0,5441	0,6121	0,3768	0,0168	0,3003
CC SVC	91	0,0056	0,4369	0,6806	0,1798	0,0000	0,3763
CC DT	91	0,0056	0,3788	0,3789	0,3789	0,0155	0,0760
CC GNB	91	0,0056	0,3356	0,3726	0,2403	0,0000	0,0683
RakelD LR	91	0,0056	0,5237	0,6344	0,3087	0,0065	0,2977
RakelD SVC	91	0,0056	0,3851	0,6209	0,1530	0,0000	0,2899
RakelD DT	91	0,0056	0,3741	0,3754	0,3690	0,0078	0,0709
RakelD GNB	91	0,0056	0,3390	0,5712	0,1292	0,0000	0,2423
BRkNN	91	0,0082	0,4096	0,4090	0,4125	0,0683	0,1469
MLkNN	91	0,0073	0,3531	0,3234	0,5597	0,0361	0,0709
MLTSVM	91	0,1034	0,2524	0,3440	0,1226	0,0013	0,0683
BR LR	107	0,0041	0,5436	0,6558	0,3228	0,0104	0,3273
BR SVC	107	0,0041	0,4036	0,6934	0,1512	0,0000	0,4073
BR DT	107	0,0056	0,3830	0,3833	0,3818	0,0078	0,0605
BR GNB	107	0,0056	0,3330	0,7041	0,1073	0,0000	0,4111
LP LR	107	0,0056	0,4428	0,4447	0,4357	0,0786	0,1623
LP SVC	107	0,0056	0,3714	0,3745	0,3598	0,0270	0,0954
LP DT	107	0,0056	0,3076	0,3076	0,3074	0,0322	0,0670

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
LP GNB	107	0,0056	0,2474	0,2412	0,2804	0,0464	0,0631
CC LR	107	0,0041	0,5336	0,5998	0,3704	0,0104	0,2861
CC SVC	107	0,0041	0,4204	0,6738	0,1679	0,0000	0,3891
CC DT	107	0,0056	0,3657	0,3656	0,3660	0,0129	0,0528
CC GNB	107	0,0056	0,3289	0,3632	0,2388	0,0000	0,0515
RakelD LR	107	0,0056	0,5234	0,6385	0,3044	0,0078	0,2951
RakelD SVC	107	0,0056	0,3794	0,6295	0,1466	0,0000	0,2796
RakelD DT	107	0,0056	0,3758	0,3772	0,3706	0,0064	0,0606
RakelD GNB	107	0,0056	0,3352	0,5892	0,1232	0,0000	0,2114
BRkNN	107	0,0056	0,4014	0,4008	0,4044	0,0683	0,1327
MLkNN	107	0,0048	0,3453	0,3154	0,5577	0,0309	0,0554
MLTSVM	107	0,0963	0,2420	0,3420	0,1118	0,0000	0,0644
BR LR	131	0,0040	0,5284	0,6426	0,3088	0,0104	0,3170
BR SVC	131	0,0040	0,3831	0,6858	0,1386	0,0000	0,3943
BR DT	131	0,0056	0,3648	0,3649	0,3645	0,0090	0,0515
BR GNB	131	0,0056	0,3198	0,6876	0,1020	0,0000	0,3827
LP LR	131	0,0056	0,4385	0,4402	0,4319	0,0773	0,1598
LP SVC	131	0,0056	0,2668	0,2658	0,2716	0,0284	0,0592
LP DT	131	0,0056	0,2887	0,2886	0,2892	0,0348	0,0799
LP GNB	131	0,0056	0,2453	0,2389	0,2795	0,0464	0,0580
CC LR	131	0,0040	0,5188	0,5853	0,3570	0,0104	0,2861
CC SVC	131	0,0040	0,4020	0,6621	0,1564	0,0000	0,3789
CC DT	131	0,0056	0,3501	0,3510	0,3469	0,0155	0,0438
CC GNB	131	0,0056	0,3213	0,3523	0,2378	0,0000	0,0438
RakelD LR	131	0,0056	0,5084	0,6270	0,2895	0,0065	0,3042
RakelD SVC	131	0,0056	0,3609	0,6236	0,1345	0,0000	0,2900
RakelD DT	131	0,0056	0,3701	0,3735	0,3569	0,0064	0,0580
RakelD GNB	131	0,0056	0,3306	0,6046	0,1176	0,0000	0,2474
BRkNN	131	0,0056	0,3911	0,3908	0,3925	0,0657	0,1263
MLkNN	131	0,0047	0,3363	0,3062	0,5550	0,0309	0,0489
MLTSVM	131	0,0875	0,2295	0,3408	0,0997	0,0000	0,0618
BR LR	169	0,0027	0,5140	0,6357	0,2911	0,0155	0,2990
BR SVC	169	0,0027	0,3637	0,6863	0,1263	0,0000	0,3608
BR DT	169	0,0052	0,3545	0,3550	0,3528	0,0064	0,0502

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR GNB	169	0,0052	0,3135	0,6699	0,1003	0,0000	0,3634
LP LR	169	0,0052	0,4194	0,4206	0,4149	0,0734	0,1469
LP SVC	169	0,0052	0,3142	0,3132	0,3212	0,0309	0,0786
LP DT	169	0,0052	0,3137	0,3144	0,3108	0,0348	0,0683
LP GNB	169	0,0052	0,2202	0,2130	0,2586	0,0438	0,0541
CC LR	169	0,0027	0,4965	0,5635	0,3366	0,0206	0,2784
CC SVC	169	0,0027	0,3869	0,6584	0,1460	0,0000	0,3557
CC DT	169	0,0052	0,3415	0,3420	0,3399	0,0103	0,0438
CC GNB	169	0,0052	0,3094	0,3369	0,2334	0,0000	0,0374
RakelD LR	169	0,0052	0,4943	0,6082	0,2828	0,0064	0,2719
RakelD SVC	169	0,0052	0,3312	0,6059	0,1179	0,0000	0,2861
RakelD DT	169	0,0052	0,3590	0,3629	0,3443	0,0064	0,0464
RakelD GNB	169	0,0052	0,3198	0,6010	0,1116	0,0000	0,2565
BRkNN	169	0,0052	0,3818	0,3813	0,3842	0,0644	0,1250
MLkNN	169	0,0046	0,3284	0,2982	0,5547	0,0296	0,0451
MLTSVM	169	0,0697	0,1885	0,2921	0,0782	0,0000	0,0374
BR LR	281	0,0027	0,4922	0,6183	0,2711	0,0103	0,2680
BR SVC	281	0,0027	0,3400	0,6680	0,1147	0,0000	0,3402
BR DT	281	0,0050	0,3501	0,3540	0,3354	0,0078	0,0515
BR GNB	281	0,0050	0,3080	0,6435	0,0999	0,0000	0,3183
LP LR	281	0,0050	0,4040	0,4054	0,3988	0,0696	0,1443
LP SVC	281	0,0050	0,3082	0,3082	0,3089	0,0232	0,0644
LP DT	281	0,0050	0,2843	0,2846	0,2833	0,0283	0,0541
LP GNB	281	0,0050	0,2041	0,1961	0,2485	0,0438	0,0515
CC LR	281	0,0027	0,4800	0,5517	0,3159	0,0103	0,2629
CC SVC	281	0,0027	0,3660	0,6392	0,1351	0,0000	0,3247
CC DT	281	0,0050	0,3253	0,3264	0,3218	0,0052	0,0335
CC GNB	281	0,0050	0,2987	0,3224	0,2306	0,0000	0,0284
RakelD LR	281	0,0045	0,4767	0,5920	0,2682	0,0052	0,2698
RakelD SVC	281	0,0050	0,3099	0,6069	0,1051	0,0000	0,2693
RakelD DT	281	0,0050	0,3296	0,3326	0,3184	0,0039	0,0425
RakelD GNB	281	0,0050	0,3062	0,5851	0,1054	0,0000	0,2319
BRkNN	281	0,0050	0,3675	0,3665	0,3717	0,0593	0,1134
MLkNN	281	0,0044	0,3168	0,2862	0,5547	0,0296	0,0412

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLTSVM	281	0,0601	0,1697	0,2846	0,0651	0,0000	0,0374

# Anexo B

## Experimento 2

Tabla B.1: Métricas Experimento 2 - Descriptores ResNet50.

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR LR	5	0,2378	0,7753	0,7943	0,7078	0,4665	0,9046
BR SVC	5	0,1955	0,6683	0,6955	0,5781	0,2977	0,8492
BR DT	5	0,2254	0,6201	0,6219	0,6150	0,2900	0,7848
BR GNB	5	0,1917	0,5600	0,6432	0,3693	0,0851	0,8248
LP LR	5	0,2484	0,7331	0,7559	0,6552	0,4382	0,8505
LP SVC	5	0,1910	0,5984	0,6362	0,4843	0,2126	0,7732
LP DT	5	0,2365	0,5797	0,5785	0,5857	0,3441	0,7023
LP GNB	5	0,1932	0,5651	0,6152	0,4271	0,1585	0,7551
CC LR	5	0,2572	0,7763	0,7917	0,7206	0,5219	0,8853
CC SVC	5	0,1966	0,6711	0,6955	0,5887	0,3118	0,8389
CC DT	5	0,2322	0,5927	0,5886	0,6104	0,3544	0,7100
CC GNB	5	0,1917	0,5564	0,6366	0,3703	0,0889	0,8183
RakelD LR	5	0,2409	0,7679	0,7946	0,6774	0,4588	0,8956
RakelD SVC	5	0,1994	0,6424	0,6749	0,5426	0,2745	0,8299
RakelD DT	5	0,2370	0,5843	0,5782	0,6112	0,3338	0,7152
RakelD GNB	5	0,1904	0,6096	0,7161	0,3826	0,0915	0,8492
BRkNN	5	0,2537	0,6200	0,6133	0,6484	0,4059	0,7307
MLkNN	5	0,2583	0,6167	0,6087	0,6558	0,3905	0,7204
MLTSVM	5	0,3901	0,4203	0,3862	0,6551	0,2668	0,5683
BR LR	14	0,1488	0,7206	0,7704	0,5734	0,1469	0,9355
BR SVC	14	0,1473	0,6362	0,7217	0,4322	0,0554	0,9085
BR DT	14	0,1484	0,5432	0,5427	0,5466	0,0915	0,8286

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR GNB	14	0,1480	0,5503	0,7207	0,2833	0,0245	0,9059
LP LR	14	0,1505	0,5836	0,5975	0,5351	0,1520	0,7977
LP SVC	14	0,1480	0,4223	0,4375	0,3723	0,0400	0,6675
LP DT	14	0,1417	0,4822	0,4814	0,4868	0,1224	0,7062
LP GNB	14	0,1490	0,5259	0,5305	0,5085	0,1366	0,7539
CC LR	14	0,1505	0,7181	0,7602	0,5889	0,1817	0,9137
CC SVC	14	0,1480	0,6402	0,7212	0,4423	0,0631	0,9059
CC DT	14	0,1480	0,5455	0,5436	0,5537	0,1418	0,7771
CC GNB	14	0,1480	0,5353	0,6755	0,2930	0,0245	0,8737
RakelD LR	14	0,1492	0,6790	0,7280	0,5357	0,1056	0,9291
RakelD SVC	14	0,1476	0,5752	0,6686	0,3702	0,0258	0,8814
RakelD DT	14	0,1450	0,5237	0,5242	0,5232	0,0902	0,7951
RakelD GNB	14	0,1480	0,5562	0,6739	0,3281	0,0090	0,8969
BRkNN	14	0,1590	0,5518	0,5509	0,5555	0,1894	0,7680
MLkNN	14	0,1560	0,5245	0,5050	0,6214	0,1727	0,7281
MLTSVM	14	0,2494	0,3553	0,3440	0,4097	0,0528	0,6147
BR LR	26	0,1222	0,6597	0,7335	0,4706	0,0284	0,9317
BR SVC	26	0,1216	0,5666	0,7176	0,3081	0,0039	0,9252
BR DT	26	0,1206	0,4838	0,4835	0,4852	0,0322	0,8118
BR GNB	26	0,1222	0,4788	0,7265	0,2032	0,0000	0,9214
LP LR	26	0,1248	0,5284	0,5385	0,4917	0,0992	0,7706
LP SVC	26	0,1192	0,4110	0,4103	0,4199	0,0374	0,6830
LP DT	26	0,1213	0,4049	0,4078	0,3939	0,0464	0,6662
LP GNB	26	0,1209	0,4318	0,4187	0,4939	0,0966	0,6701
CC LR	26	0,1222	0,6528	0,7092	0,4956	0,0541	0,9046
CC SVC	26	0,1222	0,5703	0,7150	0,3157	0,0064	0,9175
CC DT	26	0,1208	0,4738	0,4713	0,4842	0,0593	0,7745
CC GNB	26	0,1222	0,4512	0,5486	0,2643	0,0039	0,8209
RakelD LR	26	0,1222	0,6141	0,6943	0,4204	0,0245	0,9291
RakelD SVC	26	0,1222	0,5252	0,6715	0,2815	0,0064	0,9175
RakelD DT	26	0,1223	0,4608	0,4618	0,4570	0,0296	0,7964
RakelD GNB	26	0,1222	0,4901	0,6652	0,2399	0,0013	0,9227
BRkNN	26	0,1266	0,4947	0,4932	0,5006	0,1083	0,7500
MLkNN	26	0,1248	0,4524	0,4271	0,5940	0,0967	0,6984

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLTSVM	26	0,1772	0,3212	0,3348	0,2772	0,0065	0,6623
BR LR	34	0,0067	0,6296	0,7134	0,4284	0,0193	0,9304
BR SVC	34	0,0067	0,5394	0,7159	0,2716	0,0000	0,9330
BR DT	34	0,0067	0,4625	0,4621	0,4642	0,0232	0,8235
BR GNB	34	0,0067	0,4475	0,7293	0,1760	0,0000	0,9317
LP LR	34	0,0093	0,5203	0,5274	0,4940	0,1160	0,7964
LP SVC	34	0,0067	0,3667	0,3638	0,3843	0,0374	0,6314
LP DT	34	0,0080	0,4086	0,4110	0,3995	0,0490	0,6972
LP GNB	34	0,0089	0,3821	0,3645	0,4740	0,0683	0,6405
CC LR	34	0,0067	0,6254	0,6869	0,4606	0,0361	0,9034
CC SVC	34	0,0067	0,5434	0,7114	0,2796	0,0026	0,9240
CC DT	34	0,0080	0,4406	0,4381	0,4514	0,0387	0,7513
CC GNB	34	0,0067	0,4165	0,4919	0,2582	0,0000	0,7977
RakelD LR	34	0,0067	0,5866	0,6750	0,3849	0,0232	0,9253
RakelD SVC	34	0,0067	0,4933	0,6538	0,2490	0,0000	0,9149
RakelD DT	34	0,0067	0,4399	0,4436	0,4258	0,0155	0,8003
RakelD GNB	34	0,0067	0,4473	0,6213	0,2110	0,0000	0,8943
BRkNN	34	0,0119	0,4712	0,4700	0,4762	0,0941	0,7539
MLkNN	34	0,0149	0,4223	0,3956	0,5814	0,0748	0,6972
MLTSVM	34	0,1669	0,3199	0,3548	0,2302	0,0078	0,7036
BR LR	54	0,0066	0,5887	0,6839	0,3784	0,0078	0,9407
BR SVC	54	0,0066	0,4789	0,7091	0,2085	0,0000	0,9498
BR DT	54	0,0066	0,4161	0,4149	0,4219	0,0129	0,8479
BR GNB	54	0,0066	0,3928	0,7335	0,1375	0,0000	0,9497
LP LR	54	0,0092	0,4797	0,4838	0,4642	0,0915	0,8015
LP SVC	54	0,0079	0,3412	0,3350	0,3778	0,0387	0,6637
LP DT	54	0,0092	0,3580	0,3594	0,3526	0,0451	0,6869
LP GNB	54	0,0120	0,3252	0,3034	0,4572	0,0606	0,5863
CC LR	54	0,0066	0,5774	0,6425	0,4110	0,0206	0,9098
CC SVC	54	0,0066	0,4872	0,6993	0,2204	0,0000	0,9382
CC DT	54	0,0079	0,4130	0,4122	0,4168	0,0258	0,7861
CC GNB	54	0,0066	0,3694	0,4242	0,2437	0,0000	0,7861
RakelD LR	54	0,0066	0,5546	0,6570	0,3416	0,0090	0,9485
RakelD SVC	54	0,0066	0,4452	0,6436	0,1997	0,0000	0,9497
RakelD DT	54	0,0062	0,4176	0,4190	0,4128	0,0090	0,8389

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
RakelD GNB	54	0,0066	0,3858	0,5874	0,1627	0,0000	0,9098
BRkNN	54	0,0118	0,4364	0,4351	0,4416	0,0760	0,7616
MLkNN	54	0,0135	0,3823	0,3538	0,5664	0,0541	0,6997
MLTSVM	54	0,1314	0,2884	0,3512	0,1684	0,0026	0,7423
BR LR	63	0,0064	0,5786	0,6764	0,3668	0,0078	0,9381
BR SVC	63	0,0064	0,4623	0,7053	0,1947	0,0000	0,9523
BR DT	63	0,0064	0,4035	0,4034	0,4038	0,0103	0,8222
BR GNB	63	0,0064	0,3766	0,7278	0,1287	0,0000	0,9497
LP LR	63	0,0076	0,4678	0,4710	0,4556	0,0786	0,8015
LP SVC	63	0,0056	0,3237	0,3234	0,3423	0,0309	0,6018
LP DT	63	0,0064	0,3382	0,3389	0,3352	0,0386	0,6753
LP GNB	63	0,0081	0,3146	0,2941	0,4362	0,0593	0,5799
CC LR	63	0,0064	0,5678	0,6334	0,4015	0,0206	0,9021
CC SVC	63	0,0064	0,4724	0,6943	0,2076	0,0000	0,9395
CC DT	63	0,0064	0,3930	0,3923	0,3967	0,0194	0,7797
CC GNB	63	0,0064	0,3593	0,4094	0,2412	0,0000	0,7796
RakelD LR	63	0,0064	0,5486	0,6545	0,3331	0,0078	0,9497
RakelD SVC	63	0,0064	0,4289	0,6370	0,1862	0,0000	0,9381
RakelD DT	63	0,0064	0,4014	0,4048	0,3883	0,0129	0,8157
RakelD GNB	63	0,0064	0,3726	0,5817	0,1532	0,0013	0,9201
BRkNN	63	0,0090	0,4266	0,4254	0,4318	0,0734	0,7565
MLkNN	63	0,0094	0,3724	0,3435	0,5632	0,0490	0,6920
MLTSVM	63	0,1206	0,2766	0,3490	0,1515	0,0013	0,7449
BR LR	72	0,0059	0,5697	0,6693	0,3574	0,0078	0,9420
BR SVC	72	0,0059	0,4482	0,7024	0,1833	0,0000	0,9562
BR DT	72	0,0059	0,3982	0,3984	0,3983	0,0103	0,8273
BR GNB	72	0,0059	0,3622	0,7228	0,1210	0,0000	0,9549
LP LR	72	0,0072	0,4661	0,4694	0,4538	0,0838	0,8067
LP SVC	72	0,0072	0,3320	0,3322	0,3392	0,0322	0,6353
LP DT	72	0,0055	0,3312	0,3307	0,3330	0,0438	0,6792
LP GNB	72	0,0076	0,2994	0,2796	0,4194	0,0580	0,5644
CC LR	72	0,0059	0,5594	0,6266	0,3914	0,0168	0,9046
CC SVC	72	0,0059	0,4589	0,6880	0,1971	0,0000	0,9407
CC DT	72	0,0059	0,3924	0,3926	0,3922	0,0129	0,7913

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
CC GNB	72	0,0059	0,3525	0,3990	0,2404	0,0000	0,7835
RakelD LR	72	0,0059	0,5385	0,6458	0,3236	0,0064	0,9355
RakelD SVC	72	0,0059	0,4192	0,6393	0,1764	0,0000	0,9407
RakelD DT	72	0,0072	0,3830	0,3845	0,3775	0,0065	0,8157
RakelD GNB	72	0,0059	0,3548	0,5699	0,1414	0,0000	0,9162
BRkNN	72	0,0085	0,4198	0,4187	0,4244	0,0683	0,7565
MLkNN	72	0,0089	0,3652	0,3360	0,5626	0,0425	0,6920
MLTSVM	72	0,1126	0,2666	0,3475	0,1384	0,0013	0,7565
BR LR	82	0,0056	0,5601	0,6627	0,3461	0,0064	0,9407
BR SVC	82	0,0056	0,4364	0,7004	0,1742	0,0000	0,9549
BR DT	82	0,0056	0,3953	0,3960	0,3926	0,0078	0,8299
BR GNB	82	0,0056	0,3518	0,7189	0,1158	0,0000	0,9549
LP LR	82	0,0082	0,4604	0,4642	0,4467	0,0825	0,8054
LP SVC	82	0,0056	0,3339	0,3298	0,3655	0,0310	0,6843
LP DT	82	0,0082	0,3144	0,3142	0,3151	0,0335	0,6379
LP GNB	82	0,0078	0,2820	0,2622	0,4043	0,0567	0,5529
CC LR	82	0,0056	0,5540	0,6234	0,3834	0,0168	0,9034
CC SVC	82	0,0056	0,4473	0,6846	0,1876	0,0000	0,9395
CC DT	82	0,0056	0,3846	0,3852	0,3823	0,0180	0,7848
CC GNB	82	0,0056	0,3411	0,3803	0,2418	0,0000	0,7719
RakelD LR	82	0,0056	0,5277	0,6366	0,3135	0,0065	0,9394
RakelD SVC	82	0,0056	0,4090	0,6376	0,1682	0,0000	0,9407
RakelD DT	82	0,0069	0,3797	0,3818	0,3720	0,0052	0,8209
RakelD GNB	82	0,0056	0,3462	0,5716	0,1346	0,0000	0,9201
BRkNN	82	0,0082	0,4146	0,4137	0,4186	0,0683	0,7564
MLkNN	82	0,0073	0,3585	0,3290	0,5599	0,0387	0,6907
MLTSVM	82	0,1078	0,2596	0,3459	0,1302	0,0013	0,7590
BR LR	91	0,0056	0,5512	0,6555	0,3372	0,0052	0,9407
BR SVC	91	0,0056	0,4220	0,6976	0,1638	0,0000	0,9588
BR DT	91	0,0056	0,3901	0,3902	0,3900	0,0129	0,8260
BR GNB	91	0,0056	0,3423	0,7141	0,1112	0,0000	0,9562
LP LR	91	0,0082	0,4548	0,4580	0,4430	0,0799	0,7938
LP SVC	91	0,0056	0,3236	0,3224	0,3297	0,0284	0,6688
LP DT	91	0,0095	0,3189	0,3186	0,3204	0,0361	0,6817

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
LP GNB	91	0,0065	0,2707	0,2531	0,3751	0,0567	0,5387
CC LR	91	0,0056	0,5441	0,6121	0,3768	0,0168	0,8995
CC SVC	91	0,0056	0,4369	0,6806	0,1798	0,0000	0,9407
CC DT	91	0,0056	0,3788	0,3789	0,3789	0,0155	0,7938
CC GNB	91	0,0056	0,3356	0,3726	0,2403	0,0000	0,7719
RakelD LR	91	0,0056	0,5237	0,6344	0,3087	0,0065	0,9368
RakelD SVC	91	0,0056	0,3851	0,6209	0,1530	0,0000	0,9420
RakelD DT	91	0,0056	0,3741	0,3754	0,3690	0,0078	0,8170
RakelD GNB	91	0,0056	0,3390	0,5712	0,1292	0,0000	0,9111
BRkNN	91	0,0082	0,4096	0,4090	0,4125	0,0683	0,7564
MLkNN	91	0,0073	0,3531	0,3234	0,5597	0,0361	0,6894
MLTSVM	91	0,1034	0,2524	0,3440	0,1226	0,0013	0,7603
BR LR	107	0,0041	0,5436	0,6558	0,3228	0,0104	0,9433
BR SVC	107	0,0041	0,4036	0,6934	0,1512	0,0000	0,9536
BR DT	107	0,0056	0,3830	0,3833	0,3818	0,0078	0,8363
BR GNB	107	0,0056	0,3330	0,7041	0,1073	0,0000	0,9498
LP LR	107	0,0056	0,4428	0,4447	0,4357	0,0786	0,7886
LP SVC	107	0,0056	0,3714	0,3745	0,3598	0,0270	0,7307
LP DT	107	0,0056	0,3076	0,3076	0,3074	0,0322	0,6289
LP GNB	107	0,0056	0,2474	0,2412	0,2804	0,0464	0,5078
CC LR	107	0,0041	0,5336	0,5998	0,3704	0,0104	0,8891
CC SVC	107	0,0041	0,4204	0,6738	0,1679	0,0000	0,9304
CC DT	107	0,0056	0,3657	0,3656	0,3660	0,0129	0,7693
CC GNB	107	0,0056	0,3289	0,3632	0,2388	0,0000	0,7654
RakelD LR	107	0,0056	0,5234	0,6385	0,3044	0,0078	0,9433
RakelD SVC	107	0,0056	0,3794	0,6295	0,1466	0,0000	0,9510
RakelD DT	107	0,0056	0,3758	0,3772	0,3706	0,0064	0,8170
RakelD GNB	107	0,0056	0,3352	0,5892	0,1232	0,0000	0,9227
BRkNN	107	0,0056	0,4014	0,4008	0,4044	0,0683	0,7526
MLkNN	107	0,0048	0,3453	0,3154	0,5577	0,0309	0,6830
MLTSVM	107	0,0963	0,2420	0,3420	0,1118	0,0000	0,7654
BR LR	131	0,0040	0,5284	0,6426	0,3088	0,0104	0,9407
BR SVC	131	0,0040	0,3831	0,6858	0,1386	0,0000	0,9510
BR DT	131	0,0056	0,3648	0,3649	0,3645	0,0090	0,8196

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
BR GNB	131	0,0056	0,3198	0,6876	0,1020	0,0000	0,9498
LP LR	131	0,0056	0,4385	0,4402	0,4319	0,0773	0,7990
LP SVC	131	0,0056	0,2668	0,2658	0,2716	0,0284	0,5786
LP DT	131	0,0056	0,2887	0,2886	0,2892	0,0348	0,6069
LP GNB	131	0,0056	0,2453	0,2389	0,2795	0,0464	0,5181
CC LR	131	0,0040	0,5188	0,5853	0,3570	0,0104	0,8866
CC SVC	131	0,0040	0,4020	0,6621	0,1564	0,0000	0,9279
CC DT	131	0,0056	0,3501	0,3510	0,3469	0,0155	0,7694
CC GNB	131	0,0056	0,3213	0,3523	0,2378	0,0000	0,7655
RakelD LR	131	0,0056	0,5084	0,6270	0,2895	0,0065	0,9459
RakelD SVC	131	0,0056	0,3609	0,6236	0,1345	0,0000	0,9471
RakelD DT	131	0,0056	0,3701	0,3735	0,3569	0,0064	0,8118
RakelD GNB	131	0,0056	0,3306	0,6046	0,1176	0,0000	0,9356
BRkNN	131	0,0056	0,3911	0,3908	0,3925	0,0657	0,7513
MLkNN	131	0,0047	0,3363	0,3062	0,5550	0,0309	0,6817
MLTSVM	131	0,0875	0,2295	0,3408	0,0997	0,0000	0,7732
BR LR	169	0,0027	0,5140	0,6357	0,2911	0,0155	0,9588
BR SVC	169	0,0027	0,3637	0,6863	0,1263	0,0000	0,9742
BR DT	169	0,0052	0,3545	0,3550	0,3528	0,0064	0,8247
BR GNB	169	0,0052	0,3135	0,6699	0,1003	0,0000	0,9498
LP LR	169	0,0052	0,4194	0,4206	0,4149	0,0734	0,7887
LP SVC	169	0,0052	0,3142	0,3132	0,3212	0,0309	0,6456
LP DT	169	0,0052	0,3137	0,3144	0,3108	0,0348	0,6856
LP GNB	169	0,0052	0,2202	0,2130	0,2586	0,0438	0,4884
CC LR	169	0,0027	0,4965	0,5635	0,3366	0,0206	0,8814
CC SVC	169	0,0027	0,3869	0,6584	0,1460	0,0000	0,9485
CC DT	169	0,0052	0,3415	0,3420	0,3399	0,0103	0,7693
CC GNB	169	0,0052	0,3094	0,3369	0,2334	0,0000	0,7603
RakelD LR	169	0,0052	0,4943	0,6082	0,2828	0,0064	0,9446
RakelD SVC	169	0,0052	0,3312	0,6059	0,1179	0,0000	0,9433
RakelD DT	169	0,0052	0,3590	0,3629	0,3443	0,0064	0,8428
RakelD GNB	169	0,0052	0,3198	0,6010	0,1116	0,0000	0,9407
BRkNN	169	0,0052	0,3818	0,3813	0,3842	0,0644	0,7513
MLkNN	169	0,0046	0,3284	0,2982	0,5547	0,0296	0,6817

Continua en la página siguiente

Método	Etiquetas	<i>Hamming Score</i>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	EMR	5-MR
MLTSVM	169	0,0697	0,1885	0,2921	0,0782	0,0000	0,7423
BR LR	281	0,0027	0,4922	0,6183	0,2711	0,0103	0,9588
BR SVC	281	0,0027	0,3400	0,6680	0,1147	0,0000	0,9742
BR DT	281	0,0050	0,3501	0,3540	0,3354	0,0078	0,8312
BR GNB	281	0,0050	0,3080	0,6435	0,0999	0,0000	0,9498
LP LR	281	0,0050	0,4040	0,4054	0,3988	0,0696	0,7810
LP SVC	281	0,0050	0,3082	0,3082	0,3089	0,0232	0,6959
LP DT	281	0,0050	0,2843	0,2846	0,2833	0,0283	0,6585
LP GNB	281	0,0050	0,2041	0,1961	0,2485	0,0438	0,4703
CC LR	281	0,0027	0,4800	0,5517	0,3159	0,0103	0,8918
CC SVC	281	0,0027	0,3660	0,6392	0,1351	0,0000	0,9485
CC DT	281	0,0050	0,3253	0,3264	0,3218	0,0052	0,7642
CC GNB	281	0,0050	0,2987	0,3224	0,2306	0,0000	0,7603
RakelD LR	281	0,0045	0,4767	0,5920	0,2682	0,0052	0,9450
RakelD SVC	281	0,0050	0,3099	0,6069	0,1051	0,0000	0,9575
RakelD DT	281	0,0050	0,3296	0,3326	0,3184	0,0039	0,8002
RakelD GNB	281	0,0050	0,3062	0,5851	0,1054	0,0000	0,9420
BRkNN	281	0,0050	0,3675	0,3665	0,3717	0,0593	0,7513
MLkNN	281	0,0044	0,3168	0,2862	0,5547	0,0296	0,6817
MLTSVM	281	0,0601	0,1697	0,2846	0,0651	0,0000	0,7449

# Anexo C

## Experimento 3

Tabla C.1: AUC por métrica - Experimento 3. **Observación:** los datos se ordenan en forma decreciente según AUC de *Micro F<sub>2</sub> Score*.

Origen des- cripto- res	Cjto.	Método	AUC <i>Micro</i> <i>F<sub>2</sub></i>	AUC <i>Micro</i> <i>Recall</i>	AUC <i>Micro</i> <i>Preci- sion</i>	AUC <i>Ham- ming</i> <i>Score</i>	AUC <i>EMR</i>	AUC <i>5-MR</i>
ResNet18	14	BR LR	244,0482	284,6690	157,0036	6,8159	13,1961	164,5468
ResNet18	15	BR LR	243,9790	280,1737	162,0962	7,0063	10,7615	153,8686
ResNet50	12	BR LR	243,3054	283,7648	156,7695	6,8565	9,2593	169,7548
ResNet18	13	BR LR	242,4738	276,9210	163,1859	6,8806	16,4395	159,3991
ResNet18	12	BR LR	242,3744	276,7339	163,2460	6,9224	14,9021	159,2278
ResNet18	8	BR LR	242,2549	276,3050	163,4716	6,9087	11,6229	153,6062
ResNet50	11	BR LR	241,9622	282,5533	155,2452	6,8380	6,8488	166,5641
ResNet50	12	RakelD LR	241,1392	278,1919	158,7396	7,4740	9,5499	151,6026
ResNet18	8	BR LR	240,6531	277,7730	158,6728	6,8227	11,3149	162,7899
ResNet50	13	BR LR	240,5909	290,1687	145,1622	6,8148	6,3043	169,9970
ResNet50	10	BR LR	240,5013	282,8999	152,1502	6,8215	6,2076	158,8710
ResNet18	12	CC LR	239,3588	259,0768	183,9347	6,9443	23,0996	137,6598
ResNet18	2	BR LR	239,0618	272,2963	161,9080	6,8137	12,1863	153,4964
ResNet18	5	CC LR	238,7767	260,8821	178,9389	6,9958	18,1287	143,9791
ResNet18	11	BR LR	238,1627	269,0925	164,1061	6,9521	11,9760	146,0434
ResNet50	11	RakelD LR	238,1474	277,4457	153,6773	7,3734	7,7990	151,4070
ResNet18	10	BR LR	237,9248	277,4443	153,0424	6,8098	13,6417	152,7787
ResNet18	8	CC LR	237,8899	259,6552	178,7377	7,0756	21,9170	139,5582
ResNet18	15	CC LR	237,6578	257,3168	182,1383	7,0073	20,9539	131,3237

Continua en la página siguiente

Origen des- cripto- res	Cjto.	Método	AUC <i>Micro</i> $F_2$	AUC <i>Micro</i> <i>Recall</i>	AUC <i>Micro</i> <i>Preci- sion</i>	AUC <i>Ham- ming</i> <i>Score</i>	AUC EMR	AUC 5-MR
ResNet50	11	CC LR	237,5242	259,2066	178,3590	6,8387	22,5347	133,1408
ResNet50	15	BR LR	237,5172	288,6557	141,2636	6,7994	6,4530	173,1086
ResNet50	10	RakelD LR	237,2070	276,7406	152,5594	7,4702	7,7693	149,9713
ResNet18	4	BR LR	236,7273	269,9136	159,9726	7,0572	10,9606	146,8428
ResNet18	13	RakelD LR	236,1072	273,2472	154,2124	7,3919	11,7497	149,8892
ResNet18	2	CC LR	235,9559	256,0619	180,0390	6,8853	16,9459	134,3546
ResNet18	1	BR LR	235,6894	271,8011	155,1281	6,7823	7,9772	153,8868
ResNet18	15	RakelD LR	235,6590	276,4450	149,3026	7,3659	10,3356	152,8726
ResNet18	6	BR LR	235,5181	281,9329	143,8314	6,7430	9,2319	167,6142
ResNet18	14	RakelD LR	235,3875	279,6191	145,4911	7,3853	10,0375	159,1616
ResNet18	14	CC LR	235,2179	256,8552	176,3399	6,9109	23,2899	141,7512
ResNet18	3	CC LR	235,1926	263,7669	164,7532	6,8893	15,5755	143,8390
ResNet50	14	BR LR	235,0268	294,5459	132,6483	6,7088	5,6995	179,7124
ResNet50	1	BR LR	234,3102	279,5434	144,0932	6,7138	7,0970	160,1801
ResNet18	13	CC LR	233,7870	251,7533	182,1542	6,9505	26,9393	136,8888
ResNet18	11	RakelD LR	233,7220	271,3056	151,4723	7,3677	9,2826	142,8457
ResNet18	7	CC LR	233,7036	253,3608	178,6685	6,9392	24,3458	129,9039
ResNet18	3	BR LR	233,4550	282,4160	139,5644	6,7743	8,6556	157,4400
ResNet18	5	RakelD LR	233,4475	269,7220	152,8970	7,3984	10,0175	143,1309
ResNet50	12	CC LR	233,2248	254,2613	175,7771	6,9302	22,4026	131,8717
ResNet18	1	CC LR	233,1990	255,5523	173,2100	6,8768	19,3566	142,9551
ResNet50	13	RakelD LR	233,1968	278,5824	143,0168	7,3593	6,6055	149,3580
ResNet18	2	RakelD LR	233,1150	270,1232	151,8592	7,3410	8,3952	142,0081
ResNet18	7	BR LR	233,0546	264,3687	159,0474	6,8388	11,8035	135,8737
ResNet18	12	RakelD LR	232,9466	271,9881	149,2645	7,3809	9,3068	146,3224
ResNet50	1	CC LR	232,8062	257,3864	168,9272	6,7703	15,6879	139,1652
ResNet50	10	CC LR	232,1799	254,0696	173,0574	6,8854	18,4767	126,9832
ResNet50	8	BR LR	231,7319	276,5156	142,4887	6,8124	8,4420	153,2386
ResNet18	10	CC LR	231,5664	252,1826	174,8360	6,9738	22,3452	128,3253
ResNet18	4	RakelD LR	231,4747	266,9866	152,0267	7,3779	7,5270	137,6262
ResNet18	8	RakelD LR	231,4369	268,4791	150,4339	7,4245	9,8450	139,3938
ResNet18	4	CC LR	231,1337	252,7750	172,7306	7,0801	18,2111	125,9037
ResNet18	10	RakelD LR	229,6968	274,2824	140,5615	7,3470	9,2840	144,4893

Continua en la página siguiente

Origen des- cripto- res	Cjto.	Método	AUC <i>Micro</i> $F_2$	AUC <i>Micro</i> <i>Recall</i>	AUC <i>Micro</i> <i>Preci- sion</i>	AUC <i>Ham- ming</i> <i>Score</i>	AUC EMR	AUC 5-MR
ResNet18	11	CC LR	229,3510	246,9819	178,6644	6,9465	18,0785	124,8051
ResNet50	8	RakelD LR	229,3316	272,2490	142,2750	7,3367	7,5647	143,5688
ResNet18	1	RakelD LR	229,3223	269,1963	145,2376	7,2627	8,0957	139,5277
ResNet18	6	CC LR	228,9860	255,9972	161,6778	6,7713	13,6320	129,5618
ResNet50	15	RakelD LR	228,8175	277,6894	136,1777	7,1846	6,0973	152,1005
ResNet50	13	CC LR	228,0684	251,9179	166,0718	6,8780	16,2240	129,9031
ResNet50	14	RakelD LR	227,6297	283,9403	129,3120	7,3496	5,2923	159,0844
ResNet50	1	RakelD LR	227,5673	271,0454	140,3568	7,3529	6,3156	141,3511
ResNet18	7	RakelD LR	227,4799	261,2307	150,8196	7,3454	7,8996	131,0003
ResNet50	2	BR LR	227,3952	290,1901	124,6305	6,7506	6,4477	168,6589
ResNet18	9	BR LR	226,2958	285,6387	126,4425	6,6764	9,1115	169,4150
ResNet50	4	BR LR	225,9175	273,7200	134,9809	6,7567	4,9680	142,3958
ResNet50	8	CC LR	225,8778	252,2098	159,9603	6,8388	13,6367	130,1123
ResNet18	0	BR LR	225,2722	270,7146	137,0474	6,7280	8,9019	137,5695
ResNet50	15	CC LR	225,2027	247,6111	165,6416	6,8127	21,3925	116,5295
ResNet18	6	RakelD LR	224,6144	275,8325	130,5224	7,2402	5,6597	151,9993
ResNet50	5	BR LR	224,1049	272,4695	133,1109	6,7577	6,0733	144,5913
ResNet50	5	CC LR	223,2357	251,1589	155,3372	6,7742	13,8914	126,9459
ResNet50	4	RakelD LR	223,1676	266,7157	136,6595	7,3337	5,3514	132,8835
ResNet18	3	RakelD LR	221,4913	271,8105	128,7680	7,2713	6,2314	143,1968
ResNet50	4	CC LR	220,7963	248,6899	153,1465	6,7521	14,4357	133,0782
ResNet18	0	CC LR	220,5730	247,9083	153,9742	6,8408	11,7713	128,6107
ResNet50	14	CC LR	220,0139	244,7492	157,0786	6,7205	16,4703	119,0164
ResNet50	7	BR LR	218,7699	275,8725	122,0986	6,6638	5,9072	155,3608
ResNet50	6	BR LR	218,1495	281,8473	117,6734	6,7490	6,6013	153,9736
ResNet50	5	RakelD LR	217,3187	266,7415	126,4129	7,3239	5,0095	135,3313
ResNet18	9	RakelD LR	217,0020	275,0993	120,0127	7,1699	5,8464	143,2785
ResNet50	2	RakelD LR	216,0106	274,0001	119,3395	7,3491	4,9040	146,7578
ResNet18	0	RakelD LR	215,8151	260,5631	129,5796	7,3627	6,7457	129,4534
ResNet50	2	CC LR	215,5049	243,8800	147,8014	6,8262	12,2622	124,2015
ResNet18	9	CC LR	213,4201	241,2949	146,6649	6,7808	13,2101	109,6865
ResNet50	6	CC LR	212,8918	244,0931	141,6436	6,7808	9,7260	113,3516
ResNet50	7	RakelD LR	212,5634	263,6028	121,9768	7,2309	5,2599	128,1900

Continua en la página siguiente

Origen des- cripto- res	Cjto.	Método	AUC <i>Micro</i> $F_2$	AUC <i>Micro</i> <i>Recall</i>	AUC <i>Micro</i> <i>Preci- sion</i>	AUC <i>Ham- ming</i> <i>Score</i>	AUC EMR	AUC 5-MR
ResNet50	7	CC LR	212,1343	238,9538	146,9501	6,7218	12,1165	108,2800
ResNet50	6	RakelD LR	211,6148	268,7587	116,9718	7,3074	7,8672	138,1806
ResNet50	0	BR LR	205,6211	267,8715	109,2359	6,6164	8,0545	133,8566
ResNet50	0	RakelD LR	195,1486	253,9616	103,5730	7,1431	4,7135	117,4088
ResNet50	3	BR LR	193,6495	270,0499	94,2392	6,5894	3,6733	140,7733
ResNet50	0	CC LR	193,5079	225,7969	124,0629	6,6852	11,7957	98,4096
ResNet50	9	BR LR	189,8347	272,3710	89,3714	6,6081	3,3148	143,5417
ResNet50	9	RakelD LR	184,1067	258,0259	89,2431	7,2147	3,5265	121,1898
ResNet50	3	RakelD LR	182,5027	255,0046	88,1758	7,1814	2,8746	120,8858
ResNet50	9	CC LR	180,1034	210,4303	114,9041	6,6822	6,3390	81,1865
ResNet50	3	CC LR	174,0072	201,8643	112,6417	6,6179	8,2451	84,3278

# Anexo D

## Experimento 4

Tabla D.1: Métricas Experimento 4 - Descriptores AlexNet.

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
0	5	0,6386	0,7800	0,3916	0,3577	0,0980	0,6843
0	14	0,4395	0,4975	0,3077	0,2543	0,0078	0,2783
0	26	0,4426	0,5264	0,2754	0,2306	0,0000	0,2255
0	34	0,4337	0,5571	0,2382	0,2032	0,0000	0,2203
0	54	0,4208	0,6157	0,1894	0,1706	0,0000	0,2332
0	63	0,4131	0,6221	0,1775	0,1615	0,0000	0,2036
0	72	0,4064	0,6289	0,1685	0,1547	0,0000	0,2023
0	82	0,3892	0,6360	0,1537	0,1421	0,0000	0,2294
0	91	0,3847	0,6815	0,1405	0,1329	0,0000	0,2938
0	107	0,3739	0,6962	0,1312	0,1249	0,0000	0,3235
0	131	0,3603	0,7071	0,1218	0,1168	0,0000	0,3518
0	169	0,3276	0,7438	0,1012	0,0983	0,0000	0,4355
0	281	0,2698	0,7890	0,0743	0,0730	0,0000	0,5142
1	5	0,5029	0,5459	0,4289	0,3166	0,0799	0,4678
1	14	0,4269	0,4740	0,3241	0,2462	0,0013	0,2474
1	26	0,3554	0,3942	0,2636	0,1975	0,0000	0,1173
1	34	0,3441	0,3912	0,2443	0,1857	0,0000	0,0992
1	54	0,3566	0,4330	0,2132	0,1758	0,0000	0,1018
1	63	0,3425	0,4375	0,1862	0,1609	0,0000	0,0928
1	72	0,3544	0,4606	0,1900	0,1627	0,0000	0,1031
1	82	0,3696	0,4647	0,2092	0,1737	0,0000	0,0915
1	91	0,3533	0,4608	0,1876	0,1604	0,0000	0,0992
1	107	0,3521	0,4658	0,1806	0,1557	0,0000	0,0838

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
1	131	0,3438	0,5163	0,1479	0,1332	0,0000	0,1340
1	169	0,3518	0,5379	0,1485	0,1343	0,0000	0,1301
1	281	0,3190	0,6727	0,1033	0,0992	0,0000	0,3093
2	5	0,5895	0,6365	0,4608	0,3924	0,2178	0,4987
2	14	0,3620	0,3505	0,4252	0,2835	0,0670	0,1611
2	26	0,3461	0,3505	0,3450	0,2444	0,0284	0,1056
2	34	0,3572	0,4185	0,2361	0,1884	0,0013	0,1095
2	54	0,3657	0,4073	0,2598	0,2004	0,0000	0,0774
2	63	0,3468	0,4254	0,2050	0,1708	0,0000	0,0953
2	72	0,3477	0,4169	0,2103	0,1738	0,0000	0,0799
2	82	0,3301	0,4458	0,1736	0,1470	0,0000	0,1095
2	91	0,3253	0,4522	0,1542	0,1370	0,0000	0,0954
2	107	0,3497	0,4373	0,1950	0,1630	0,0000	0,0606
2	131	0,3330	0,4812	0,1500	0,1336	0,0000	0,1018
2	169	0,3445	0,5506	0,1385	0,1271	0,0000	0,1533
2	281	0,3116	0,6408	0,1024	0,0981	0,0000	0,2616
3	5	0,5741	0,6056	0,4763	0,3951	0,2578	0,4639
3	14	0,3439	0,3354	0,4038	0,2631	0,0528	0,1572
3	26	0,3593	0,3858	0,2931	0,2130	0,0013	0,1044
3	34	0,3864	0,4425	0,2651	0,2063	0,0000	0,1095
3	54	0,3654	0,4200	0,2449	0,1930	0,0000	0,0825
3	63	0,3598	0,4280	0,2217	0,1804	0,0000	0,0747
3	72	0,3604	0,4474	0,2120	0,1726	0,0000	0,0851
3	82	0,3300	0,4356	0,1732	0,1470	0,0000	0,0838
3	91	0,3604	0,4730	0,1861	0,1604	0,0000	0,0954
3	107	0,3586	0,4838	0,1790	0,1536	0,0000	0,0799
3	131	0,3528	0,5171	0,1561	0,1393	0,0000	0,1314
3	169	0,3598	0,6194	0,1350	0,1258	0,0000	0,2126
3	281	0,3035	0,7239	0,0924	0,0896	0,0000	0,3931
4	5	0,5946	0,6489	0,4522	0,3850	0,1894	0,5090
4	14	0,4038	0,4276	0,3810	0,2679	0,0348	0,2126
4	26	0,3660	0,3837	0,3395	0,2362	0,0155	0,1095
4	34	0,3912	0,4376	0,2758	0,2142	0,0000	0,1070
4	54	0,3774	0,4400	0,2424	0,1943	0,0000	0,0902
4	63	0,3699	0,4338	0,2361	0,1885	0,0000	0,0889
4	72	0,3780	0,4723	0,2115	0,1765	0,0000	0,0915

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
4	82	0,3664	0,4790	0,1926	0,1633	0,0000	0,0941
4	91	0,3498	0,4563	0,1814	0,1543	0,0000	0,0786
4	107	0,3534	0,4860	0,1703	0,1501	0,0000	0,0966
4	131	0,3600	0,5404	0,1543	0,1389	0,0000	0,1340
4	169	0,3427	0,5659	0,1333	0,1230	0,0000	0,1546
4	281	0,3167	0,6638	0,1028	0,0985	0,0000	0,2899
5	5	0,6045	0,6759	0,4373	0,3798	0,1559	0,5387
5	14	0,3912	0,3963	0,3870	0,2834	0,0477	0,2113
5	26	0,3556	0,3570	0,3549	0,2396	0,0077	0,1005
5	34	0,3696	0,3929	0,3037	0,2187	0,0000	0,0863
5	54	0,3781	0,4763	0,2138	0,1780	0,0000	0,0992
5	63	0,3779	0,4586	0,2224	0,1843	0,0000	0,0967
5	72	0,3579	0,4726	0,1906	0,1616	0,0000	0,1147
5	82	0,3662	0,4566	0,2066	0,1709	0,0000	0,0812
5	91	0,3538	0,4648	0,1867	0,1577	0,0000	0,0915
5	107	0,3775	0,5522	0,1680	0,1511	0,0000	0,1469
5	131	0,3653	0,5800	0,1475	0,1354	0,0000	0,1624
5	169	0,3609	0,5930	0,1415	0,1312	0,0000	0,1907
5	281	0,3224	0,6765	0,1046	0,1006	0,0000	0,3299
6	5	0,5489	0,5932	0,4631	0,3577	0,1753	0,4794
6	14	0,4129	0,4270	0,3812	0,2860	0,0528	0,2178
6	26	0,3692	0,3862	0,3188	0,2280	0,0000	0,0992
6	34	0,3785	0,4153	0,2819	0,2191	0,0026	0,1083
6	54	0,3679	0,4366	0,2270	0,1858	0,0000	0,0851
6	63	0,3671	0,4570	0,2072	0,1730	0,0000	0,0979
6	72	0,3507	0,4580	0,1861	0,1580	0,0000	0,1018
6	82	0,3746	0,4864	0,1962	0,1671	0,0000	0,0851
6	91	0,3528	0,4544	0,1890	0,1600	0,0000	0,0774
6	107	0,3607	0,4836	0,1804	0,1546	0,0000	0,0851
6	131	0,3593	0,5595	0,1483	0,1360	0,0000	0,1547
6	169	0,3524	0,5850	0,1372	0,1270	0,0000	0,2010
6	281	0,3203	0,6735	0,1039	0,0997	0,0000	0,3157
7	5	0,5140	0,5360	0,4529	0,3453	0,1830	0,4072
7	14	0,3806	0,3763	0,4062	0,2888	0,0644	0,1843
7	26	0,3563	0,3588	0,3470	0,2404	0,0077	0,0940
7	34	0,3449	0,4166	0,2475	0,1862	0,0026	0,1224

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
7	54	0,3340	0,3812	0,2272	0,1782	0,0000	0,0747
7	63	0,3474	0,4452	0,1908	0,1632	0,0000	0,1082
7	72	0,3668	0,4447	0,2197	0,1817	0,0000	0,0966
7	82	0,3634	0,4465	0,2120	0,1742	0,0000	0,0876
7	91	0,3445	0,4359	0,1909	0,1590	0,0000	0,0670
7	107	0,3654	0,5368	0,1641	0,1466	0,0000	0,1276
7	131	0,3495	0,5335	0,1475	0,1358	0,0000	0,1315
7	169	0,3451	0,5354	0,1427	0,1301	0,0000	0,1121
7	281	0,3237	0,6860	0,1042	0,1004	0,0000	0,3441
8	5	0,5877	0,6788	0,4219	0,3439	0,0876	0,5528
8	14	0,3748	0,3838	0,3528	0,2580	0,0310	0,2178
8	26	0,3867	0,4160	0,3145	0,2292	0,0052	0,1224
8	34	0,3918	0,4408	0,2725	0,2122	0,0000	0,1044
8	54	0,3641	0,4442	0,2151	0,1774	0,0000	0,0953
8	63	0,3688	0,4339	0,2306	0,1859	0,0000	0,0850
8	72	0,3797	0,4790	0,2127	0,1778	0,0000	0,0902
8	82	0,3408	0,4084	0,2068	0,1706	0,0000	0,0657
8	91	0,3564	0,5131	0,1696	0,1475	0,0000	0,1134
8	107	0,3682	0,5111	0,1760	0,1540	0,0000	0,0954
8	131	0,3565	0,5323	0,1552	0,1396	0,0000	0,1418
8	169	0,3584	0,5830	0,1416	0,1309	0,0000	0,1817
8	281	0,3166	0,7042	0,0994	0,0960	0,0000	0,3621
9	5	0,5813	0,6356	0,4480	0,3674	0,1520	0,4974
9	14	0,4069	0,4334	0,3291	0,2690	0,0284	0,2371
9	26	0,3838	0,4205	0,2987	0,2286	0,0064	0,1366
9	34	0,3770	0,4228	0,2656	0,2058	0,0013	0,1018
9	54	0,3803	0,4418	0,2495	0,1984	0,0000	0,0941
9	63	0,3746	0,4579	0,2175	0,1814	0,0000	0,1044
9	72	0,3647	0,4537	0,2078	0,1737	0,0000	0,0876
9	82	0,3458	0,4428	0,1878	0,1577	0,0000	0,0928
9	91	0,3533	0,4687	0,1780	0,1541	0,0000	0,0876
9	107	0,3503	0,4545	0,1830	0,1555	0,0000	0,0657
9	131	0,3648	0,5624	0,1530	0,1390	0,0000	0,1520
9	169	0,3607	0,6150	0,1361	0,1273	0,0000	0,2268
9	281	0,3007	0,6978	0,0928	0,0904	0,0000	0,3608
10	5	0,4361	0,4870	0,3478	0,2796	0,0580	0,4523

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
10	14	0,3345	0,3211	0,4091	0,2585	0,0657	0,1533
10	26	0,2903	0,2915	0,3051	0,2008	0,0103	0,0838
10	34	0,3571	0,4150	0,2373	0,1874	0,0000	0,1082
10	54	0,3337	0,3866	0,2176	0,1758	0,0000	0,0850
10	63	0,2811	0,3802	0,1481	0,1256	0,0000	0,0979
10	72	0,3293	0,4646	0,1615	0,1409	0,0000	0,1289
10	82	0,3323	0,4285	0,1823	0,1527	0,0000	0,0954
10	91	0,3540	0,4457	0,1998	0,1660	0,0000	0,0838
10	107	0,3520	0,4728	0,1744	0,1512	0,0000	0,0799
10	131	0,3481	0,4983	0,1604	0,1413	0,0000	0,1095
10	169	0,3480	0,5457	0,1431	0,1303	0,0000	0,1559
10	281	0,3220	0,6581	0,1060	0,1016	0,0000	0,2899
11	5	0,6101	0,7000	0,4151	0,3644	0,1044	0,5722
11	14	0,4032	0,4117	0,4044	0,2954	0,0670	0,2049
11	26	0,3152	0,3215	0,2997	0,2056	0,0103	0,0928
11	34	0,3127	0,3542	0,2194	0,1708	0,0000	0,0812
11	54	0,2980	0,3259	0,2243	0,1674	0,0000	0,0515
11	63	0,3603	0,4269	0,2288	0,1833	0,0000	0,0851
11	72	0,3320	0,3856	0,2278	0,1784	0,0000	0,0721
11	82	0,3236	0,4351	0,1689	0,1444	0,0000	0,1096
11	91	0,3539	0,4562	0,1877	0,1598	0,0000	0,0953
11	107	0,3441	0,4436	0,1829	0,1555	0,0000	0,0773
11	131	0,3509	0,4875	0,1660	0,1456	0,0000	0,0979
11	169	0,3339	0,5306	0,1360	0,1242	0,0000	0,1263
11	281	0,3190	0,6404	0,1064	0,1017	0,0000	0,2706
12	5	0,5758	0,6465	0,4368	0,3480	0,0954	0,5206
12	14	0,2240	0,2044	0,4294	0,1936	0,0631	0,1134
12	26	0,2968	0,3046	0,2903	0,1954	0,0142	0,0786
12	34	0,3371	0,3660	0,2666	0,2013	0,0013	0,0851
12	54	0,2710	0,3505	0,1485	0,1262	0,0000	0,0838
12	63	0,3404	0,4287	0,1979	0,1645	0,0000	0,1082
12	72	0,3220	0,4035	0,1808	0,1518	0,0000	0,0838
12	82	0,3505	0,3996	0,2425	0,1867	0,0000	0,0606
12	91	0,3226	0,4130	0,1768	0,1501	0,0000	0,0786
12	107	0,3516	0,4483	0,1893	0,1595	0,0000	0,0619
12	131	0,3395	0,4545	0,1738	0,1488	0,0000	0,0825

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
12	169	0,3516	0,5552	0,1446	0,1320	0,0000	0,1611
12	281	0,3306	0,6297	0,1144	0,1083	0,0000	0,2577
13	5	0,5468	0,6003	0,4269	0,3456	0,1327	0,4910
13	14	0,3171	0,3057	0,3821	0,2415	0,0580	0,1431
13	26	0,2092	0,2228	0,1848	0,1178	0,0000	0,0593
13	34	0,2266	0,2441	0,1897	0,1298	0,0000	0,0528
13	54	0,3402	0,3893	0,2319	0,1809	0,0000	0,0863
13	63	0,3296	0,3643	0,2416	0,1828	0,0000	0,0567
13	72	0,3437	0,4018	0,2301	0,1800	0,0000	0,0735
13	82	0,3465	0,4343	0,1960	0,1636	0,0000	0,0979
13	91	0,3369	0,4085	0,2022	0,1666	0,0000	0,0631
13	107	0,3596	0,4632	0,1918	0,1623	0,0000	0,0696
13	131	0,3502	0,4833	0,1725	0,1495	0,0000	0,0979
13	169	0,3429	0,5452	0,1385	0,1269	0,0000	0,1495
13	281	0,3316	0,6214	0,1163	0,1101	0,0000	0,2423
14	5	0,5923	0,6472	0,4536	0,3870	0,2010	0,5077
14	14	0,1859	0,1715	0,3356	0,1640	0,0567	0,1044
14	26	0,2655	0,2728	0,2988	0,1777	0,0103	0,0760
14	34	0,2652	0,2788	0,2434	0,1643	0,0026	0,0632
14	54	0,3240	0,3641	0,2591	0,1879	0,0013	0,0709
14	63	0,2742	0,3340	0,1712	0,1376	0,0000	0,0786
14	72	0,3288	0,3992	0,2036	0,1633	0,0000	0,0851
14	82	0,3419	0,3937	0,2291	0,1798	0,0000	0,0670
14	91	0,3381	0,4189	0,1944	0,1606	0,0000	0,0631
14	107	0,3428	0,3923	0,2300	0,1793	0,0000	0,0348
14	131	0,3664	0,5112	0,1721	0,1520	0,0000	0,1095
14	169	0,3512	0,5701	0,1388	0,1281	0,0000	0,1740
14	281	0,3189	0,5511	0,1243	0,1138	0,0000	0,1985
15	5	0,5291	0,5487	0,4830	0,3636	0,2075	0,4098
15	14	0,2771	0,2703	0,3160	0,2056	0,0374	0,1392
15	26	0,2628	0,2439	0,3818	0,2178	0,0348	0,0708
15	34	0,3604	0,4049	0,2602	0,1989	0,0000	0,0992
15	54	0,3110	0,3399	0,2390	0,1810	0,0000	0,0554
15	63	0,3210	0,3885	0,1966	0,1624	0,0000	0,0928
15	72	0,3258	0,3965	0,1920	0,1602	0,0000	0,0799
15	82	0,3479	0,4306	0,1969	0,1646	0,0000	0,0799

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
15	91	0,3026	0,3897	0,1734	0,1412	0,0000	0,0670
15	107	0,3398	0,4218	0,1924	0,1582	0,0000	0,0490
15	131	0,3518	0,4449	0,1957	0,1623	0,0000	0,0683
15	169	0,3440	0,5098	0,1545	0,1363	0,0000	0,1186
15	281	0,3310	0,6277	0,1152	0,1090	0,0000	0,2461

Tabla D.2: Métricas Experimento 4 - Descriptores Imágenes *flatten*.

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
0	5	0,6386	0,7800	0,3916	0,3577	0,0980	0,6843
0	14	0,4395	0,4975	0,3077	0,2543	0,0078	0,2783
0	26	0,4426	0,5264	0,2754	0,2306	0,0000	0,2255
0	34	0,4337	0,5571	0,2382	0,2032	0,0000	0,2203
0	54	0,4208	0,6157	0,1894	0,1706	0,0000	0,2332
0	63	0,4131	0,6221	0,1775	0,1615	0,0000	0,2036
0	72	0,4064	0,6289	0,1685	0,1547	0,0000	0,2023
0	82	0,3892	0,6360	0,1537	0,1421	0,0000	0,2294
0	91	0,3847	0,6815	0,1405	0,1329	0,0000	0,2938
0	107	0,3739	0,6962	0,1312	0,1249	0,0000	0,3235
0	131	0,3603	0,7071	0,1218	0,1168	0,0000	0,3518
0	169	0,3276	0,7438	0,1012	0,0983	0,0000	0,4355
0	281	0,2698	0,7890	0,0743	0,0730	0,0000	0,5142
1	5	0,5029	0,5459	0,4289	0,3166	0,0799	0,4678
1	14	0,4269	0,4740	0,3241	0,2462	0,0013	0,2474
1	26	0,3554	0,3942	0,2636	0,1975	0,0000	0,1173
1	34	0,3441	0,3912	0,2443	0,1857	0,0000	0,0992
1	54	0,3566	0,4330	0,2132	0,1758	0,0000	0,1018
1	63	0,3425	0,4375	0,1862	0,1609	0,0000	0,0928
1	72	0,3544	0,4606	0,1900	0,1627	0,0000	0,1031
1	82	0,3696	0,4647	0,2092	0,1737	0,0000	0,0915
1	91	0,3533	0,4608	0,1876	0,1604	0,0000	0,0992
1	107	0,3521	0,4658	0,1806	0,1557	0,0000	0,0838
1	131	0,3438	0,5163	0,1479	0,1332	0,0000	0,1340
1	169	0,3518	0,5379	0,1485	0,1343	0,0000	0,1301
1	281	0,3190	0,6727	0,1033	0,0992	0,0000	0,3093

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
2	5	0,5895	0,6365	0,4608	0,3924	0,2178	0,4987
2	14	0,3620	0,3505	0,4252	0,2835	0,0670	0,1611
2	26	0,3461	0,3505	0,3450	0,2444	0,0284	0,1056
2	34	0,3572	0,4185	0,2361	0,1884	0,0013	0,1095
2	54	0,3657	0,4073	0,2598	0,2004	0,0000	0,0774
2	63	0,3468	0,4254	0,2050	0,1708	0,0000	0,0953
2	72	0,3477	0,4169	0,2103	0,1738	0,0000	0,0799
2	82	0,3301	0,4458	0,1736	0,1470	0,0000	0,1095
2	91	0,3253	0,4522	0,1542	0,1370	0,0000	0,0954
2	107	0,3497	0,4373	0,1950	0,1630	0,0000	0,0606
2	131	0,3330	0,4812	0,1500	0,1336	0,0000	0,1018
2	169	0,3445	0,5506	0,1385	0,1271	0,0000	0,1533
2	281	0,3116	0,6408	0,1024	0,0981	0,0000	0,2616
3	5	0,5741	0,6056	0,4763	0,3951	0,2578	0,4639
3	14	0,3439	0,3354	0,4038	0,2631	0,0528	0,1572
3	26	0,3593	0,3858	0,2931	0,2130	0,0013	0,1044
3	34	0,3864	0,4425	0,2651	0,2063	0,0000	0,1095
3	54	0,3654	0,4200	0,2449	0,1930	0,0000	0,0825
3	63	0,3598	0,4280	0,2217	0,1804	0,0000	0,0747
3	72	0,3604	0,4474	0,2120	0,1726	0,0000	0,0851
3	82	0,3300	0,4356	0,1732	0,1470	0,0000	0,0838
3	91	0,3604	0,4730	0,1861	0,1604	0,0000	0,0954
3	107	0,3586	0,4838	0,1790	0,1536	0,0000	0,0799
3	131	0,3528	0,5171	0,1561	0,1393	0,0000	0,1314
3	169	0,3598	0,6194	0,1350	0,1258	0,0000	0,2126
3	281	0,3035	0,7239	0,0924	0,0896	0,0000	0,3931
4	5	0,5946	0,6489	0,4522	0,3850	0,1894	0,5090
4	14	0,4038	0,4276	0,3810	0,2679	0,0348	0,2126
4	26	0,3660	0,3837	0,3395	0,2362	0,0155	0,1095
4	34	0,3912	0,4376	0,2758	0,2142	0,0000	0,1070
4	54	0,3774	0,4400	0,2424	0,1943	0,0000	0,0902
4	63	0,3699	0,4338	0,2361	0,1885	0,0000	0,0889
4	72	0,3780	0,4723	0,2115	0,1765	0,0000	0,0915
4	82	0,3664	0,4790	0,1926	0,1633	0,0000	0,0941
4	91	0,3498	0,4563	0,1814	0,1543	0,0000	0,0786
4	107	0,3534	0,4860	0,1703	0,1501	0,0000	0,0966

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
4	131	0,3600	0,5404	0,1543	0,1389	0,0000	0,1340
4	169	0,3427	0,5659	0,1333	0,1230	0,0000	0,1546
4	281	0,3167	0,6638	0,1028	0,0985	0,0000	0,2899
5	5	0,6045	0,6759	0,4373	0,3798	0,1559	0,5387
5	14	0,3912	0,3963	0,3870	0,2834	0,0477	0,2113
5	26	0,3556	0,3570	0,3549	0,2396	0,0077	0,1005
5	34	0,3696	0,3929	0,3037	0,2187	0,0000	0,0863
5	54	0,3781	0,4763	0,2138	0,1780	0,0000	0,0992
5	63	0,3779	0,4586	0,2224	0,1843	0,0000	0,0967
5	72	0,3579	0,4726	0,1906	0,1616	0,0000	0,1147
5	82	0,3662	0,4566	0,2066	0,1709	0,0000	0,0812
5	91	0,3538	0,4648	0,1867	0,1577	0,0000	0,0915
5	107	0,3775	0,5522	0,1680	0,1511	0,0000	0,1469
5	131	0,3653	0,5800	0,1475	0,1354	0,0000	0,1624
5	169	0,3609	0,5930	0,1415	0,1312	0,0000	0,1907
5	281	0,3224	0,6765	0,1046	0,1006	0,0000	0,3299
6	5	0,5489	0,5932	0,4631	0,3577	0,1753	0,4794
6	14	0,4129	0,4270	0,3812	0,2860	0,0528	0,2178
6	26	0,3692	0,3862	0,3188	0,2280	0,0000	0,0992
6	34	0,3785	0,4153	0,2819	0,2191	0,0026	0,1083
6	54	0,3679	0,4366	0,2270	0,1858	0,0000	0,0851
6	63	0,3671	0,4570	0,2072	0,1730	0,0000	0,0979
6	72	0,3507	0,4580	0,1861	0,1580	0,0000	0,1018
6	82	0,3746	0,4864	0,1962	0,1671	0,0000	0,0851
6	91	0,3528	0,4544	0,1890	0,1600	0,0000	0,0774
6	107	0,3607	0,4836	0,1804	0,1546	0,0000	0,0851
6	131	0,3593	0,5595	0,1483	0,1360	0,0000	0,1547
6	169	0,3524	0,5850	0,1372	0,1270	0,0000	0,2010
6	281	0,3203	0,6735	0,1039	0,0997	0,0000	0,3157
7	5	0,5140	0,5360	0,4529	0,3453	0,1830	0,4072
7	14	0,3806	0,3763	0,4062	0,2888	0,0644	0,1843
7	26	0,3563	0,3588	0,3470	0,2404	0,0077	0,0940
7	34	0,3449	0,4166	0,2475	0,1862	0,0026	0,1224
7	54	0,3340	0,3812	0,2272	0,1782	0,0000	0,0747
7	63	0,3474	0,4452	0,1908	0,1632	0,0000	0,1082
7	72	0,3668	0,4447	0,2197	0,1817	0,0000	0,0966

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
7	82	0,3634	0,4465	0,2120	0,1742	0,0000	0,0876
7	91	0,3445	0,4359	0,1909	0,1590	0,0000	0,0670
7	107	0,3654	0,5368	0,1641	0,1466	0,0000	0,1276
7	131	0,3495	0,5335	0,1475	0,1358	0,0000	0,1315
7	169	0,3451	0,5354	0,1427	0,1301	0,0000	0,1121
7	281	0,3237	0,6860	0,1042	0,1004	0,0000	0,3441
8	5	0,5877	0,6788	0,4219	0,3439	0,0876	0,5528
8	14	0,3748	0,3838	0,3528	0,2580	0,0310	0,2178
8	26	0,3867	0,4160	0,3145	0,2292	0,0052	0,1224
8	34	0,3918	0,4408	0,2725	0,2122	0,0000	0,1044
8	54	0,3641	0,4442	0,2151	0,1774	0,0000	0,0953
8	63	0,3688	0,4339	0,2306	0,1859	0,0000	0,0850
8	72	0,3797	0,4790	0,2127	0,1778	0,0000	0,0902
8	82	0,3408	0,4084	0,2068	0,1706	0,0000	0,0657
8	91	0,3564	0,5131	0,1696	0,1475	0,0000	0,1134
8	107	0,3682	0,5111	0,1760	0,1540	0,0000	0,0954
8	131	0,3565	0,5323	0,1552	0,1396	0,0000	0,1418
8	169	0,3584	0,5830	0,1416	0,1309	0,0000	0,1817
8	281	0,3166	0,7042	0,0994	0,0960	0,0000	0,3621
9	5	0,5813	0,6356	0,4480	0,3674	0,1520	0,4974
9	14	0,4069	0,4334	0,3291	0,2690	0,0284	0,2371
9	26	0,3838	0,4205	0,2987	0,2286	0,0064	0,1366
9	34	0,3770	0,4228	0,2656	0,2058	0,0013	0,1018
9	54	0,3803	0,4418	0,2495	0,1984	0,0000	0,0941
9	63	0,3746	0,4579	0,2175	0,1814	0,0000	0,1044
9	72	0,3647	0,4537	0,2078	0,1737	0,0000	0,0876
9	82	0,3458	0,4428	0,1878	0,1577	0,0000	0,0928
9	91	0,3533	0,4687	0,1780	0,1541	0,0000	0,0876
9	107	0,3503	0,4545	0,1830	0,1555	0,0000	0,0657
9	131	0,3648	0,5624	0,1530	0,1390	0,0000	0,1520
9	169	0,3607	0,6150	0,1361	0,1273	0,0000	0,2268
9	281	0,3007	0,6978	0,0928	0,0904	0,0000	0,3608
10	5	0,4361	0,4870	0,3478	0,2796	0,0580	0,4523
10	14	0,3345	0,3211	0,4091	0,2585	0,0657	0,1533
10	26	0,2903	0,2915	0,3051	0,2008	0,0103	0,0838
10	34	0,3571	0,4150	0,2373	0,1874	0,0000	0,1082

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
10	54	0,3337	0,3866	0,2176	0,1758	0,0000	0,0850
10	63	0,2811	0,3802	0,1481	0,1256	0,0000	0,0979
10	72	0,3293	0,4646	0,1615	0,1409	0,0000	0,1289
10	82	0,3323	0,4285	0,1823	0,1527	0,0000	0,0954
10	91	0,3540	0,4457	0,1998	0,1660	0,0000	0,0838
10	107	0,3520	0,4728	0,1744	0,1512	0,0000	0,0799
10	131	0,3481	0,4983	0,1604	0,1413	0,0000	0,1095
10	169	0,3480	0,5457	0,1431	0,1303	0,0000	0,1559
10	281	0,3220	0,6581	0,1060	0,1016	0,0000	0,2899
11	5	0,6101	0,7000	0,4151	0,3644	0,1044	0,5722
11	14	0,4032	0,4117	0,4044	0,2954	0,0670	0,2049
11	26	0,3152	0,3215	0,2997	0,2056	0,0103	0,0928
11	34	0,3127	0,3542	0,2194	0,1708	0,0000	0,0812
11	54	0,2980	0,3259	0,2243	0,1674	0,0000	0,0515
11	63	0,3603	0,4269	0,2288	0,1833	0,0000	0,0851
11	72	0,3320	0,3856	0,2278	0,1784	0,0000	0,0721
11	82	0,3236	0,4351	0,1689	0,1444	0,0000	0,1096
11	91	0,3539	0,4562	0,1877	0,1598	0,0000	0,0953
11	107	0,3441	0,4436	0,1829	0,1555	0,0000	0,0773
11	131	0,3509	0,4875	0,1660	0,1456	0,0000	0,0979
11	169	0,3339	0,5306	0,1360	0,1242	0,0000	0,1263
11	281	0,3190	0,6404	0,1064	0,1017	0,0000	0,2706
12	5	0,5758	0,6465	0,4368	0,3480	0,0954	0,5206
12	14	0,2240	0,2044	0,4294	0,1936	0,0631	0,1134
12	26	0,2968	0,3046	0,2903	0,1954	0,0142	0,0786
12	34	0,3371	0,3660	0,2666	0,2013	0,0013	0,0851
12	54	0,2710	0,3505	0,1485	0,1262	0,0000	0,0838
12	63	0,3404	0,4287	0,1979	0,1645	0,0000	0,1082
12	72	0,3220	0,4035	0,1808	0,1518	0,0000	0,0838
12	82	0,3505	0,3996	0,2425	0,1867	0,0000	0,0606
12	91	0,3226	0,4130	0,1768	0,1501	0,0000	0,0786
12	107	0,3516	0,4483	0,1893	0,1595	0,0000	0,0619
12	131	0,3395	0,4545	0,1738	0,1488	0,0000	0,0825
12	169	0,3516	0,5552	0,1446	0,1320	0,0000	0,1611
12	281	0,3306	0,6297	0,1144	0,1083	0,0000	0,2577
13	5	0,5468	0,6003	0,4269	0,3456	0,1327	0,4910

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
13	14	0,3171	0,3057	0,3821	0,2415	0,0580	0,1431
13	26	0,2092	0,2228	0,1848	0,1178	0,0000	0,0593
13	34	0,2266	0,2441	0,1897	0,1298	0,0000	0,0528
13	54	0,3402	0,3893	0,2319	0,1809	0,0000	0,0863
13	63	0,3296	0,3643	0,2416	0,1828	0,0000	0,0567
13	72	0,3437	0,4018	0,2301	0,1800	0,0000	0,0735
13	82	0,3465	0,4343	0,1960	0,1636	0,0000	0,0979
13	91	0,3369	0,4085	0,2022	0,1666	0,0000	0,0631
13	107	0,3596	0,4632	0,1918	0,1623	0,0000	0,0696
13	131	0,3502	0,4833	0,1725	0,1495	0,0000	0,0979
13	169	0,3429	0,5452	0,1385	0,1269	0,0000	0,1495
13	281	0,3316	0,6214	0,1163	0,1101	0,0000	0,2423
14	5	0,5923	0,6472	0,4536	0,3870	0,2010	0,5077
14	14	0,1859	0,1715	0,3356	0,1640	0,0567	0,1044
14	26	0,2655	0,2728	0,2988	0,1777	0,0103	0,0760
14	34	0,2652	0,2788	0,2434	0,1643	0,0026	0,0632
14	54	0,3240	0,3641	0,2591	0,1879	0,0013	0,0709
14	63	0,2742	0,3340	0,1712	0,1376	0,0000	0,0786
14	72	0,3288	0,3992	0,2036	0,1633	0,0000	0,0851
14	82	0,3419	0,3937	0,2291	0,1798	0,0000	0,0670
14	91	0,3381	0,4189	0,1944	0,1606	0,0000	0,0631
14	107	0,3428	0,3923	0,2300	0,1793	0,0000	0,0348
14	131	0,3664	0,5112	0,1721	0,1520	0,0000	0,1095
14	169	0,3512	0,5701	0,1388	0,1281	0,0000	0,1740
14	281	0,3189	0,5511	0,1243	0,1138	0,0000	0,1985
15	5	0,5291	0,5487	0,4830	0,3636	0,2075	0,4098
15	14	0,2771	0,2703	0,3160	0,2056	0,0374	0,1392
15	26	0,2628	0,2439	0,3818	0,2178	0,0348	0,0708
15	34	0,3604	0,4049	0,2602	0,1989	0,0000	0,0992
15	54	0,3110	0,3399	0,2390	0,1810	0,0000	0,0554
15	63	0,3210	0,3885	0,1966	0,1624	0,0000	0,0928
15	72	0,3258	0,3965	0,1920	0,1602	0,0000	0,0799
15	82	0,3479	0,4306	0,1969	0,1646	0,0000	0,0799
15	91	0,3026	0,3897	0,1734	0,1412	0,0000	0,0670
15	107	0,3398	0,4218	0,1924	0,1582	0,0000	0,0490
15	131	0,3518	0,4449	0,1957	0,1623	0,0000	0,0683

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
15	169	0,3440	0,5098	0,1545	0,1363	0,0000	0,1186
15	281	0,3310	0,6277	0,1152	0,1090	0,0000	0,2461

# Anexo E

## Experimento 5

Tabla E.1: Métricas Experimento 5.

Conjunto	Tto.	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	EMR	5-MR
0	sin balance	0,2103	0,1790	0,7239	0,1756	0,0026	0,0039
0	con balance	0,5605	0,7412	0,2857	0,2782	0,0013	0,4755
1	sin balance	0,2078	0,1751	0,8422	0,1756	0,0052	0,0052
1	con balance	0,5604	0,7405	0,2867	0,2800	0,0000	0,4781
2	sin balance	0,1775	0,1492	0,8180	0,1456	0,0039	0,0052
2	con balance	0,5854	0,7048	0,3506	0,3277	0,0052	0,4304
3	sin balance	0,1923	0,1629	0,7645	0,1585	0,0039	0,0052
3	con balance	0,5517	0,7575	0,2722	0,2652	0,0026	0,5271
4	sin balance	0,1779	0,1491	0,8437	0,1476	0,0052	0,0052
4	con balance	0,5682	0,7002	0,3270	0,3063	0,0000	0,4214
5	sin balance	0,1843	0,1541	0,8694	0,1526	0,0039	0,0039
5	con balance	0,5795	0,6849	0,3598	0,3312	0,0090	0,4072
6	sin balance	0,1717	0,1434	0,8462	0,1398	0,0039	0,0039
6	con balance	0,5708	0,7363	0,3016	0,2902	0,0000	0,4819
7	sin balance	0,1695	0,1414	0,8368	0,1372	0,0026	0,0026
7	con balance	0,5781	0,6907	0,3510	0,3256	0,0013	0,4111
8	sin balance	0,1910	0,1607	0,8131	0,1567	0,0039	0,0039
8	con balance	0,5844	0,6981	0,3557	0,3299	0,0052	0,4111
9	sin balance	0,1791	0,1513	0,6923	0,1470	0,0026	0,0026
9	con balance	0,5090	0,7534	0,2233	0,2199	0,0000	0,4949
10	sin balance	0,1792	0,1505	0,8251	0,1463	0,0039	0,0052
10	con balance	0,5853	0,7134	0,3411	0,3220	0,0052	0,4356

Continua en la página siguiente

<b>Conjunto</b>	<b>Tto.</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
11	sin balance	0,1696	0,1414	0,8596	0,1395	0,0039	0,0039
11	con balance	0,2948	0,3568	0,1739	0,1657	0,0052	0,2242
12	sin balance	0,1730	0,1444	0,8555	0,1412	0,0039	0,0039
12	con balance	0,5768	0,7196	0,3230	0,3072	0,0052	0,4536
13	sin balance	0,1935	0,1637	0,7152	0,1598	0,0039	0,0039
13	con balance	0,5909	0,7146	0,3519	0,3295	0,0052	0,4330
14	sin balance	0,1629	0,1356	0,8500	0,1315	0,0039	0,0039
14	con balance	0,5840	0,7517	0,3094	0,3036	0,0013	0,5026
15	sin balance	0,1478	0,1229	0,8358	0,1185	0,0026	0,0026
15	con balance	0,5834	0,7696	0,2967	0,2943	0,0013	0,5180

# Anexo F

## Experimento 6

Tabla F.1: Métricas Experimento 6.

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<i>EMR</i>	<i>5-MR</i>
0	5	0,8258	0,8546	0,7308	0,6711	0,5206	0,8157
8	5	0,8154	0,8354	0,7566	0,6936	0,5489	0,8054
13	5	0,8217	0,8314	0,7862	0,7048	0,5954	0,7861
0	14	0,7614	0,8063	0,6231	0,5724	0,2062	0,6250
8	14	0,7474	0,7849	0,6333	0,5694	0,2217	0,5941
13	14	0,7690	0,7980	0,6720	0,6128	0,2784	0,6108
0	26	0,6748	0,7658	0,4617	0,4246	0,0374	0,5193
8	26	0,6863	0,7370	0,5389	0,4735	0,0824	0,4858
13	26	0,7168	0,7674	0,5677	0,5103	0,1005	0,5232
0	34	0,6455	0,7692	0,3938	0,3765	0,0310	0,5180
8	34	0,6676	0,7162	0,5274	0,4594	0,0657	0,4253
13	34	0,6835	0,7485	0,5077	0,4580	0,0567	0,4961
0	54	0,6058	0,7848	0,3240	0,3128	0,0052	0,5515
8	54	0,6330	0,7066	0,4486	0,4050	0,0399	0,4201
13	54	0,6507	0,7390	0,4409	0,4061	0,0296	0,4755
0	63	0,5914	0,7655	0,3124	0,3107	0,0103	0,5206
8	63	0,6267	0,7149	0,4199	0,3850	0,0155	0,4407
13	63	0,6395	0,7366	0,4189	0,3892	0,0193	0,4794
0	72	0,5946	0,7440	0,3312	0,3189	0,0026	0,4845
8	72	0,6106	0,6975	0,4078	0,3718	0,0155	0,4227
13	72	0,6302	0,7344	0,4024	0,3770	0,0142	0,4755
0	82	0,5753	0,7554	0,2960	0,2892	0,0013	0,5039
8	82	0,6064	0,7217	0,3723	0,3475	0,0065	0,4536

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
13	82	0,6144	0,7369	0,3696	0,3508	0,0078	0,4858
0	91	0,5631	0,7568	0,2853	0,2764	0,0026	0,5000
8	91	0,5806	0,7082	0,3385	0,3182	0,0078	0,4407
13	91	0,6084	0,7248	0,3708	0,3507	0,0090	0,4742
0	107	0,5403	0,7496	0,2560	0,2505	0,0000	0,4910
8	107	0,5784	0,7156	0,3288	0,3124	0,0000	0,4265
13	107	0,5840	0,7517	0,3094	0,3036	0,0013	0,5026
0	131	0,4294	0,8124	0,1780	0,1777	0,0000	0,6057
8	131	0,5453	0,7139	0,2850	0,2718	0,0013	0,4330
13	131	0,5598	0,7190	0,2971	0,2858	0,0013	0,4562
0	169	0,4339	0,7927	0,1666	0,1688	0,0000	0,5837
8	169	0,4989	0,7168	0,2261	0,2198	0,0000	0,4497
13	169	0,5279	0,7234	0,2547	0,2482	0,0000	0,4729
0	281	0,1726	0,8938	0,0462	0,0493	0,0000	0,7873
8	281	0,3397	0,7818	0,1108	0,1132	0,0000	0,5618
13	281	0,4321	0,7417	0,1620	0,1632	0,0000	0,4974

# Anexo G

## Experimento 7

Tabla G.1: Métricas Experimento 7.

<b>Etiquetas</b>	<b>Umbral</b>	<i>Micro F<sub>2</sub></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
5	0,25	0,8319	0,8525	0,7593	0,7175	0,5812	0,8183
5	0,5	0,8217	0,8314	0,7862	0,7048	0,5954	0,7861
5	0,75	0,7978	0,7909	0,8284	0,7332	0,6134	0,7590
14	0,25	0,7744	0,8267	0,6186	0,5787	0,2126	0,6611
14	0,5	0,7690	0,7980	0,6720	0,6128	0,2784	0,6108
14	0,75	0,7529	0,7644	0,7106	0,6135	0,2835	0,5786
26	0,25	0,7164	0,8041	0,4998	0,4712	0,0593	0,5915
26	0,5	0,7168	0,7674	0,5677	0,5103	0,1005	0,5232
26	0,75	0,6895	0,7031	0,6406	0,5261	0,1482	0,4484
34	0,25	0,6833	0,7950	0,4374	0,4193	0,0322	0,5567
34	0,5	0,6835	0,7485	0,5077	0,4580	0,0567	0,4961
34	0,75	0,6703	0,7127	0,5569	0,4678	0,0696	0,4330
54	0,25	0,6464	0,7786	0,3856	0,3728	0,0180	0,5387
54	0,5	0,6507	0,7390	0,4409	0,4061	0,0296	0,4755
54	0,75	0,6380	0,6773	0,5178	0,4403	0,0567	0,3956
63	0,25	0,6320	0,7734	0,3655	0,3560	0,0103	0,5335
63	0,5	0,6395	0,7366	0,4189	0,3892	0,0193	0,4794
63	0,75	0,6310	0,6780	0,4940	0,4272	0,0412	0,3788
72	0,25	0,6077	0,7730	0,3282	0,3258	0,0078	0,5477
72	0,5	0,6302	0,7344	0,4024	0,3770	0,0142	0,4755
72	0,75	0,6207	0,6644	0,4917	0,4214	0,0374	0,3737
82	0,25	0,6013	0,7755	0,3169	0,3136	0,0039	0,5515

Continua en la página siguiente

<b>Etiquetas</b>	<b>Umbral</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
82	0,5	0,6144	0,7369	0,3696	0,3508	0,0078	0,4858
82	0,75	0,6196	0,6807	0,4571	0,4023	0,0206	0,4085
91	0,25	0,5840	0,7678	0,2991	0,2947	0,0039	0,5348
91	0,5	0,6084	0,7248	0,3708	0,3507	0,0090	0,4742
91	0,75	0,6051	0,6723	0,4346	0,3840	0,0207	0,3931
107	0,25	0,5692	0,7605	0,2845	0,2818	0,0000	0,5219
107	0,5	0,5840	0,7517	0,3094	0,3036	0,0013	0,5026
107	0,75	0,5883	0,6676	0,3996	0,3574	0,0090	0,3750
131	0,25	0,5442	0,7554	0,2577	0,2574	0,0039	0,5051
131	0,5	0,5598	0,7190	0,2971	0,2858	0,0013	0,4562
131	0,75	0,5780	0,6702	0,3742	0,3402	0,0129	0,3879
169	0,25	0,4990	0,7664	0,2085	0,2100	0,0000	0,5399
169	0,5	0,5279	0,7234	0,2547	0,2482	0,0000	0,4729
169	0,75	0,5364	0,6694	0,2998	0,2788	0,0000	0,3917
281	0,25	0,4106	0,7494	0,1468	0,1482	0,0000	0,5077
281	0,5	0,4321	0,7417	0,1620	0,1632	0,0000	0,4974
281	0,75	0,4711	0,6734	0,2163	0,2085	0,0000	0,3931

# Anexo H

## Experimento 9

Tabla H.1: Métricas Experimento 9.

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
0	5	0,5022	0,4905	0,5903	0,4657	0,3544	0,4601
3	5	0,5371	0,5145	0,6656	0,4894	0,3879	0,4871
15	5	0,0000	0,0000	0,0000	0,2010	0,2010	0,2010
0	14	0,3482	0,3320	0,4633	0,2865	0,1070	0,2049
3	14	0,3748	0,3503	0,5305	0,3263	0,1405	0,1894
15	14	0,0000	0,0000	0,0000	0,0580	0,0580	0,0580
0	26	0,2488	0,2367	0,3404	0,1917	0,0245	0,0851
3	26	0,2498	0,2386	0,3101	0,1877	0,0426	0,0928
15	26	0,0113	0,0092	0,1875	0,0309	0,0232	0,0232
0	34	0,2410	0,2331	0,2790	0,1767	0,0155	0,0619
3	34	0,1912	0,1763	0,2910	0,1458	0,0193	0,0554
15	34	0,0000	0,0000	0,0000	0,0180	0,0180	0,0180
0	54	0,1878	0,1793	0,2337	0,1308	0,0155	0,0477
3	54	0,1656	0,1571	0,2183	0,1209	0,0142	0,0438
15	54	0,0318	0,0259	0,4248	0,0426	0,0193	0,0193
0	63	0,1859	0,1752	0,2460	0,1244	0,0142	0,0425
3	63	0,1730	0,1592	0,2678	0,1264	0,0116	0,0296
15	63	0,0000	0,0000	0,0000	0,0103	0,0103	0,0103
0	72	0,1719	0,1588	0,2580	0,1216	0,0142	0,0335
3	72	0,1544	0,1434	0,2251	0,1137	0,0129	0,0284
15	72	0,0038	0,0032	0,0120	0,0102	0,0078	0,0103
0	82	0,1416	0,1325	0,1954	0,0990	0,0065	0,0219

Continua en la página siguiente

<b>Conjunto</b>	<b>Etiquetas</b>	<i>Micro <math>F_2</math></i>	<i>Micro Recall</i>	<i>Micro Precision</i>	<i>Hamming Score</i>	<b>EMR</b>	<b>5-MR</b>
3	82	0,1710	0,1595	0,2414	0,1193	0,0090	0,0232
15	82	0,0084	0,0068	0,3935	0,0153	0,0090	0,0090
0	91	0,1415	0,1350	0,1832	0,0995	0,0090	0,0219
3	91	0,1901	0,1812	0,2423	0,1329	0,0090	0,0245
15	91	0,0196	0,0162	0,1394	0,0231	0,0077	0,0077
0	107	0,1481	0,1439	0,1688	0,1005	0,0013	0,0065
3	107	0,1839	0,1697	0,2809	0,1404	0,0116	0,0193
15	107	0,0254	0,0212	0,1904	0,0241	0,0064	0,0077
0	131	0,1342	0,1316	0,1482	0,0884	0,0039	0,0090
3	131	0,0631	0,0593	0,0856	0,0391	0,0026	0,0065
15	131	0,0335	0,0288	0,0958	0,0290	0,0052	0,0052
0	169	0,0652	0,0630	0,0803	0,0446	0,0052	0,0052
3	169	0,0334	0,0322	0,0421	0,0238	0,0052	0,0065
15	169	0,0410	0,0363	0,0896	0,0332	0,0090	0,0090
0	281	0,0843	0,0811	0,1015	0,0570	0,0039	0,0039
3	281	0,0429	0,0398	0,0620	0,0278	0,0026	0,0026
15	281	0,0727	0,0645	0,1533	0,0507	0,0052	0,0052