

Evaluation of Weakly Supervised Learning Paradigms on Automatic Visual Inspection

J. Arrieta, C. Mera, and A. Espinosa

Abstract—In pattern recognition supervised learning algorithms has been used to model the relationships among the characteristics of a group of objects (or patterns) and their class labels. This model is build based on a set of training images that have been labelled by an expert in a specific problem domain. In general, the labelling process consists in defining the object's class label as well as demarcating the precise part of the image in which it is located the object of interest. In many computer vision applications, for example in automatic visual inspection, the labelling process may be a laborious task which requires an extensive work, it could even become impractical when a large number of training images are needed to train the pattern recognition system. This limitation has led to the development of new learning algorithms that allow some ambiguity or "weakness" in the way class labels are assigned. These algorithms are known as *weakly supervised learning* algorithms. In this paper we present an evaluation of different weakly supervised learning paradigms grouped according to the representation of the object of interest (in one or several characteristic vectors) associated to its class label (single or multiple label). We perform different experiments in the automatic visual inspection context, using a benchmark dataset of texture images with artificial defects. We test different weakly supervised learning algorithms. Results show that representing an object using multiple instances, presents better results in the identification of defective texture image.

Index Terms—Weakly supervised learning, Computer vision, Automatic visual inspection systems.

I. INTRODUCCIÓN

LOS algoritmos de aprendizaje supervisado han sido utilizados en el área de reconocimiento de patrones para generalizar y aprender las relaciones que existen entre un grupo de objetos y sus etiquetas de clase. Generalmente, dicho proceso de generalización se realiza a partir de un conjunto de imágenes de entrenamiento en el que los objetos que se desean reconocer deben estar demarcados y relacionados con sus respectivas etiquetas de clase. Este proceso, denominado anotación manual, o simplemente supervisión, es en muchos casos una tarea laboriosa y costosa que requiere del esfuerzo y del tiempo de expertos anotadores humanos [1]–[5] lo que en algunas aplicaciones constituye una limitación en el entrenamiento de los sistemas de reconocimiento de patrones.

J. Arrieta, Universidad Nacional de Colombia - Sede Medellín, Medellín, Colombia, jmarrieta@unal.edu.co.

C. Mera, Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia, carlosmera@itm.edu.co.

A. Espinosa, Universidad Nacional de Colombia - Sede Medellín, Medellín, Colombia, aespinos@unal.edu.co.

Corresponding author: Jose Arrieta.

Esta y otras limitaciones han motivado el desarrollo de nuevos algoritmos de aprendizaje en los que se puede aprender a partir de conjuntos de datos en los que hay cierta "debilidad" o "ambigüedad" en la forma en la que se asignan las etiquetas de clase, o lo que es lo mismo, permiten cierta flexibilidad en el tipo de supervisión que se hace sobre los datos de entrenamiento. En la literatura existen diversos algoritmos que siguen esta tendencia y son denominados algoritmos de aprendizaje "débil". Ejemplos de estos son aquellos algoritmos del paradigma de aprendizaje semi-supervisado [4] en el que se utilizan tanto datos etiquetados como datos sin etiquetar para aprender a reconocer los patrones en los mismos. También existen los algoritmos del paradigma de aprendizaje de múltiples instancias (MIL) [6] en los que en lugar de clasificar *instancias* (o vectores de características) individuales, se clasifican conjuntos de instancias, denominados *bolsas*.

En la literatura, García *et al.* [7] han sido de los primeros autores en plantear una discusión sobre los algoritmos de etiquetado débil. En su trabajo, estos autores plantean que los algoritmos de aprendizaje pueden ser diferenciados de acuerdo al nivel de supervisión que se proporciona de los datos durante el proceso de aprendizaje, es así que proponen un *framework* que permite definir funciones objetivo flexibles, en el modelo de aprendizaje, que se acomodan naturalmente a variaciones del nivel de supervisión de los datos. Si bien es cierto que agrupar los problemas de clasificación desde una perspectiva común puede ser beneficioso, en muchos casos es preferible mantener las diferencias entre los algoritmos para discernir qué algoritmo puede ser más apropiado a cada problema particular de aprendizaje. En esta misma línea de trabajo, Yi *et al.* [8] proponen cuantificar el grado de supervisión de un problema de aprendizaje con base en las etiquetas que son incorrectamente asignadas por un experto y definen una forma de calcular la probabilidad que una etiqueta sea errónea con base en una cuantificación del nivel de supervisión. Como conclusión, Yi *et al.* indican que hay una correlación entre el grado de supervisión, la precisión del clasificador y el tiempo de cómputo, tal que, a medida que aumenta el grado de supervisión, suele aumentar la precisión y disminuir el tiempo de cómputo necesario para generar el modelo de clasificación.

Por otra parte, Hernández *et al.* [9] propusieron una taxonomía que permite situar los algoritmos de aprendizaje débil en tres ejes no excluyentes:

- En el primer eje de la taxonomía, los algoritmos pueden clasificarse como fuertemente supervisados, cuando se dispone de las etiquetas de clase de todas las muestras usadas en el entrenamiento; semi-supervisados cuando sólo se cuenta con las etiquetas de clase de algunas de las muestras de entrenamiento; o no supervisados cuando no se

cuenta con información sobre las etiquetas de clase de ninguna de las muestras usadas para el entrenamiento.

- En el segundo eje de la taxonomía, los algoritmos están catalogados con base en la información *a priori* que se pueda conocer de las etiquetas de clase de las muestras a ser clasificadas. En el modelo de clasificación tradicional se asume que no existe información alguna de las muestras a ser clasificadas, sin embargo, existen otros modelos de clasificación en los que el algoritmo de aprendizaje requiere, de manera explícita, un conjunto de etiquetas de clase candidatas para la muestra a ser clasificada.

- Finalmente, en el tercer eje de la taxonomía se agrupan los algoritmos con base en la relación que existe entre los datos del conjunto de entrenamiento y las etiquetas de clase que estos pueden tener asignadas. En este caso se consideran los algoritmos en los que, por ejemplo, un objeto puede tener uno o varios vectores de características (o *instancias*) y a la vez este puede tener asignadas una o varias etiquetas de clase.

De acuerdo con Hernández *et al.* [9], el último eje de su taxonomía puede dividirse en cuatro categorías. La primera, denominado *Una-Instancia Única-Etiqueta* (SISL) [4], la cual define el caso en el que cada muestra del conjunto de entrenamiento sólo puede tener asignada una sola etiqueta de clase, lo que en el contexto del reconocimiento de patrones significa que un objeto está representado únicamente por un solo vector de características que solo puede tener asociado una etiqueta de clase. En la segunda categoría, llamada *Una-Instancia Múltiples-Etiquetas* (SIML) [10], una muestra del conjunto de entrenamiento puede tener asignadas una o más etiquetas de clase, es decir, un objeto que es representado por un vector de características puede pertenecer a una o más clases. La tercera categoría, *Múltiples-Instancias Única-Etiqueta* (MISL) [11], cubre el caso en el que un objeto puede estar representado por múltiples vectores de características, aunque este solo puede tener asociado una etiqueta de clase. En la cuarta categoría, un objeto puede estar representado por múltiples vectores de características y a la vez puede pertenecer a múltiples clases. Esta configuración se denomina *Múltiples-Instancias Múltiples-Etiquetas* (MIML) [12].

Finalmente, aunque en los últimos años ha habido una tendencia en aumento del uso de este tipo algoritmos de aprendizaje emergentes, este aún se considera un campo de estudio activo con un estado del arte aún en desarrollo, especialmente en lo que se refiere a la comparación del desempeño entre los algoritmos de aprendizaje débil en dominios de aplicación específicos. Con base en lo anterior, en este trabajo se propone utilizar como punto de partida el tercer eje de la taxonomía de Hernández *et al.* [9] con el fin de hacer una asociación entre el concepto de debilidad y la anotación manual de las imágenes en el área del reconocimiento de patrones. En relación con esta asociación, se revisan y comparan algunos de los algoritmos de aprendizaje débil. Particularmente, la comparación se hace en el contexto del problema de inspección visual automática usando imágenes con texturas.

Este trabajo está organizado de la siguiente manera. En la Sección 2 se presenta la definición de las relaciones objeto-etiqueta que se usan en los algoritmos de aprendizaje débil agrupados en el tercer eje de la taxonomía propuesta por

Hernández *et al.* [9]. En la Sección 3 se revisan, en el contexto del reconocimiento de patrones, algunos de los algoritmos de aprendizaje débil de acuerdo las subcategorías propuestas por Hernández. En la Sección 4 se describe como modelar el problema de inspección visual automática desde el enfoque del paradigma de aprendizaje débil y se presentan los resultados experimentales en la detección de defectos sobre imágenes texturizadas. En la Sección 5 se presentan y se discuten los resultados de los experimentos. Finalmente, las conclusiones y el trabajo futuro se presentan en la Sección 6.

II. APRENDIZAJE DÉBIL Y SU RELACIÓN OBJETO-ETIQUETA

En los últimos años ha habido un incremento en el número de trabajos que están usando algoritmos de aprendizaje débil en diversos dominios de aplicación específicos [5], [13]–[16], sin embargo, poca atención se prestado a la categorización y definición de los algoritmos en este paradigma de aprendizaje emergente.

Existen múltiples definiciones del concepto “aprendizaje débil”, lo que genera falta de precisión en el contexto de la visión por computador. Una de estas definiciones es la Torresani *et al.* [17], en la cual los algoritmos de aprendizaje débil son considerados como aquellos en los que el modelo de clasificación se construye usando ejemplos que están parcialmente anotados o etiquetados. Por otro lado, Gu *et al.* [18] definen la debilidad refiriéndose a la ambigüedad presente en la delimitación y demarcación del objeto de interés y la definición de su etiqueta, mientras que Lefort *et al.* [19] se refiere al aprendizaje débil como aquel aprendizaje en el que el conjunto de datos de entrenamiento tiene asociada información *a priori* en la que se define la probabilidad de pertenencia de cada instancia a cada una de las clases disponibles. Por su parte, Yi *et al.* [8] definen los algoritmos de aprendizaje débil como aquellos en los cuales existe una alta probabilidad de asignar, erróneamente, la etiqueta de clase de un dato. Por último, Hernández *et al.* [9] hacen referencia a los algoritmos de aprendizaje débil como aquellos donde se utilizan conjuntos de datos en los que existe una falta de supervisión en los datos proporcionados.

Como se puede observar, las definiciones anteriores comparten la misma temática: el aprendizaje débil. Sin embargo, cada una de ellas aborda el aprendizaje débil desde una perspectiva diferente. En el caso de este trabajo, nos centramos en aquellos algoritmos de aprendizaje débil que permiten cierta ambigüedad en la representación de los objetos.

Se debe tener presente que en el contexto del reconocimiento de patrones la relación entre un objeto y su etiqueta, la relación objeto-etiqueta introducida en [9] y [12], está definida con base en el número de vectores de características usados para representar cada objeto y el número de etiquetas de clase que este puede tener asociadas a él. Partiendo de esta relación, los algoritmos de aprendizaje supervisado pueden pertenecer a uno de los cuatro grupos mostrados en la Fig. 1 y que se describen a continuación.

A. Una Instancia – Única Etiqueta (SISL)

En el aprendizaje supervisado tradicional, el conjunto de

datos para entrenar un clasificador toma la forma de la ecuación (1)

$$B = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, \quad (1)$$

donde $x_i \in \mathcal{X}$ es un vector de características (o instancia) que representa al objeto i , cuya etiqueta de clase es $y_i \in \mathcal{Y}$. Con base en esta representación, el algoritmo de aprendizaje debe encontrar una función de proyección (o clasificador) $f: \mathcal{X} \rightarrow \mathcal{Y}$ tal que, por ejemplo, minimice el error de clasificación. Con base en esta notación, un objeto, sin importar su complejidad, sólo puede ser representado por una instancia y sólo puede tener asociada una etiqueta de clase. En este tipo de algoritmos la ambigüedad en el etiquetado se presenta, o bien cuando en el conjunto de entrenamiento existen instancias sin una etiqueta asociada o cuando no hay certeza de la etiqueta asignada a cada instancia [20].

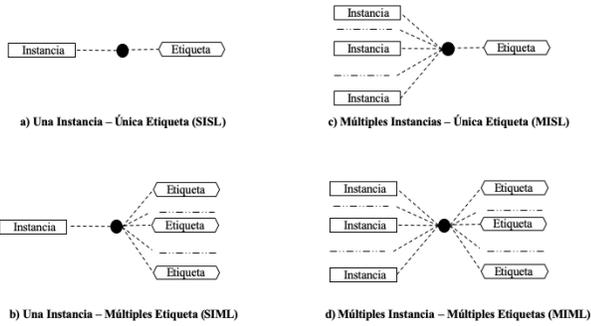


Fig. 1. Representación de diversos grupos de paradigmas de aprendizaje débil según su relación objeto etiqueta. Imagen modificada de [12].

B. Una Instancia – Múltiples Etiquetas (SIML)

En esta categoría de algoritmos de aprendizaje débil, los objetos son aún representados por una única instancia, pero estos pueden tener asignadas múltiples etiquetas de clases, ver Fig. 1c. En este caso, el objetivo es estimar una función de proyección $f: \mathcal{X} \rightarrow 2^y$ a partir de un conjunto de entrenamiento de la forma de la ecuación (2).

$$B = \{(x_1, \mathcal{Y}_1), (x_2, \mathcal{Y}_2), \dots, (x_m, \mathcal{Y}_m)\}. \quad (2)$$

En este caso, $\mathcal{Y}_i \subseteq \Omega$ es un conjunto de etiquetas de clase $\{\psi_{i1}, \psi_{i2}, \dots, \psi_{i, l_i}\}$, $\psi_{ik} \in \Omega$ y $k = 1, 2, \dots, l_i$. Donde l_i denota el número de clases en Ω . Note que la ambigüedad se presenta por las múltiples etiquetas que puede tener un solo objeto. En la literatura se hace referencia a esta categoría como el paradigma de aprendizaje con múltiples etiquetas [21].

C. Múltiples Instancias – Única Etiqueta (MISL)

Esta categoría de algoritmos de aprendizaje débil hace referencia principalmente a los algoritmos del paradigma de aprendizaje de múltiples instancias introducido por Dietterich *et al.* [22]. Aquí, un objeto puede estar representado por múltiples instancias (vectores de características) pero solo puede tener asociado a él una única etiqueta de clase, como ilustra la Fig. 1b. En este caso, los algoritmos deben estimar una función de proyección $f: 2^x \rightarrow \mathcal{Y}$ a partir de un conjunto de entrenamiento con la forma de la ecuación (3)

$$B = \{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}, \quad (3)$$

donde $X_i = \{x_{i1}, x_{i2}, \dots, x_{i, n_i}\} \subseteq \mathcal{X}$ es una bolsa (o conjunto de instancias) cuya etiqueta de clase es $y_i \in \mathcal{Y}$. En esta categoría de algoritmos de aprendizaje débil, la ambigüedad se presenta por la falta de información que existe con respecto a la etiqueta de cada instancia, puesto que la etiqueta es asignada a la bolsa como un todo y no a las instancias individuales.

D. Múltiples Instancias – Múltiples Etiquetas (MIML)

En esta categoría, un objeto puede ser representado por múltiples instancias y puede ser etiquetado con múltiples etiquetas de clases [12], [23]. En este sentido, el objetivo es aprender una función de proyección $f: 2^x \rightarrow 2^y$ a partir de un conjunto de datos con la forma de la ecuación (4)

$$B = \{(X_1, \mathcal{Y}_1), (X_2, \mathcal{Y}_2), \dots, (X_m, \mathcal{Y}_m)\}, \quad (4)$$

donde $X_i \subseteq \mathcal{X}$ es una bolsa de instancias y $\mathcal{Y}_i \subseteq \Omega$ es el conjunto de etiquetas de clase asignadas a la bolsa. En esta categoría, la ambigüedad radica en la falta de información generada por el desconocimiento que existe de la etiqueta de clase de cada una de las instancias, además del número de etiquetas diferentes que un solo objeto puede llegar a tener.

III. RECONOCIMIENTO DE PATRONES EN IMÁGENES UTILIZANDO ALGORITMOS DE APRENDIZAJE DÉBIL

En la literatura existen diversos trabajos en los que se proponen o utilizan algoritmos de aprendizaje débil para el reconocimiento de patrones en imágenes [24]. A continuación, se revisan algunos de estos algoritmos, agrupándolos en las cuatro categorías descritas anteriormente.

A. Una Instancia – Una Etiqueta (SISL)

Dentro de este grupo de paradigmas de aprendizaje débil SISL, un objeto o patrón consiste de una región de interés en una imagen determinada. Cada objeto es entonces representado por un vector de características de bajo nivel que describe la región que lo contiene y tiene asociado una etiqueta de clase que representa la clase de objetos a la que este pertenece. En esta categoría existen diversos algoritmos en diferentes paradigmas de aprendizaje para estimar la función de proyección $f: \mathcal{X} \rightarrow \mathcal{Y}$.

A1. Aprendizaje Semi-supervisado

Este paradigma es empleado cuando los datos etiquetados son usualmente difíciles de obtener, costosos o consumen mucho tiempo por parte de un anotador humano. En este sentido, los algoritmos de este paradigma utilizan grandes cantidades de datos sin etiquetar junto con algunos datos etiquetados, a partir de los cuales se construye y se refina el modelo de clasificación. A continuación, se describen algunas de las aproximaciones de aprendizaje semi-supervisado más comunes.

A1.1 Auto-Entrenamiento (Self-Training)

En *self-training* un clasificador es construido utilizando las instancias etiquetadas. Este clasificador es luego usado para estimar las etiquetas de las instancias sin etiqueta. Las

instancias con las predicciones más confiables son agregadas al conjunto de entrenamiento y el proceso de entrenamiento se repite de nuevo. Es importante anotar que el clasificador usa sus predicciones para enseñarse a sí mismo [4]. Por ejemplo, Zhao *et al* proponen en [25] un algoritmo de análisis de discriminante lineal (LDA) basado en *self-training* para el reconocimiento de rostros.

A1.2 Co-Entrenamiento (Co-Training)

Co-training asume que las características se pueden separar en dos grupos, tal que cada subconjunto de características es utilizado para construir dos clasificadores, uno por cada subconjunto. Después, al igual que con *self-training*, cada clasificador predice las etiquetas de los objetos sin etiquetar y las predicciones más confiables son usadas para re-entrenar el clasificador contrario[4]. Un ejemplo de uso de este algoritmo es presentado en [26] para reconocer y diferenciar entre diferentes tipos de objetos en imágenes.

A1.3 Métodos basados en Grafos

Los algoritmos de aprendizaje de este tipo definen un grafo en el que los nodos representan objetos etiquetados u objetos no etiquetados del conjunto de entrenamiento, mientras que las aristas definen la similitud entre los objetos que estas conectan [4]. Un caso de uso de este tipo de algoritmos es descrito en [27], en el que los autores proponen identificar una enfermedad llamada glositis a partir de imágenes de la lengua. Otro ejemplo es presentado en [28] para reconocer las imágenes de una base de datos denominada Caltech-4. Un elemento que cabe anotar de los métodos de aprendizaje semi-supervisado basados en grafos, es que estos asumen cierta distribución estadística sobre los datos de entrenamiento de tal manera que se considera que puntos cercanos comparten etiquetas similares.

A2. Positivo sin etiqueta / Positive-unlabeled (PU)

El paradigma de aprendizaje *positive-unlabeled (PU)*, se centra en problemas de clasificación binaria en los que es difícil conseguir muestras etiquetadas de una de las dos clases de interés [29]. A diferencia de los algoritmos de aprendizaje de una clase (*one-class classifiers*), los algoritmos en el paradigma PU usando dos conjuntos de entrenamiento, uno con la clase positiva y otro en el que hay instancias positivas y negativas, aunque no se conoce específicamente cuales son cuales. En el reconocimiento de patrones, en [29] utilizan el algoritmo *positive naïve Bayes* (PNB) y *positive tree augmented naïve Bayes* (PTAN) en el reconocimiento de caracteres.

A3. Etiquetas parciales / Partial labels

En este paradigma, cada instancia tiene asociada un conjunto de etiquetas candidatas de las cuales sólo una es correcta, contrario al aprendizaje con múltiples etiquetas en el que todas las etiquetas pueden ser válidas. En este sentido, los algoritmos en el paradigma de etiquetas parciales buscan construir un modelo que, además de generalizar, pueda desambiguar las etiquetas de los objetos. Una aplicación de este paradigma de aprendizaje consiste en el etiquetado automático de rostros en una colección de imágenes y videos de programas televisivos [30]. En dicha propuesta las etiquetas candidatas son obtenidas de los guiones escritos para los programas. Otro ejemplo de aplicación de este paradigma se presenta en [31]. En dicho

trabajo los autores proponen un algoritmo denominado IPAL (*Instance-based Partial Label learning*) el cual construye un grafo asimétrico ponderado sobre las instancias de entrenamiento realizando un análisis de afinidad entre estas, luego de esto el algoritmo IPAL trata de identificar la etiqueta correcta de cada instancia por medio de propagación de la etiqueta.

A4. Relaciones de equivalencia parciales / Partial equivalence relations

En este paradigma de aprendizaje los datos de entrenamiento se obtienen en pequeños grupos, donde los objetos de cada grupo están sujetos a pertenecer a una misma clase, aunque dicha clase es desconocida. Esto genera una relación de pareja del tipo 'x es similar a y' para todos los puntos x e y que pertenecen al mismo grupo [32]. Esta forma de aprendizaje se ha usado, por ejemplo, para aprender discriminativamente de grupos de imágenes de rostros estándar, considerando que cada grupo corresponde a diferentes imágenes de una misma persona [32].

A5. Aprendiendo de las multitudes / Learning from crowds

Esta aproximación de aprendizaje hace referencia a situaciones donde se cuenta con múltiples expertos o anotadores, pero donde sólo se elige una etiqueta dentro de las sugeridas. Raykar *et al* proponen en [20] un método probabilístico que ayuda a determinar cuáles son los mejores expertos y de esta forma asigna los pesos de las etiquetas que estos asignan a los datos de entrenamiento. Los autores emplean dicho método en el problema de detección de cáncer en mamografías.

B. Una Instancia – Múltiples Etiquetas (SIML)

Dentro de este grupo de paradigmas de aprendizaje débil SIML, un objeto o una región de interés en la imagen es representado por un vector de características que tiene asociado una o múltiples etiquetas de clase. De esta forma, el objetivo del algoritmo de aprendizaje consiste en estimar una función que proyecte un vector de características (o instancia) a un subconjunto de las etiquetas de clase disponibles. Entre los paradigmas de aprendizaje débil en esta categoría se encuentran:

B1. Aprendizaje de múltiples etiquetas

En este paradigma se permite que una instancia tenga asociada varias etiquetas de clase. En general, los algoritmos de este paradigma siguen una estrategia *one vs all*, donde se entrena un clasificador binario para cada clase, no obstante, se debe considerar que esta estrategia ignora la relación que existe entre las etiquetas, lo que es una limitación. Una de las propuestas en este paradigma, en el contexto del reconocimiento de objetos, explota la idea de combinar las etiquetas con el fin de aumentar el poder de clasificación en el espacio de las etiquetas [33]. Por otro lado, en [34] los autores proponen un esquema de aprendizaje basado en ensambles de clasificadores que usan múltiples etiquetas. Dicha propuesta es evaluada utilizando el conjunto de imágenes de *Corel* con el fin de reconocer los objetos en las mismas. En la misma línea del trabajo anterior, en [21] los autores presentan una extensión de un algoritmo basado en ensambles de árboles de decisión en el

que cada árbol se entrena sin utilizar la información de las etiquetas de clase, como resultado cada árbol genera una distribución de probabilidad que, ponderada con la de los otros árboles, ayuda a determinar las etiquetas de clase de un objeto. Siguiendo una aproximación diferente, los autores en [10] proponen un algoritmo de *clustering* de múltiples etiquetas denominado CBMLC. Los autores reportaron que el algoritmo es efectivo cuando se busca clasificar imágenes de escenas naturales que han sido anotadas hasta con 6 etiquetas.

B2. Aprendizaje de múltiples etiquetas semi-supervisado

El paradigma de aprendizaje semi-supervisado con múltiples etiquetas usa, al igual que el aprendizaje semi-supervisado tradicional, tanto instancias etiquetadas como instancias sin etiquetar para entrenar el clasificador. Un ejemplo de este tipo de algoritmos es el presentado en [35], donde el proceso de aprendizaje se lleva a cabo en un sub-especie de las instancias siguiendo la aproximación usual de los algoritmos de aprendizaje semi-supervisados.

Por otro lado, los autores en [36] presentan una propuesta, basada en el algoritmo k-NN, para disminuir el ruido generado por las instancias sin etiqueta que son usadas para re-entrenar el clasificador. Este algoritmo fue evaluado en el contexto de reconocimiento de imágenes de escenas naturales con múltiples etiquetas que incluyen desiertos, mares, árboles, playas, bosques entre otros, además de otras imágenes sin etiquetas.

C. Múltiples Instancias – Una Etiqueta (MISL)

Dentro del grupo de paradigmas de aprendizaje débil MISL, un objeto puede estar representado por varias regiones de interés en la imagen y, por lo tanto, este está descrito por una colección de múltiples vectores de características (o *instancias*) llamada *bolsa*. En esta representación cada bolsa tiene asociada una etiqueta de clase. Así, los algoritmos en este paradigma buscan una función que proyecte el conjunto de instancias de cada bolsa a una sólo etiqueta de clase que describe al objeto como un todo. Algunos paradigmas en este grupo son:

C1. Aprendizaje de múltiples Instancias (MIL)

En MIL, en lugar de clasificar instancias individuales, como se hace en el aprendizaje supervisado tradicional, se clasifican grupos de instancias, teniendo en cuenta que la etiqueta de cada instancia individual es desconocida. En términos de reconocimiento de patrones, cada imagen del conjunto de entrenamiento poder estar representada por una bolsa cuyas instancias se obtienen de los segmentos o zonas de interés de esta. Una de las aplicaciones iniciales de MIL en el reconocimiento de patrones usando imágenes corresponde al uso del algoritmo mi-SVM [37] para la clasificación de imágenes en la base de datos de imágenes de Corel. Otras aplicaciones de MIL incluyen la detección y clasificación de regiones cancerígenas en imágenes histopatológicas [37, 38] y en imágenes de la piel [40].

C2. Aprendizaje de múltiples instancias generalizado (G-MIL)

G-MIL es una versión generalizada del paradigma MIL que extiende el supuesto standard a fin de considerar, no solo una instancia, sino un grupo de instancias para definir la etiqueta de la bolsa [41][42]. Específicamente, Scott *et al.* evaluaron el

desempeño del algoritmo G-MIL1 [41] en la recuperación de imágenes por contenido. Unos años más tarde, los mismos autores presentaron un algoritmo llamado G-MIL2 [43] el cual presenta una mejora de G-MIL1 en términos del tiempo de cómputo y consumo de memoria. Otras variantes de GMIL han sido presentadas en [44] y [45].

C3. Conjuntos con etiquetados candidatos / *Candidate labeling sets*

A diferencia de MIL, los algoritmos en el paradigma *candidate labeling* asocian a una bolsa un conjunto de posibles etiquetas teniendo en cuenta que sólo una de estas es la etiqueta real. *Candidate labeling* se distingue del grupo MIML en que en este último todas las etiquetas asignadas una bolsa son etiquetas reales de la misma, es decir, no hay incertidumbre el conjunto de etiquetas asignadas a cada bolsa. Un algoritmo representativo del paradigma *candidate labeling* se denomina MMS (*Maximum Margin Set learning*) y es introducido en [46] en el contexto del reconocimiento de rostros en imágenes de noticias.

D. Múltiples Instancias – Múltiples Etiquetas (MIML)

En esta categoría de paradigmas un objeto puede estar representado por varias regiones de interés en la imagen y, por tanto, está descrito por múltiples vectores de características agrupados en bolsas. La diferencia con los algoritmos en la categoría MISL es que una bolsa tiene asociada múltiples etiquetas de clase que pueden representar grupos de objetos diferentes. En este caso, los algoritmos estiman una función que proyecta un conjunto de instancias (una bolsa) a un subconjunto del conjunto de etiquetas de clase. Dos algoritmos en este grupo son los algoritmos MIMLBOOST y MIMLSVM [12] los cuales son evaluados en el reconocimiento de escenas que contienen múltiples objetos, mientras que Chen *et al.* [23] usan un tipo de red neuronal para MIML en la clasificación de imágenes en la base de datos de Corel.

Una de las limitaciones de MIML es que se asume que una bolsa está etiquetada con todas las etiquetas de los objetos que esta contiene, no obstante, en las aplicaciones reales los usuarios pueden omitir etiquetas. Para cubrir esta limitación, Zhou *et al.* proponen MIMLwel [12] un algoritmo que busca tratar con este tipo de etiquetado débil.

IV. EXPERIMENTOS

Empleando los algoritmos descritos en la sección previa se realizaron experimentos sobre un banco imágenes que ya ha sido usado en inspección visual automática. El proceso de inspeccionar visualmente de forma automática, típicamente se lleva a cabo en cuatro etapas que abarcan: la adquisición de la imagen, el pre procesamiento, la extracción de características y el reconocimiento de los defectos. En esta última etapa, se requiere de un conjunto de imágenes de entrenamiento, previamente etiquetadas, para enseñar a la máquina a reconocer los posibles defectos. En muchos casos prácticos, la disponibilidad de los expertos es limitada o la cantidad de imágenes sin etiquetar disponibles supera la cantidad de imágenes etiquetadas o las imágenes han sido etiquetadas débilmente, es decir, se conoce que imágenes tienen defectos, pero se desconoce la ubicación de estos.

En este sentido, la inspección visual automática se puede modelar como un problema de aprendizaje débil debido a que es común contar con una gran cantidad de imágenes sin etiquetar o etiquetadas débilmente. Es así que en los experimentos presentados más adelante se comparan diferentes algoritmos los cuales pertenecen a los diferentes grupos de paradigmas de aprendizaje débil descritos en la sección anterior. Dicha comparación busca determinar cuál de los algoritmos evaluados es el más apropiado para el problema de inspección visual en el dominio de aplicación seleccionado.

El esquema general del proceso de entrenamiento y pruebas se presenta en la Fig. 2. El proceso inicia con un conjunto de imágenes de entrenamiento débilmente etiquetadas, la región débilmente etiquetada se particiona en bloques a partir de los cuales se extraen las características para formar las instancias. Estas últimas constituyen el conjunto de vectores de características con el que se entrena el clasificador. El uso del clasificador sigue un proceso similar, como lo muestra la parte inferior de la Fig 2.

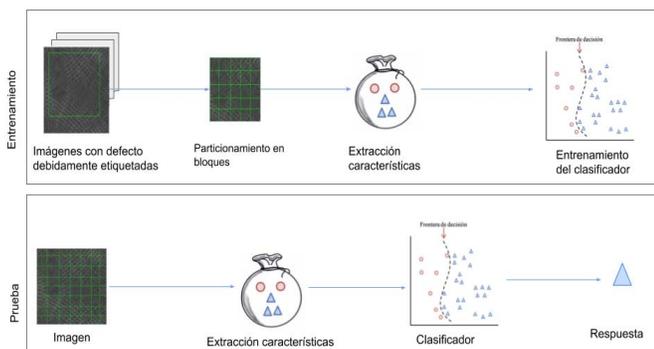


Fig. 2. Proceso general de entrenamiento y clasificación de defectos.

A continuación, se describe la configuración tanto de los experimentos como de los conjuntos de datos a utilizar.

A. Conjunto de Datos de Prueba

Los experimentos se llevaron a cabo empleando el conjunto de datos del DAGM [47], el cual consiste en diez conjuntos de imágenes con diversas clases de texturas y diferentes tipos de defectos artificiales. Este conjunto de datos fue diseñado, artificialmente, para simular un problema de inspección sobre objetos con texturas.

Cada conjunto contiene 1000 imágenes texturizadas sin defectos y 150 imágenes con defectos. En los experimentos realizados en este trabajo se utilizaron 5 clases de texturas que corresponden a las clases 1, 2, 3, 4 y 5, generadas cada una con un modelo de textura y un defecto diferente como lo indican en [47]. Con el fin de utilizar los algoritmos de múltiples etiquetas en los experimentos, algunas de las imágenes se modificaron para contener dos tipos de defectos diferentes, el defecto original y un defecto inducido que se creó aplicando un filtro de suavizado. El tamaño de la máscara de filtrado es de 9x9 píxeles para las clases 1 y 5, de tamaño 11x11 para las clases 2 y 3, y de tamaño de 13x13 para la clase 4, cada uno relativo al tamaño del defecto original. Un ejemplo de las imágenes usadas se muestra en la Fig. 3.

B. Generación de Instancias

La generación de instancias se realizó con base en un etiquetado débil de las imágenes. Para ello, se demarcaron las regiones que encierran los defectos en la imagen. Se debe considerar que este tipo de etiquetado es débil dado que los defectos no se demarcan de manera precisa por los que las zonas que demarcadas contienen tanto píxeles defectuosos como no defectuosos. Estas zonas se definieron automáticamente con la ayuda del *ground truth* de las imágenes. Para cada zona de defecto usó una estrategia de ventana deslizante para generar las instancias a partir de las características extraídas de bloques de tamaño 32x32 píxeles.

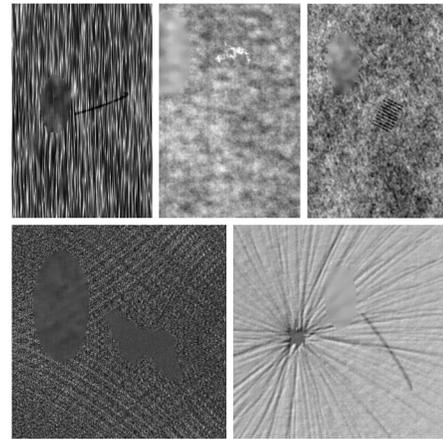


Fig. 3. Ejemplos de las imágenes con los defectos originales y los defectos inducidos.

Para aquellos algoritmos en los paradigmas de múltiples instancias, el conjunto de instancias generado a partir de una región defectuosa se agrupa en una bolsa, es decir, cada defecto conforma una bolsa por sí mismo. Por el contrario, en los algoritmos de una sola instancia, cada bloque de 32x32 píxeles es tratado de manera independiente como una instancia positiva, es decir, un bloque que contiene parte del defecto. En las imágenes negativas, es decir, las imágenes sin defectos, una bolsa se constituye de todas las instancias generadas deslizando la ventana por toda la imagen.

C. Extracción de Características

Para calcular el vector de características que representa cada instancia se utilizaron los algoritmos de extracción de características pertenecientes a varias librerías (balu-python [48], scikit-image [49] y mahotas [50]). Dentro de las características extraídas de las imágenes se encuentran características que consideraran los cambios de intensidad de la luz, textura y forma. A continuación, se listan las características calculadas.

- **Características básicas de Intensidad** [51]: Estas son características básicas de intensidad como la media, la desviación estándar, kurtosis, oblicuidad, media laplaciana y gradiente de límite medio.
- **Características de textura con LBP** [52]: Describe la textura de una imagen o región basado en la relación entre el valor de intensidad de cada píxel en la imagen con sus vecinos.

- **Características de textura de Haralick** [53]: Estas características se obtienen a partir de las matrices de adyacencia de los píxeles.
- **Estadísticas de Adyacencia de Umbral (TAS)** [54]: Esta característica define un umbral para la imagen y cuenta el número de píxeles por encima de este umbral que tengan un número dado de píxeles adyacentes por encima del umbral definido.
- **Momentos de Zernike** [55]: los cuales son una medida global que cuantifica como se distribuyen los píxeles en el bloque.
- **Características de forma con HOG** [56] : En general, este descriptor cuenta las ocurrencias de las orientaciones del gradiente en bloques de tamaño fijo en la imagen y ha sido uno de los descriptores más efectivos para capturar la información de forma en los bordes o regiones de la imagen.

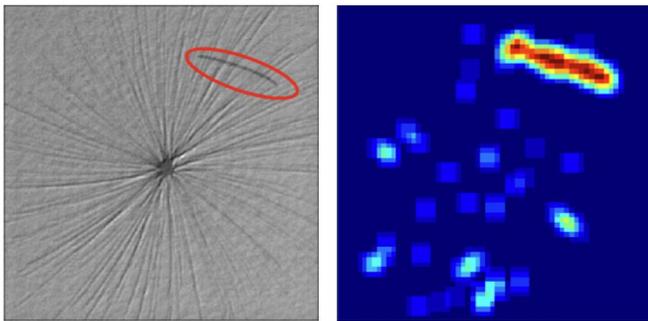


Fig. 4. Ejemplo de Imagen de entrada y salida con el algoritmo de clasificación.

V. RESULTADOS

Para la evaluación se adoptó una validación cruzada estratificada con 10 particiones. Se seleccionaron diferentes algoritmos para cada uno de los paradigmas de aprendizaje débiles indicados en la Fig. 1. Dicha selección se realizó con base en la disponibilidad de una implementación en los lenguajes de programación Python o Matlab. Para medir el desempeño de los algoritmos se usó la métrica del área bajo la curva (AUC). La selección de esta métrica se debe a que es la métrica más usada cuando se utilizan conjuntos de datos desbalanceados, como suelen ser los conjuntos de datos obtenidos de problemas de inspección visual automática [57] [24]. En la Tabla 1 se resumen los resultados obtenidos correspondientes al promedio del AUC obtenido en todas las repeticiones de la validación cruzada. En dicha tabla, la primera columna hace referencia a los algoritmos utilizados, en corchetes se encuentra el paradigma de aprendizaje débil al cual pertenece y en paréntesis los meta-parámetros usados, en caso de no señalarlo explícitamente, los valores son los usados por defecto en el toolbox. Las siguientes columnas reportan el desempeño obtenido en cada una de las clases de texturas utilizadas. Los mejores resultados son resaltados en negrita para cada conjunto de datos.

Algunas observaciones pueden ser realizadas a partir de los resultados obtenidos. Para estos conjuntos de imágenes, los paradigmas de aprendizaje débil de una sola instancia [SISL] presentan poco favorables en todas las clases como se aprecia en las primeras dos filas de la Tabla 1. Esto puede explicarse, en parte, por el desbalance que hay entre las clases, esto es un problema porque el clasificador tiende a favorecer la clase

mayoritaria y puede sesgar los resultados. Los paradigmas de aprendizaje con múltiples instancias y múltiples etiquetas [MIML] presentan mejores resultados que los paradigmas de instancia única, sin embargo, estos presentan resultados más desfavorables en comparación a los paradigmas de múltiples instancias con etiqueta única [MISL], esto puede ser por dificultad de generalizar la clasificación usando dos etiquetas. Por último, los paradigmas de múltiples instancias con etiqueta única presentaron mejores resultados en las clases, dentro de los que se destacan los algoritmos de *Simple MIL* y *Bags of Words* (BOW), debido al poder generalizar ambos defectos en una sola etiqueta, haciendo más sencillo el aprendizaje.

TABLA I
RESULTADOS OBTENIDOS PARA LA MÉTRICA AUCx100

Algoritmo	AUC				
	Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
Label propagation (kernel: knn, alpha:1.0, neighbors = 7, 30%) [SISL]	50,94	50,00	50,00	54,06	51,06
Label propagation (kernel: rbf, alpha:1.0, neighbors = 7, 30%) [SISL]	53,56	57,22	54,56	55,00	50,56
OneVsRestClassifier (SVC, kernel: rbf, C:1, degree:3.) [SIML]	57,44	50,00	46,89	53,44	47,56
OneVsRestClassifier (SVC, kernel: linear, C:1, degree:3) [SIML]	55,11	53,78	61,11	55,44	52,56
SimpleMI (type:min) [MISL]	83,94	66,72	81,89	63,56	71,33
SimpleMIL(type:avg) [MISL]	86,89	96,83	97,00	96,50	97,44
SimpleMIL(type:max) [MISL]	99,50	100,00	100,00	100,00	100,00
SimpleMI (type:extreme) [MISL]	99,50	99,50	100,00	100,00	100,00
BOW (GMM,k:10, covar_type:'diag') [MISL]	100,00	100,00	100,00	100,00	100,00
EMDD (alf:10, spoints = 10) [MISL]	70,50	61,50	59,50	62,23	61,00
MIMLBOOST [MIML]	74,00	74,00	74,00	74,00	74,00
MIMLSVM [MIML]	74,00	74,00	74,00	74,00	74,00

Finalmente, como resultado en estos conjuntos de datos provenientes de defectos en texturas, el tratar de modelar las imágenes como bolsas de instancias y dentro de las bolsas tratar de identificar el concepto objetivo, modelando la población negativa y donde el objetivo es buscar la más positiva de las instancias en las bolsas positivas, como es en el caso del algoritmo de *Bag of Words* o identificar la instancia máxima en las bolsas, resulta en una buena aproximación para detectar las imágenes con defectos. Esta aproximación resulta más cercana al ideal de un ambiente real de inspección visual donde los objetos a inspeccionar deben estar libres de cualquier defecto o combinación de ellos y solo una imagen sin defectos es la ideal, influenciando el mejor desempeño de los algoritmos del paradigma MISL.

A modo de ejemplo, la Fig 4 muestra el resultado de segmentar un defecto en una de las imágenes de prueba. Este tipo de resultados se obtuvo utilizando una estrategia de ventana deslizante en la que la imagen original se particiona en bloques

de 32x32, cada bloque se clasifica para determinar si tiene o no un defecto. La intensidad de la zona con un defecto se obtiene sumando los bloques en esa zona que son clasificados como defecto.

VI. CONCLUSIONES

En este trabajo se realizó una revisión de los paradigmas de aprendizaje débil. Estos se agruparon de acuerdo con la representación del objeto de interés y su etiqueta de clase. Se llevó a cabo una comparación práctica, en el contexto de la inspección visual automática empleando un banco de imágenes. Se usaron imágenes con diferentes texturas a fin de identificar diferentes defectos en las mismas a partir de un etiquetado débil de las imágenes.

Con base en los experimentos se puede concluir que, en el contexto de la inspección visual automática, resulta una aproximación acertada la representación de las imágenes en el paradigma de aprendizaje de Múltiples Instancias- Única Etiqueta (MISL), debido a que permite modelar las imágenes como bolsas con instancias que representan defectos y bolsas con instancias que no contienen defectos. Realizando la experimentación se encontró que el tiempo de procesamiento se redujo en esta categoría de paradigmas.

Nuestro trabajo futuro en esta área de investigación estará enfocado a la aplicación de paradigmas de aprendizaje débil en diferentes contextos de reconocimiento de patrones de imágenes donde la ambigüedad en la representación del objeto de interés esté presente.

REFERENCIAS

- [1] J. B. Li, Y. Yu, Z. M. Yang, and L. L. Tang, "Breast tissue image classification based on semi-supervised locality discriminant projection with kernels," *J. Med. Syst.*, vol. 36, no. 5, pp. 2779–2786, 2012.
- [2] C. Mera, "Un Método de Clasificación para la Detección de Defectos Superficiales en Sistemas de Inspección Visual Automática," Universidad Nacional de Colombia, 2013.
- [3] N. N. Pise and P. Kulkarni, "A Survey of Semi-Supervised Learning Methods," in *International Conference on Computational Intelligence and Security*, 2008, vol. 2, pp. 30–34.
- [4] X. Zhu, "Semi-Supervised Learning Literature Survey Contents," *Sci. York*, vol. 10, no. 1530, p. 10, 2008.
- [5] J. Song, H. Liu, F. Geng, and C. Zhang, "Weakly-Supervised Classification of Pulmonary Nodules Based on Shape Characters," in *IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing*, 2016, pp. 228–232.
- [6] J. Amores, "Multiple Instance Classification: Review, Taxonomy and Comparative Study," *Artif. Intell.*, vol. 201, pp. 81–105, 2013.
- [7] D. García-García and R. Williamson, "Degrees of supervision," *Tsc.Uc3M.Es*, pp. 1–11, 2011.
- [8] X. Yi, Z. Wang, Z. Yang, C. Caramanis, and H. Liu, "More Supervision, Less Computation: Statistical-Computational Tradeoffs in Weakly Supervised Learning," *Nips*, pp. 4475–4483, 2016.
- [9] J. Hernandez-Gonzalez, I. Inza, and J. A. Lozano, "Weak supervision and other non-standard classification problems: A taxonomy," *Pattern Recognit. Lett.*, vol. 69, pp. 49–55, 2016.
- [10] G. Nasierding, "Multi-label Classification with Clustering for Image and Text Categorization," *6th Int. Congr. Image Signal Process.*, pp. 869–874, 2013.
- [11] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artif. Intell.*, vol. 201, pp. 81–105, Aug. 2013.
- [12] Z. H. Zhou, M. L. Zhang, S. J. Huang, and Y. F. Li, "Multi-instance multi-label learning," *Artif. Intell.*, vol. 176, no. 1, pp. 2291–2320, 2012.
- [13] G. Patrini, F. Nielsen, R. Nock, and M. Carioni, "Loss factorization, weakly supervised learning and label noise robustness," *CoRR*, vol. abs/1602.0. 2016.
- [14] A. Joulin and F. Bach, "A convex relaxation for weakly supervised classifiers," in *ICML*, 2012.
- [15] A. Prest, C. Schmid, and V. Ferrari, "Weakly Supervised Learning of Interactions between Humans and Objects," vol. 34, no. 3, pp. 1–14, 2012.
- [16] D. Crandall and D. Huttenlocher, "Weakly supervised learning of part-based spatial models for visual object recognition," *Eccv*, pp. 16–29, 2006.
- [17] L. Torresani, "Weakly Supervised Learning," *Media*, vol. 42, p. 823, 2010.
- [18] F. Gu *et al.*, "Weakly supervised activity analysis with spatio-temporal localisation," *Recent Adv. Hybrid Artif. Intell. Syst. its Appl. to Real-World Probl. Pap. from HAIS 2013 Conf.*, vol. 216, pp. 778–789, 2016.
- [19] J.-M. B. R. Lefort, R. Fablet, "Weakly Supervised Learning with decision trees applied to fisheries acoustics," *2010 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 2254–2257, 2010.
- [20] V. C. Raykar *et al.*, "Learning From Crowds," *J. of Machine Learn. Res.*, vol. 11, pp. 1297–1322, 2010.
- [21] X. Zhang, Q. Yuan, S. Zhao, and W. Fan, "Multi-label classification without the multi-label cost," in *Proceedings of the 2010 SIAM International Conference on Data Mining*, 2010, pp. 778–789.
- [22] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, pp. 31–71, 1997.
- [23] Z. Chen, Z. Chi, H. Fu, and D. Feng, "Multi-instance multi-label image classification: A neural approach," *Neurocomputing*, vol. 99, pp. 298–306, 2013.
- [24] C. Mera and J. W. Branch, "A Survey on Class Imbalance Learning on Automatic Visual Inspection," *IEEE Lat. Am. Trans.*, vol. 12, no. 4, pp. 657–667, 2014.
- [25] X. Zhao, N. Evans, and J. L. Dugelay, "Semi-supervised face recognition with LDA self-training," *Proc. - Int. Conf. Image Process. ICIP*, pp. 3041–3044, 2011.
- [26] I. Koprinska, D. Deng, and F. Felix, "Image Classification Using Labelled and Unlabelled Data," *14th Eur. Signal Process. Conf.*, vol. 1, 2016.
- [27] C. Li and P. Yuen, "Semi-supervised Learning in Medical Image Database," *Adv. Knowl. Discov. Data Min.*, pp. 154–160, 2001.
- [28] L. Z. -, B. Y. -, M. Z. -, H. Z. -, and W. L. -, "Image Classification by Semi-Supervised Learning with Spatial Information," *Int. J. Adv. Comput. Technol.*, vol. 4, no. 22, pp. 87–96, 2012.
- [29] B. Calvo, P. Larrañaga, and J. A. Lozano, "Learning Bayesian classifiers from positive and unlabeled examples," *Pattern Recognit. Lett.*, vol. 28, no. 16, pp. 2375–2384, 2007.
- [30] T. Cour and B. Sapp, "Learning from Partial Labels," *J. Mach. Learn. Res.*, vol. 12, no. 2, pp. 1501–1536, 2011.
- [31] M. Zhang and F. Yu, "Solving the Partial Label Learning Problem: An Instance-based Approach," in *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015, pp. 4048–4054.
- [32] S. Kumar and H. A. Rowley, "Classification of weakly-labeled data with partial equivalence relations," in *Proceedings of the IEEE International Conference on Computer Vision*, 2007, pp. 1–8.
- [33] X. Li, F. Zhao, and Y. Guo, "Multi-label Image Classification with A Probabilistic Label Enhancement Model," in *Thirtieth Conference on Uncertainty in Artificial Intelligence*, 2014, pp. 430–439.
- [34] D. Zhang and X. Liu, "Ensemble of k-Labelset Classifiers for Multi-label Image Classification," in *Intelligent Information Processing VI*, Springer Berlin Heidelberg, pp. 364–371.
- [35] Y. Guo and D. Schuurmans, "Semi-supervised Multi-label Classification," in *Machine Learning and Knowledge Discovery in Databases*, Springer Berlin Heidelberg, 2012, pp. 355–370.

- [36] Z. Wei, H. Wang, and R. Zhao, "Semi-supervised multi-label image classification based on nearest neighbor editing," *Neurocomputing*, vol. 119, pp. 462–468, 2013.
- [37] S. Andrews, T. Hofmann, and I. Tsochantaridis, "Multiple Instance Learning with Generalized Support Vector Machines," *AAAI Eighteenth Natl. Conf. Artif. Intell.*, pp. 943–944, 2002.
- [38] M. Kandemir, C. Zhang, and F. A. Hamprecht, "Empowering multiple instance histopathology cancer diagnosis by cell graphs," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8674, pp. 228–235.
- [39] Y. Xu, J. Y. Zhu, E. Chang, and Z. Tu, "Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.
- [40] G. Zhang et al., "A Sparse Bayesian Multi-Instance Multi-Label Model for Skin Biopsy Image Analysis," *2013 IEEE Int. Conf. Bioinforma. Biomed.*, pp. 261–266, 2013.
- [41] S. SCOTT, J. ZHANG, and J. BROWN, "on Generalized Multiple-Instance Learning," *Int. J. Comput. Intell. Appl.*, vol. 05, no. 01, pp. 21–35, 2005.
- [42] M. L. Zhang, "Generalized multi-instance learning: Problems, algorithms and data sets," *Proc. 2009 WRI Glob. Congr. Intell. Syst. GCIS 2009*, vol. 3, pp. 539–543, 2009.
- [43] Q. Tao and S. Scott, "A Faster Algorithm for Generalized Multiple-instance Learning," *Seventeenth Annu. FLAIRS Conf.*, no. 1998, pp. 550–555, 2004.
- [44] Q. Tao, S. Scott, N. V. Vinodchandran, and T. T. Osugi, "SVM-based generalized multiple-instance learning via approximate box counting," in *Twenty-first international conference on Machine learning - ICML '04*, 2004, pp. 779–806.
- [45] Q. Tao, S. Scott, and N. Vinodchandran, "An extended kernel for generalized multiple-instance learning," in *16th IEEE International Conference on Tools with Artificial Intelligence*, 2004, pp. 272–277.
- [46] L. Jie and F. Orabona, "Learning from candidate labeling sets," in *Advances in Neural Information Processing Systems*, 2010, pp. 1–9.
- [47] International Association for Pattern Recognition, "Weakly Supervised Learning for Industrial Optical Inspection," 2007. [Online]. Available: <https://hci.iwr.uni-heidelberg.de/node/3616>.
- [48] "Python implementation of Balu Toolbox." [Online]. Available: <https://github.com/dipaco/balu-python>.
- [49] "scikit-image: Image processing in Python." [Online]. Available: <https://scikit-image.org/>.
- [50] "Mahotas: Computer Vision in Python." [Online]. Available: <https://mahotas.readthedocs.io/en/latest/>.
- [51] A. Kumar and G. K. H. Pang, "Defect detection in textured materials using gabor filters," *Ieee Trans. Ind. Appl.*, vol. 38, no. 2, pp. 425–440, 2002.
- [52] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray scale and rotation invariant texture classification with local binary patterns," *Comput. Vision-ECCV 2000*, vol. 24, no. 7, pp. 404–420, 2000.
- [53] K. S. R M Haralick, I Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [54] A. Ahmed et al., "Structured literature image finder: Parsing text and figures in biomedical literature," *J. Web Semant.*, vol. 8, no. 2–3, pp. 151–154, 2010.
- [55] A. Khotanzad and Y. H. Hong, "Invariant Image Recognition by Zernike Moments," *Ann. Oper. Res. Pattern Anal. Mach. Intell. IEEE Trans. Pattern Anal. Mach. Intell. J. Robot. Res. J. Robot. Res. J. ACM Networks I. J. Stoker*, vol. 12, no. 14, pp. 13–118, 1990.
- [56] N. Dalal and W. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.
- [57] C. M. B, J. Arrieta, M. Orozco-alzate, and J. Branch, "Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications," vol. 9423, pp. 724–731, 2015.



Jose Arrieta obtuvo su título de Ingeniería de Sistemas e Informática de la Universidad Nacional de Colombia sede Medellín (2015). Joven Investigador COLCIENCIAS (2016). Sus intereses de investigación están en las áreas de visión por computador, reconocimiento de patrones y aprendizaje de máquina.



Carlos Mera obtuvo su título de Magister en Ingeniería de Sistemas de la Universidad del Valle (2007) y realizó el Máster en Sistemas Inteligentes en la Universidad de Salamanca, España (2009). Obtuvo su título de Doctor en Ingeniería de la Universidad Nacional de Colombia (2017). Actualmente es profesor en el Instituto Tecnológico Metropolitano de Medellín (ITM).



Albeiro Espinosa Bedoya posee los títulos de Magister en Ingeniería de Sistemas (2004) y Doctor en Ingeniería – Sistemas e Informática (2014) de la Universidad Nacional de Colombia. Actualmente es docente de esa misma institución y está adscrito al Departamento de Ciencias de la Computación y de la Decisión.