

ALGORITMO SEMISUPERVISADO PARA EL ETIQUETADO DE IMÁGENES  
BASADO EN MÉTODOS DE APRENDIZAJE PROFUNDO

MIGUEL ALBERTO PLAZAS WADYNSKI

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERIAS FISICOMECHANICAS  
ESCUELA DE INGENIERIA DE SISTEMAS E INFORMATICA  
MAESTRÍA EN INGENIERIA DE SISTEMAS E INFORMATICA  
BUCARAMANGA

2022

ALGORITMO SEMISUPERVISADO PARA EL ETIQUETADO DE IMÁGENES  
BASADO EN MÉTODOS DE APRENDIZAJE PROFUNDO

MIGUEL ALBERTO PLAZAS WADYNSKI

Trabajo de grado para optar al título de  
Magister en Ingeniería de Sistemas e Informática

Director

Raúl Ramos-Pollán

Ph.D. en Ingeniería Informática

Codirector

Fabio Martinez Carrillo

Ph.D. en Ingeniería de Sistemas y computación

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2022

## **AGRADECIMIENTOS**

Me gustaría exaltar mi más sincero agradecimiento al profesor Fabio Martínez Carrillo por su valioso aporte y guía en el desarrollo de este trabajo de investigación. La paciencia, constancia y perseverancia son dignas de reconocimiento, así como sus valiosos consejos y recomendaciones. De igual forma, muchas gracias al profesor Raúl Ramos-Pollán por su guía, apoyo y esfuerzos.

De forma similar, me gustaría agradecer a los miembros del grupo de investigación Bivl2ab de la Universidad Industrial de Santander de Colombia por su apoyo y comprensión incondicional.

Finalmente, un agradecimiento especial a mis familiares y amigos que me han apoyado tanto durante toda esta gran aventura.

## CONTENIDO

	pág.
INTRODUCCIÓN.....	9
1. FUNDAMENTOS Y TRABAJO PREVIO.....	12
1.1 FUNDAMENTOS DE APRENDIZAJE SEMI-SUPERVISADO.....	12
1.1.1 Auto-entrenamiento.....	14
1.1.2 Co-entrenamiento.....	17
1.1.3 Co-aprendizaje democrático.....	19
1.2 TRABAJOS RELACIONADOS.....	20
2. PROBLEMA DE INVESTIGACIÓN.....	25
2.1 PREGUNTA DE INVESTIGACIÓN.....	27
3. OBJETIVOS.....	28
3.1 OBJETIVO GENERAL.....	28
3.2 OBJETIVOS ESPECÍFICOS.....	28
4. ENFOQUE PROPUESTO.....	29
4.1 ENFOQUE DE ENSEMBLE: UNA REPRESENTACIÓN PROFUNDA.....	30
4.1.1 ResNet152:.....	32
4.1.2 InceptionV3:.....	33
4.1.3 InceptionResNetV2:.....	34
4.2 ESQUEMA SEMI-SUPERVISADO.....	34
5. CONFIGURACIÓN EXPERIMENTAL.....	38
6. EVALUACIÓN Y RESULTADOS.....	43
7. CONCLUSIONES.....	53
BIBLIOGRAFÍA.....	56

## LISTA DE CUADROS

	pág.
Cuadro 1. Precisión – UCMerced.....	43
Cuadro 2. F1- Score – UCMerced.....	43
Cuadro 3. Recall – UCMerced.....	44
Cuadro 4. Precisión – AID .....	44
Cuadro 5. F1-Score - AID .....	44
Cuadro 6. Recall - AID.....	45
Cuadro 7. Precisión – WHU-RS19 .....	45
Cuadro 8. F1-Score – WHU-RS19 .....	45
Cuadro 9. Recall – WHU-RS19 .....	46
Cuadro 10. Precisión – NWPU-RESISC45 .....	46
Cuadro 11. F1-Score – NWPU-RESISC45 .....	47
Cuadro 12. Recall – NWPU-RESISC45 .....	47
Cuadro 13. Rendimiento del enfoque propuesto vs Estado del Arte .....	52

## LISTA DE FIGURAS

	pág.
Figura 1. Principio fundamental de aprendizaje semi-supervisado .....	13
Figura 2. Algoritmo 1 Auto-entrenamiento .....	15
Figura 3. Método semi-supervisado de auto-entrenamiento.....	15
Figura 4. Algoritmo 2 Co-entrenamiento .....	17
Figura 5. Método de co-entrenamiento .....	18
Figura 6. Algoritmo 3 Co-aprendizaje democrático .....	19
Figura 7. Pipeline del enfoque propuesto.....	35
Figura 8. Algoritmo 4 Aprendizaje semi-supervisado de decisiones ajustado.....	37
Figura 9. Muestras de los conjuntos de datos.....	40
Figura 10. División de los datos .....	41
Figura 11. Tiempo de entrenamiento del método propuesto .....	48
Figura 12. Muestras de matrices de confusión .....	49

## RESUMEN

**TÍTULO:** ALGORITMO SEMISUPERVISADO PARA EL ETIQUETADO DE IMÁGENES BASADO EN MÉTODOS DE APRENDIZAJE PROFUNDO\*

**AUTOR:** MIGUEL ALBERTO PLAZAS WADYNSKI\*\*

**PALABRAS CLAVE:** APRENDIZAJE SEMI-SUPERVISADO, DETECCIÓN REMOTA, TRANSFERENCIA DE APRENDIZAJE, MARCO DE TRABAJO DE ENSEMBLE

**DESCRIPCIÓN:** Las técnicas de aprendizaje semi-supervisado (SSL) exploran estrategias para el descubrimiento progresivo de la estructura oculta y latente de los datos. Para ello, estas estrategias hacen propagación de información supervisada sobre datos no etiquetados, que se utilizan posteriormente para reforzar el aprendizaje. Estos esquemas son beneficiosos en la teledetección, donde se agregan miles de imágenes nuevas todos los días y los resultados del etiquetado manual son prohibitivos. Este trabajo presenta un enfoque novedoso de aprendizaje profundo semi-supervisado basado en conjuntos que inicialmente toma un subconjunto de datos etiquetados  $\mathcal{D}_\ell$ , que representa la estructura latente de los datos, y propaga etiquetas de forma progresiva y automática desde un conjunto de datos de expansión sin etiquetar  $\mathcal{D}_u$ . La estrategia de ensamble es un conjunto de clasificadores cuyas predicciones se recopilan para derivar una predicción consolidada. Solo aquellos datos que tienen una predicción de alta confianza se consideran etiquetas recién generadas. El enfoque propuesto se validó exhaustivamente en cuatro conjuntos de datos públicos, logrando resultados apreciables en comparación con los métodos de última generación en la mayoría de las configuraciones evaluadas. Para todos los conjuntos de datos, el enfoque propuesto logró una puntuación de F1-score y un Recall de hasta un 90%, en promedio. El esquema SSL y recursivo también demostró una ganancia promedio de ~2% en la última etapa de entrenamiento en conjuntos de datos grandes.

---

\*Trabajo de grado

\*\*Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Maestría en Ingeniería de Sistemas e Informática. Director: Raúl Ramos-Pollán. Ph.D. en Ingeniería Informática. Codirector: Fabio Martínez Carrillo. Ph.D. en Ingeniería de Sistemas y computación.

## ABSTRACT

**TITLE:** SEMI-SUPERVISED ALGORITHM FOR IMAGE LABELING BASED ON DEEP LEARNING METHODS\*

**AUTHOR:** MIGUEL ALBERTO PLAZAS WADYNSKI\*\*

**KEY WORDS:** SEMI-SUPERVISED LEARNING, DEEP LEARNING, TRANSFER LEARNING, ENSEMBLE FRAMEWORK

**DESCRIPTION:** Semi-supervised learning (SSL) techniques explore the progressive discovery of the hidden latent data structure by propagating supervised information on unlabeled data which are thereafter used to reinforce learning. These schemes are beneficial in remote sensing, where thousands of new images are added every day, and manual labeling results are prohibitive. This work introduces a novel ensemble-based semi-supervised deep learning approach that initially takes a subset of labeled data  $\mathcal{D}_\ell$ , which represents the latent structure of the data, and progressively propagates labels automatically from an expanding set of unlabeled data  $\mathcal{D}_u$ . The ensemble is a set of classifiers whose predictions are collated to derive a consolidated prediction. Only those data having a high-confidence prediction are considered as newly generated labels. The proposed approach was exhaustively validated on four public datasets, achieving appreciable results compared to the state-of-the-art methods in most of the evaluated configurations. For all datasets, the proposed approach achieved a classification F1-score and recall of up to 90%, on average. The SSL and recursive scheme also demonstrated an average gain of ~2% at the last training stage in such large datasets.

---

\*Degree work

\*\*Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Raúl Ramos-Pollán, PhD. Co-advisor: Fabio Martínez, PhD.

## INTRODUCCIÓN

Las tecnologías actuales de radar para la detección remota y la observación de la Tierra producen miles de imágenes aéreas y satelitales con altas resoluciones espaciales, espectrales y temporales<sup>1</sup>. La clasificación de escenas de imágenes, inmerso en esquemas de detección remota de imágenes satelitales juega un papel crucial en una amplia gama de aplicaciones, como la determinación de la escena de la cobertura terrestre en la planificación urbana, la detección de peligros naturales<sup>2,3</sup>, protección ambiental, el mapeo de cultivos desde plataformas aéreas o espaciales, el mapeo de vegetación<sup>4,5,6</sup> y monitoreo del entorno entre muchos otros<sup>7,8</sup>.

---

<sup>1</sup> CHENG,Gong; HAN,Junwei and LU,Xiaoqiang. Remote Sensing Image Scene Classification: Benchmark and State of the Art. En: PROCEEDINGS OF THE IEEE. Oct.vol. 105, no. 10, p. 1865-1883

<sup>2</sup> MARTHA,Tapas Ranjan, et al. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 49, no. 12, p. 4928-4943

<sup>3</sup> CHENG,Gong, et al. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 34, no. 1, p. 45-59

<sup>4</sup> KIM,Minho; MADDEN,Marguerite and WARNER,Timothy A. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery. En: PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING. vol. 75, no. 7, p. 819-829

<sup>5</sup> LI,Xiaoxiao and SHAO,Guofan. Object-based urban vegetation mapping with high-resolution aerial photography as a single data source. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 34, no. 3, p. 771-789

<sup>6</sup> MISHRA,Niti B. and CREWS,Kelley A. Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with Random Forest. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 35, no. 3, p. 1175-1198

<sup>7</sup> GÓMEZ-CHOVA,Luis, et al. Multimodal classification of remote sensing images: A review and future directions. En: PROCEEDINGS OF THE IEEE. vol. 103, no. 9, p. 1560-1584

Actualmente, la clasificación de escenas de imágenes remotas es un desafío debido a la limitada resolución para representar un objeto de interés, junto con la variabilidad inherente entre clases e intraclasses. Además, las nuevas aplicaciones están en constante evolución para estas imágenes de observación de la Tierra, con una complejidad y variabilidad cada vez mayores, lo que genera problemas para representar objetos a partir de esas fuentes.

En el estado del arte se ha propuesto descriptores visuales convencionales para diferentes dominios de detección remota de imágenes satelitales para clasificar objetos; sin embargo, su eficacia se evalúa mediante el uso de conjuntos de datos pequeños y bien delimitados<sup>9</sup>. Más recientemente, se han aplicado representaciones convolucionales profundas en la detección remota de imágenes satelitales para obtener resultados de vanguardia. Sin embargo, estas estrategias supervisadas dependen en gran medida de grandes conjuntos de datos etiquetados, lo que constituye una limitación importante en la detección remota de imágenes satelitales<sup>10</sup>. Por otra parte, han surgido métodos alternativos que descubren la estructura latente de los datos mediante la propagación de información supervisada, pero que se limitan a estructuras visuales fijas, lo que constituye un desafío cuando se enfrentan nuevos datos no-etiquetados y variabilidad visual incontrolada.

---

<sup>8</sup> MA,Xinhua. Innovation-driven efficient development of the Longwangmiao Fm large-scale sulfur gas reservoir in Moxi block, Sichuan Basin. En: NATURAL GAS INDUSTRY B. vol. 3, no. 2, p. 99-107

<sup>9</sup> DALAL,Navneet and TRIGGS,Bill. Histograms of oriented gradients for human detection. 2005. p. 886-893

<sup>10</sup> ZHANG,Fan; DU,Bo and ZHANG,Liangpei. Saliency-guided unsupervised feature learning for scene classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 53, no. 4, p. 2175-2184

En este trabajo presentamos un enfoque semi-supervisado de ensemble que puede aprender una representación específica de detección remota de objetos y escenarios con sentido semántico sobre secuencias satelitales. El enfoque propuesto utiliza un conjunto de arquitecturas de aprendizaje profundo que se entrenaron inicialmente en un dominio general de imagen natural. Luego, se implementó una estrategia por transferencia de aprendizaje para ajustar progresivamente la representación visual al dominio de detección remota de imágenes satelitales<sup>11</sup>. Este trabajo consideró un esquema de extremo a extremo (*end-to-end*) que calcula la probabilidad de correspondencia de clases. Posteriormente, se llevó a cabo un proceso de decisión para predecir la clase. Una vez se propagan las etiquetas sobre un sub-conjunto sin anotaciones, se filtran aquellas que corresponden a una mayor confianza, las cuales serán utilizadas para re-entrenar y afinar la estrategia ensemble. Estas nuevas muestras ajustan la representación visual de detección remota de imágenes satelitales. El algoritmo propuesto entonces actúa de forma iterativa propagando etiquetas en nuevos conjuntos de datos y tomando las de mayor confianza para el entrenamiento. El método propuesto se evaluó utilizando diferentes configuraciones para evaluar su robustez al procesar un número reducido de datos etiquetados (~5% o ~10%).

El resto del trabajo está organizado en la exploración del estado del arte (Capítulo 2), la presentación del problema de investigación (Capítulo 3) y los objetivos planteados durante la propuesta (Capítulo 4), luego se hace una descripción detallada del método propuesto (Capítulo 5) y se presenta un esquema de validación y presentación de los resultados (Capítulo 7). Finalmente, en el Capítulo 8 se presentan algunas conclusiones y perspectivas del trabajo presentado.

---

<sup>11</sup> TAN,Chuanqi, et al. A Survey on Deep Transfer Learning. En: CoRR. vol. abs/1808.01974

# 1. FUNDAMENTOS Y TRABAJO PREVIO

## 1.1 FUNDAMENTOS DE APRENDIZAJE SEMI-SUPERVISADO

El aprendizaje semi-supervisado intenta descubrir la estructura latente de los datos a partir de representaciones parciales que se van complementando a través de la propagación de etiquetas. En este sentido el aprendizaje puede dividirse en inductivo y transductivo. El aprendizaje inductivo generaliza reglas y define un modelo de aprendizaje con las muestras de etiquetadas. Este modelo propaga las etiquetas a nuevas muestras, pero su base de conocimiento no se modifica<sup>12,13</sup>. Por otra parte, el aprendizaje transductivo<sup>14,15</sup> asocia por sus descriptores muestras no etiquetadas con las etiquetadas y luego construye un modelo general de aprendizaje con todas las muestras que tienen alta confianza de representación.

Formalmente, se tiene un conjunto de datos de entrenamiento  $\mathcal{D} = \{\mathcal{D}_\ell, \mathcal{D}_u\}$  con subgrupos definido para los datos etiquetados  $\mathcal{D}_\ell = \{(x_i, y_i)\}_{i=1}^{n_\ell}$  y un sub-conjunto no etiquetado, como:  $\mathcal{D}_u = \{x_i\}_{i=1}^{n_u}$ . La tarea de propagación de etiquetas es entonces definida por la función de aprendizaje:  $f: \mathcal{X} \rightarrow \mathcal{Y}, f \in \mathcal{F}$  que asocia a una representación particular  $\mathcal{X}$  una etiqueta respectiva  $\mathcal{Y}$ . En el aprendizaje semi-supervisado un pequeño conjunto de datos de entrenamiento  $\mathcal{D}_\ell$  esta etiquetado mientras que un gran conjunto de datos esta no-etiquetado ( $n_u \gg n_\ell$ ). Por lo

---

<sup>12</sup> LEE,Dong-Hyun. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. 2013. p. 2

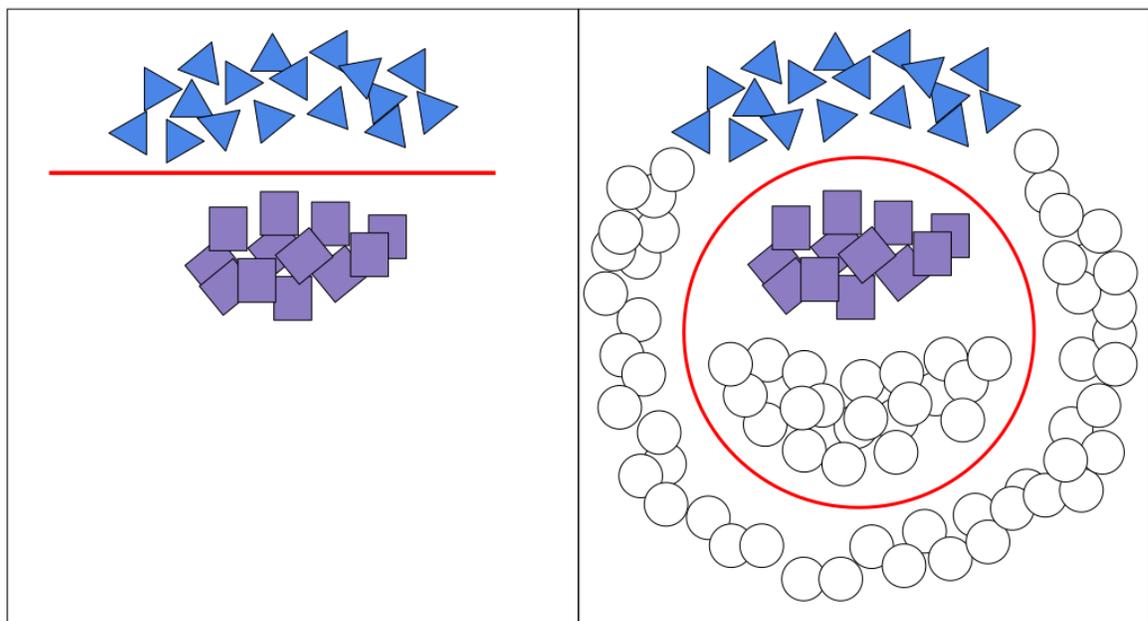
<sup>13</sup> SAJJADI,Mehdi; JAVANMARDI,Mehran and TASDIZEN,Tolga. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. 2016. p. 1163-1171

<sup>14</sup> ZHU,Xiaojin. Semi-Supervised Learning: From Gaussian Fields to Gaussian Processes Xiaojin Zhu John Lafferty Zoubin Ghahramani August 18, 2003.

<sup>15</sup> ZHOU,Dengyong, et al. Learning with local and global consistency. 2004. p. 321-328

tanto, el problema principal en el aprendizaje semi-supervisado consiste en definir pseudo-etiquetas o asociaciones para  $\mathcal{D}_u$  que permitan construir una representación completa de aprendizaje. Cada muestra de entrenamiento  $(x_i, y_i)$  se asume a ser muestreada independientemente de una distribución  $P(x, y), x \in \mathcal{X}, y \in \mathcal{Y}$  mientras las muestras no-etiquetadas son muestreadas de  $P(x) = \sum_y P(x, y)$ . La representación o propagación completa de las etiquetas permite aproximar la topología completa de los datos, variando las fronteras de separación, y permitiendo modelar de forma mas efectiva la variabilidad entre clases. Un ejemplo ilustrativo es mostrado en la Figura 1.

Figura 1. Principio fundamental de aprendizaje semi-supervisado



Fuente: elaboración propia.

En los esquemas comunes de aprendizaje semi-supervisado se utilizan diferentes estrategias de propagación y asociación de etiquetas. Además, se consideran las

siguientes premisas en el espacio donde coexisten los datos de entrenamiento<sup>16,17</sup>:

- **Suposición de suavidad:** Si dos puntos en una región de alta densidad están cerca, entonces sus salidas correspondientes también están cerca.
- **Suposición de agrupación:** Si dos puntos están en el mismo grupo, es probable que sean de la misma clase.
- **Suposición del colector:** Los datos de alta dimensión se encuentran dentro de una variedad de baja dimensión.

A partir de estas premisas se han destacado tres diferentes metodologías para realizar los esquemas de aprendizaje semi-supervisado, como son: el auto-entrenamiento, co-entrenamiento, y co-aprendizaje democrático. Cada una de estas metodologías son descritas a continuación, así como también son resaltados algunos ejemplos destacables en la literatura:

**1.1.1 Auto-entrenamiento.** Este tipo de estrategias semi-supervisadas parten de un modelo parcial de aprendizaje  $m_0$  utilizando los datos etiquetados  $\mathcal{D}_\ell$ . Seguidamente, para cada muestra sin-etiquetar  $x \in \mathcal{D}_u$  se utiliza el modelo entrenado  $m$  para proponer una pseudo-etiqueta  $m(x)$  la cual corresponde a la máxima probabilidad ( $\max(m(x))$ ) de las clases propuestas por el modelo. Si dicha probabilidad es mayor a un umbral de alta confianza  $\tau$  entonces dicha pseudo-etiqueta se considera de alta confianza  $(x, p(x))$  y es almacenada en el conjunto de muestras etiquetadas  $\mathcal{D}_\ell$ . El umbral de alta confianza  $\tau$  se configura

---

<sup>16</sup> CHAPELLE, Olivier; SCHOLKOPF, Bernhard and ZIEN, Alexander. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. En: IEEE TRANSACTIONS ON NEURAL NETWORKS. vol. 20, no. 3, p. 542

<sup>17</sup> SUBRAMANYA, A. and TALUKDAR, P. P. Graph-Based Semi-Supervised Learning. Morgan & Claypool, 2014.

inicialmente. En la Figura 2 se visualiza el Algoritmo 1, el cual describe el proceso general del auto-entrenamiento y en la Figura 3 se ilustra este método<sup>18,19</sup>

Figura 2. Algoritmo 1 Auto-entrenamiento

---

**Algoritmo 1** Auto-entrenamiento

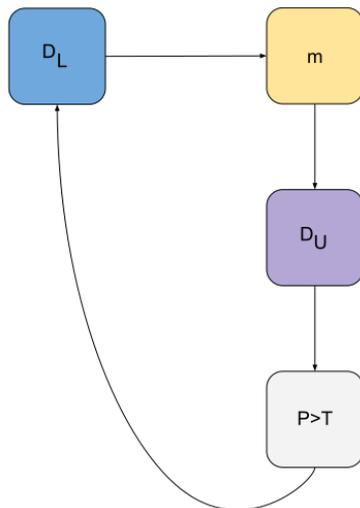
---

```
 $m_0 \leftarrow \text{entrenar\_modelo}(\mathcal{D}_l)$ 
repeat
   $m \leftarrow \text{entrenar\_modelo}(\mathcal{D}_l)$ 
  for  $x \in \mathcal{D}_u$  do
    if  $\max m(x) > \tau$  then
       $\mathcal{D}_l \leftarrow \mathcal{D}_l \cup \{(x, p(x))\}$ 
until las predicciones no sean confiables
```

---

Fuente: elaboración propia.

Figura 3. Método semi-supervisado de auto-entrenamiento



Fuente: elaboración propia.

---

<sup>18</sup> YAROWSKY,David. Unsupervised word sense disambiguation rivaling supervised methods. 1995. p. 189-196

<sup>19</sup> MCCLOSKEY,David; CHARNIAK,Eugene and JOHNSON,Mark. Effective self-training for parsing. 2006. p. 152-159

Existen numerosas estrategias que han utilizado este principio para propagar etiquetas en diferentes dominios específicos. Por ejemplo, en HAN, Wei et al.<sup>20</sup> se aplicó el auto-entrenamiento en un dominio de etiquetado y clasificación de objetos en imágenes satelitales. En este trabajo, las imágenes eran caracterizadas con vectores embebidos obtenidos de redes convolucionales del estado del arte, normalmente la VGG y ResNet. Luego estos vectores embebidos servían para entrenar una máquina de soporte vectorial (SVM) con un kernel Lineal que lograba una apropiada partición del espacio. Este trabajo logra resultados sobresalientes (~ 91.57%) en la tarea de clasificación, pero su caracterización pierde de vista el modelamiento específico del dominio, teniendo en cuenta que las características aprendidas desde un dataset general, no son acotadas o re-entrenadas para las imágenes satelitales. También, GU,Xiaowei y ANGELOV,Plamen P.<sup>21</sup> utilizaron redes previamente entrenadas, en dominios generales, como descriptores de imágenes satelitales. En este caso, se hace una aproximación local, caracterizando parches de las imágenes con capas convolucionales de primer orden. En este trabajo, el proceso semi-supervisado es un sistema profundo basado en un conjunto de reglas fuzzy, que tienen incluso la capacidad de generar nuevas clases, durante la etapa de aprendizaje semi-supervisado. Estas estrategias están limitadas al umbral de confianza definido y sujetas a un único modelo para lograr representar aprendizajes parciales. Esta dependencia puede limitar las fronteras definidas entre cada instancia. Además, según la efectividad de la clasificación está sujeto a una única hipótesis de partición de los datos.

---

<sup>20</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

<sup>21</sup> GU,Xiaowei and ANGELOV,Plamen P. A semi-supervised deep rule-based approach for remote sensing scene classification. 2019. p. 257-266

**1.1.2 Co-entrenamiento.** Las estrategias de co-entrenamiento<sup>22</sup> mitigan la dependencia de un único modelo para plantear las fronteras de separación entre clases. En este caso, los datos de entrenamiento etiquetado  $\mathcal{D}_\ell$  se dividen en dos sub-conjuntos  $\mathcal{D}_{\ell_1}$  y  $\mathcal{D}_{\ell_2}$ , los cuales sirven como recurso de inicialización de dos diferentes modelos de aprendizaje  $m_1$  y  $m_2$ . Entonces las muestras no etiquetadas son mapeadas a cada uno de los modelos, y según el grado de confianza establecido, las muestras son usadas para actualizarlos. Por ejemplo, las muestras cuya máxima confianza de predicción del modelo  $m_1$  superen un umbral  $\tau$  son adicionadas al sub-conjunto de muestras etiquetadas  $\mathcal{D}_{\ell_2}$ . En la Figura 4 se ilustra el Algoritmo 2 se detalla un pseudocódigo de la estrategia de co-entrenamiento y en la Figura 5 se ilustra este método.

Figura 4. Algoritmo 2 Co-entrenamiento

---

**Algoritmo 2** Co-entrenamiento

---

```

repeat
   $m_1 \leftarrow \text{entrenar\_clasificador}(\mathcal{D}_{l_1})$ 
   $m_2 \leftarrow \text{entrenar\_clasificador}(\mathcal{D}_{l_2})$ 
  for  $x \in \mathcal{D}_u$  do
    if  $\max m_1(x) > \tau$  and  $\max m_2(x) < \tau$  then
       $\mathcal{D}_{l_2} \leftarrow \mathcal{D}_{l_2} \cup \{(x, p_1(x))\}$ 
    if  $\max m_2(x) > \tau$  and  $\max m_1(x) < \tau$  then
       $\mathcal{D}_{l_1} \leftarrow \mathcal{D}_{l_1} \cup \{(x, p_2(x))\}$ 
  until las predicciones no son confiables
  basándose en un clasificador

```

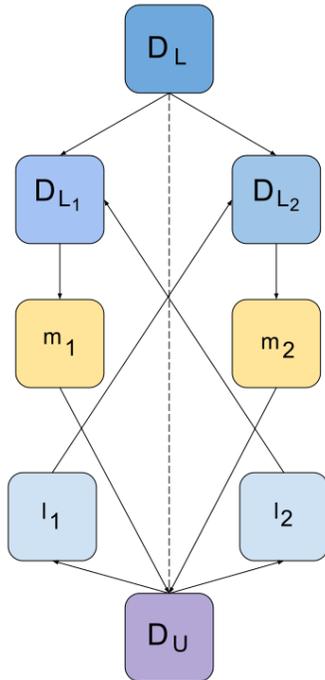
---

Fuente: elaboración propia.

---

<sup>22</sup> BLUM, Avrim and MITCHELL, Tom. Combining labeled and unlabeled data with co-training. 1998. p. 92-100

Figura 5. Método de co-entrenamiento



Fuente: elaboración propia.

Un ejemplo particular del principio de co-entrenamiento, es propuesto en HAN, Wei et al.<sup>23</sup> donde se entrenan SVM independientes y con posibilidad de hipótesis de separación diferente, según la definición del kernel. Cada uno de estos modelos establece una función de separación para los datos. Dado un nuevo conjunto no etiquetado, las muestras con mayor confianza sirven para robustecer y definir con mayor claridad las fronteras de cada modelo. Además, a través de la propagación se puede definir la mejor hipótesis de separación de clases, expresada en cada SVM. Estos métodos sin embargo permanecen restringidos a los umbrales que definen la confianza de cada muestra. Además, sus descriptores de representación en general son genéricos perdiendo poder de representación en el dominio específico.

---

<sup>23</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

**1.1.3 Co-aprendizaje democrático.** En un caso particular del co-entrenamiento, se pueden definir múltiples modelos, bajo diferentes hipótesis de representación, y el grado de confianza de una muestra en particular es asociada por un proceso de votación entre estos esquemas de aprendizaje. En este caso, la totalidad de muestras  $\mathcal{D}_\ell$  son utilizadas para inicializar  $m_i$  modelos. Cada muestra  $x$  en  $x \in \mathcal{D}_u$  se mapea a cada modelo  $m_i$ . Entonces se define el subconjunto  $\mathcal{M}$  que contiene los modelos que coinciden en la predicción de la clase  $j$ . Si la sumatoria de los intervalos de confianza de correspondencia a la clase  $j$  es superior a los de sumatoria de no correspondencia, entonces la etiqueta es agregada al conjunto  $\mathcal{D}_\ell$ . En la Figura 6 se ilustra el Algoritmo 3 se detalla un pseudocódigo de la estrategia de co-aprendizaje democrático.

Figura 6. Algoritmo 3 Co-aprendizaje democrático

---

**Algoritmo 3** Co-aprendizaje democrático

---

```

repeat
  for  $i \in \{1..n\}$  do
     $m_i \leftarrow \text{entrenar\_modelo}(\mathcal{D}_l)$ 
  for  $x \in \mathcal{D}_u$  do
    for  $j \in \{1..C\}$  do
       $M \leftarrow \{i \mid p_i(x) = j\}$ 
      if  $|M| > n/2$  and  $\sum_{i \in M} w_i > \sum_{i \notin M} w_i$  then
         $\mathcal{D}_l \leftarrow \mathcal{D}_l \cup \{(x, j)\}$ 
  until ningún modelo  $m_i$  cambia
  aplicar voto mayoritario ponderado  $m_i$ 

```

---

Fuente: elaboración propia

Un ejemplo particular de este enfoque es reportado en Yalniz et. al.<sup>24</sup> utilizando el paradigma profesor-estudiante, sobre un dominio específico de imágenes

---

<sup>24</sup> YALNIZ,I. Zeki, et al. Billion-scale semi-supervised learning for image classification. En: arXiv PREPRINT arXiv:1905.00546.

satelitales y el dataset YFCC100m<sup>25</sup>. En este caso, se entrena una red profesor con los datos  $\mathcal{D}_\ell$ . Luego se propagan las etiquetas de  $\mathcal{D}_u$  en esta red, y se seleccionan las  $k$  con mayor confianza. Estas muestras de alta confianza son utilizadas para entrenar una red *estudiante*, la cual es refinada con las muestras etiquetadas  $\mathcal{D}_\ell$ . Este tipo de esquema ha reportado un rendimiento superior a la arquitectura ResNet50, usando un método puramente supervisado. Este esquema sin embargo requiere una cantidad considerable de datos  $\mathcal{D}_u$ , aproximadamente 100 millones de nuevas imágenes.

Estos esquemas han resultado ser eficaces en diferentes dominios, pero siguen siendo dependientes de una amplia cantidad de información para definir la estructura de los datos y propagar las nuevas etiquetas. Sin embargo, existen dominios, donde se requieren los esquemas supervisados, pero no se cuenta con una suficiente población muestral para aprender sobre el dominio. Por lo tanto, aún hoy en día se trabaja en esquemas de aprendizaje más compactos que logren una representación apropiada y una propagación de etiquetas en un nuevo dominio, pero utilizando información limitada.

## 1.2 TRABAJOS RELACIONADOS

El aprendizaje semi-supervisado intenta descubrir la estructura latente de los datos a partir de representaciones parciales que se van complementando a través de la propagación de etiquetas. En este sentido el aprendizaje puede dividirse en inductivo y transductivo. La clasificación de escenas de imágenes remotas se ha abordado mediante descriptores visuales de ingeniería y representaciones de aprendizaje profundo, incluidos métodos SSL. Los trabajos iniciales han abordado la clasificación de imágenes en la detección remota de imágenes satelitales

---

<sup>25</sup> KALKOWSKI, Sebastian, et al. Real-time analysis and visualization of the YFCC100M dataset. 2015. p. 25-30

utilizando ingeniería de características las cuales modelan señales individuales asociadas con las clases. Tales señales se basan en general en histogramas de color<sup>26</sup>, descriptores de textura<sup>27,28,29</sup>, y también incluyen un análisis de detalle local de los puntos SIFT<sup>30</sup> o histograma de gradientes orientados (HOG)<sup>31</sup>. Estos métodos basados en ingeniería de características han sido abordados únicamente en conjuntos limitados de datos y su explotación se encuentra restringida a la validación en escenas con poca variabilidad de primitivas visuales.

Los métodos de aprendizaje profundo han surgido como una alternativa robusta para aprender representaciones visuales complejas, utilizando grandes conjuntos de datos. De la familia de métodos supervisados inspirados en esquemas de transferencia de aprendizaje se encuentra en la literatura trabajos como el de Grant et. al} los cuales adaptaron una red neuronal convolucional previamente entrenada (ResNet50, GoogleNet y CaffeNet) para la clasificación de escenas de imágenes satelitales en conjuntos de datos clásicos tales como UCMerced<sup>32</sup>. En

---

<sup>26</sup> SWAIN,Michael J. and BALLARD,Dana H. Color indexing. En: INTERNATIONAL JOURNAL OF COMPUTER VISION. vol. 7, no. 1, p. 11-32

<sup>27</sup> HARALICK,Robert M.; SHANMUGAM,Karthikeyan and DINSTEIN,Its' Hak. Textural features for image classification. En: IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. no. 6, p. 610-621

<sup>28</sup> JAIN,Anil K.; RATHA,Nalini K. and LAKSHMANAN,Sridhar. Object detection using Gabor filters. En: PATTERN RECOGNITION. vol. 30, no. 2, p. 295-309

<sup>29</sup> OJALA,Timo; PIETIKAINEN,Matti and MAENPAA,Topi. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. En: IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. vol. 24, no. 7, p. 971-987

<sup>30</sup> LOWE,David G. Distinctive image features from scale-invariant keypoints. En: INTERNATIONAL JOURNAL OF COMPUTER VISION. vol. 60, no. 2, p. 91-110

<sup>31</sup> DALAL,Navneet and TRIGGS,Bill. Histograms of oriented gradients for human detection. 2005. p. 886-893

<sup>32</sup> SCOTT,Grant J., et al. Training deep convolutional neural networks for land–cover classification of high-resolution imagery. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 14, no. 4, p. 549-553

este trabajo se sigue un esquema clásico de transferencia, haciendo un ajuste iterativo, sobre capas superiores, para la adaptación del dominio de imágenes satelitales. De forma similar, Zhan et. al propuso una adaptación de dominio, alineando distribuciones de características como un paso intermedio no-supervisado, calculado a partir embebidos extraídos de arquitecturas tales como: ResNet50 y Alexnet. En este caso, cada imagen es representado como los vectores embebidos de dos redes del estado del arte, los cuales proyectan descriptores embebidos con respecto a cada imagen observada. Una vez acotada la descripción por estos vectores, los autores proponen una estrategia de alineamiento de características profundas, donde se conservan únicamente las características de mayor grado de correlación entre los vectores embebidos correspondientes. Finalmente, entrenan un clasificador NN (Nearest Neighbor) utilizando la nueva representación de características alineada<sup>33</sup>.

Por otro lado, inspirados en la familia de métodos supervisados basados en características globales y locales, Qiqi et. al proponen una estrategia de fusión de características locales y globales. Desde una perspectiva local utilizan la combinación de ingeniería de características como la desviación media y estándar (MSD) de las imágenes, así como también, la transformación de función invariante de escala (SIFT). Desde una perspectiva global utilizan características profundas extraídas a partir de una arquitectura CaffeNet. Después de consolidar estas estrategias global y local con dichas características entrenan una Máquina de Soporte Vectorial (SVM) para realizar la tarea de clasificación<sup>34</sup>. No obstante, estos enfoques solo utilizan vectores de embebidos extraídos de conjuntos de datos generales, lo cual limita el modelado de variabilidad de las aplicaciones de

---

<sup>33</sup> ZHANG,Jun, et al. Domain Adaptation Based on Correlation Subspace Dynamic Distribution Alignment for Remote Sensing Image Scene Classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

<sup>34</sup> ZHU,Qiqi, et al. A deep-local-global feature fusion framework for high spatial resolution imagery scene classification. En: REMOTE SENSING. vol. 10, no. 4, p. 568

detección remota de imágenes satelitales. A pesar de los avances en tales enfoques supervisados, estos métodos requieren un número significativo de imágenes manualmente etiquetadas para construir representaciones visuales robustas, definir límites sólidos entre clases y obtener una representación confiable<sup>35</sup>. Adicionalmente, estos conjuntos de datos especializados requieren personal experimentado para realizar las tareas de etiquetado.

En la detección remota de imágenes satelitales también han surgido varios enfoques basados en esquemas semi-supervisados que utilizan estrategias de auto-entrenamiento y co-entrenamiento. De la familia de métodos basados en auto-entrenamiento encontramos a Dai et. al<sup>36</sup> quien propone la extracción de vectores embebidos tomados de la arquitectura ResNet50 bajo esquemas semi-supervisados. Estos vectores se utilizan en clasificadores clásicos de aprendizaje de máquina, incluida la máquina de soporte vectorial (SVM), K-vecinos más cercanos (K-NN) y la regresión logística (LR). Para la propagación de etiquetas, se utilizan las muestras que maximizan la distancia entre clases y a estas se les asigna una misma clase. En cuanto a la familia de métodos basados en la estrategia de transferencia de aprendizaje bajo un esquema semi-supervisado de auto-entrenamiento Li et. al propuso un enfoque basado en el principio de aprendizaje de dominio asociativo para mapear la distribución de conocimiento de un conjunto de entrenamiento a un conjunto sin etiquetar. Aquí, se usa un conjunto de vectores embebidos para aprender una distribución de datos entre clases, la cual es reforzada en un nuevo conjunto sin etiquetar para recuperar muestras

---

<sup>35</sup> ZHANG,Fan; DU,Bo and ZHANG,Liangpei. Saliency-guided unsupervised feature learning for scene classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 53, no. 4, p. 2175-2184

<sup>36</sup> DAI,Xueyuan, et al. Semi-Supervised Scene Classification for Remote Sensing Images Based on CNN and Ensemble Learning. IEEE, Jul 2018. p. 4732-4735

cercanas que el modelo entrenado puede admitir<sup>37</sup>. La hipótesis subyacente de este enfoque es prometedora, pero el modelado de variabilidad es limitado para una distribución de dominio particular, con resultados desalentadores en conjuntos de datos públicos. Otros enfoques han utilizado representaciones visuales de grafos para definir estructuras de datos, propagar etiquetas y preservar relaciones entre clases<sup>38</sup>. Este enfoque es interesante, pero permanece restringido a la admisión de variabilidad de las representaciones visuales de las mismas clases<sup>39</sup>. En cuanto a la familia de métodos de auto-entrenamiento y co-entrenamiento Han et. al<sup>40</sup> proponen un método interesante basado en esencialmente en la estrategia de extracción de características profundas utilizando arquitecturas como la ResNet y VGG y la utilización de estas características profundas para el entrenamiento de múltiples máquinas de soporte vectorial (SVM) bajo los principios de co-entrenamiento para efectuar el proceso de etiquetado semi-supervisado. Estos métodos muestran una tendencia en partir de representaciones robustas en dominios de imágenes naturales y buscar una proyección hacia dominios con muestras más restringidas. A pesar de los avances logrados en el estado del arte, en general los mecanismos de adaptación siguen siendo restringidos, con respecto al manejo de la variabilidad intra e inter-clase. Una alternativa para abordar esta variabilidad puede ser construir esquemas auto-supervisados o semi-supervisados que descubran la topología de los datos sin hacer acotaciones o mantener sesgos por las anotaciones.

---

<sup>37</sup> LI, Youyou, et al. Semi-Supervised Remote Sensing Classification Via Associative Transfer. 2018. p. 2475-2478

<sup>38</sup> ALHICHRI, Haikel, et al. Tile-based semisupervised classification of large-scale VHR remote sensing images. En: JOURNAL OF SENSORS. vol. 2018

<sup>39</sup> MEIER, Raphael, et al. Patient-specific semi-supervised learning for postoperative brain tumor segmentation. 2014. p. 714-721

<sup>40</sup> HAN, Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

## 2. PROBLEMA DE INVESTIGACIÓN

Los modelos de aprendizaje, hoy en día, resultan ser efectivos según el conjunto de muestras que se usen para su entrenamiento<sup>41</sup>. Un mayor número de muestras en un problema supervisado por lo general permite modelar la variabilidad de cada clase, y definir apropiadamente las fronteras en un problema multiclase<sup>42</sup>. Sin embargo, el etiquetado de dichos datos es tedioso, sujeto a errores y a la variabilidad según los actores que participan en esta tarea<sup>43</sup>. En algunos dominios específicos, como en medicina y evaluación de procesos químicos, la marcación de estos datos requiere observadores expertos, lo cual demanda grandes recursos para el desarrollo de esta actividad<sup>44</sup>. Otras tareas, como las observaciones físicas, cada día reciben nuevos y numerosos datos, lo cual hace imposible la tarea de etiquetado y dificulta entender la topología donde se representan cada una de las instancias que se quiere modelar<sup>45</sup>.

El utilizar métodos de aprendizaje semi-supervisado junto con métodos de aprendizaje profundo ha demostrado resultados destacables para reconocer la estructura parcial de los datos y recursivamente propagar etiquetas sobre un

---

<sup>41</sup> YUAN,Qiangqiang, et al. Deep learning in environmental remote sensing: Achievements and challenges. En: REMOTE SENSING OF ENVIRONMENT. vol. 241, p. 111716

<sup>42</sup> ANUSE,Alwin and VYAS,Vibha. A novel training algorithm for convolutional neural network. En: COMPLEX & INTELLIGENT SYSTEMS. vol. 2, no. 3, p. 221-234

<sup>43</sup> ARVANITI,Eirini, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. En: SCIENTIFIC REPORTS. vol. 8, no. 1, p. 1-11

<sup>44</sup> BROSCH, Tom and TAM, Roger. Efficient training of convolutional deep belief networks in the frequency domain for application to high-resolution 2D and 3D images. En: NEURAL COMPUTATION. vol. 27, no. 1, p. 211-227

<sup>45</sup> CASTELLUCCIO, Marco, et al. Land use classification in remote sensing images by convolutional neural networks. En: arXiv PREPRINT arXiv:1508.00092.

nuevo conjunto de datos<sup>46</sup>. A pesar de la relevancia de las metodologías semi-supervisadas de aprendizaje aún existen diversos desafíos para ser abordados en el dominio científico. Por ejemplo, el uso en el entrenamiento de nuevos datos etiquetados automáticamente depende de un umbral, el cual en muchos de los casos es forzado de manera fija<sup>47</sup>. Además, cada una de las muestras son mapeadas a dominios generales de representación, por ejemplo, mapeando las imágenes a estructuras convoluciones que han sido entrenadas para imágenes naturales. Estos descriptores profundos, pierden descripción del dominio, y pueden hacer que la tarea de aprendizaje sea más costosa<sup>48</sup>. Nuevas tendencias han resultado del co-entrenamiento de diferentes modelos, los cuales logran un acuerdo a través de procesos típicos de votación. Estas estrategias han resultado efectivas, superando incluso, arquitecturas entrenadas bajo esquemas supervisados<sup>49</sup>. Sin embargo, la estructuración de las fronteras en este dominio requiere un muy amplio conjunto de imágenes etiquetadas y no-etiquetadas para aprender una representación efectiva. En dominios especializados, donde el número de muestras es limitado, estos modelos podrían fallar para lograr una representación semi-supervisada. Es particularmente destacable que el dominio de las imágenes satelitales resulta ser un buen candidato para desarrollar el

---

<sup>46</sup> FENG,Ziyi; HUANG,Guanhua and CHI,Daocai. Classification of the Complex Agricultural Planting Structure with a Semi-Supervised Extreme Learning Machine Framework. En: REMOTE SENSING. vol. 12, no. 22, p. 3708

<sup>47</sup> LIU,Jiantao, et al. Urban green plastic cover mapping based on VHR remote sensing images and a deep semi-supervised learning framework. En: ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION. vol. 9, no. 9, p. 527

<sup>48</sup> LIU,Yanfei; ZHONG,Yanfei and QIN,Qianqing. Scene classification based on multiscale convolutional neural network. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 56, no. 12, p. 7109-7121

<sup>49</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

estudio propuesto debido a que hay potencialmente muchos más datos<sup>50,51</sup>, sin embargo, en ciertas aplicaciones hay una carencia de etiquetas en las mismas. Por cual planteamos la pregunta de investigación:

## **2.1 PREGUNTA DE INVESTIGACIÓN**

¿Cuáles son las condiciones para utilizar aprendizaje semi-supervisado basado en aprendizaje profundo en dominios con un número de muestras etiquetadas limitado?

---

<sup>50</sup> HUANG, Yanbo, et al. Agricultural remote sensing big data: Management and applications. En: JOURNAL OF INTEGRATIVE AGRICULTURE. vol. 17, no. 9, p. 1915-1931

<sup>51</sup> MA, Lei, et al. Deep learning in remote sensing applications: A meta-analysis and review. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 152, p. 166-177

## **3. OBJETIVOS**

### **3.1 OBJETIVO GENERAL**

Desarrollar un modelo de aprendizaje semi-supervisado para el etiquetado automático de por lo menos el dominio de imágenes satelitales con datos limitados.

### **3.2 OBJETIVOS ESPECÍFICOS**

- Seleccionar un conjunto de datos correspondiente por lo menos al dominio de imágenes satelitales.
- Identificar un esquema de representación y entrenamiento del conjunto de imágenes seleccionado
- Proponer una estrategia de integración semi-supervisada para la propagación de etiquetas
- Evaluar condiciones de la estrategia propuesta y su impacto en el desempeño del modelo
- Validar estrategia propuesta usando métricas de clasificación

## 4. ENFOQUE PROPUESTO

Los algoritmos SSL han surgido como una herramienta poderosa para construir representaciones de datos parciales y dinámicas que mejoran el aprendizaje al propagar el conocimiento sobre datos sin etiquetar<sup>52</sup>. Esta metodología permite aprender sobre un pequeño conjunto de etiquetas y luego propagar etiquetas de forma incremental para volver a entrenar los modelos y lograr la solidez de las representaciones visuales<sup>53</sup>. Formalmente, los datos de entrenamiento se pueden expresar como  $\mathcal{D} = \{\mathcal{D}_\ell, \mathcal{D}_u\}$ , donde  $\mathcal{D}_\ell = \{(x_l^{(i)}, y_l^{(i)})\}_{i=1}^{n_l}$  es el conjunto etiquetado y  $\mathcal{D}_u = \{x_u^{(i)}\}_{i=1}^{n_u}$  es el conjunto sin etiquetar. Aquí,  $n_u$  y  $n_l$  son las muestras etiquetadas y sin etiquetar, respectivamente, con  $n_u \gg n_l$ .

El SSL tiene como objetivo definir pseudo-etiquetas de alta confianza en  $\mathcal{D}_u$  usando el modelo entrenado en  $\mathcal{D}_\ell$ . Las imágenes recién etiquetadas se utilizan para volver a entrenar y ajustar la topología de la representación de  $\mathcal{D}$ . La estructura general del enfoque propuesto se ilustra en la Figura 7. En las siguientes subsecciones se proporciona una descripción detallada de cada etapa propuesta. El contenido completo de este capítulo ha sido publicado en Journal of Applied Remote Sensing<sup>54</sup>. El cual es un Journal destacable debido a su categoría de A2 según Publindex, un Impact Factor de 1.530 y Q2 según Scimago Journal & Country Rank.

---

<sup>52</sup> BAI,Wenjia, et al. Semi-supervised learning for network-based cardiac MR image segmentation. 2017. p. 253-260

<sup>53</sup> TRIGUERO,Isaac; GARCÍA,Salvador and HERRERA,Francisco. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. En: KNOWLEDGE AND INFORMATION SYSTEMS. vol. 42, no. 2, p. 245-284

<sup>54</sup> PLAZAS,Miguel; RAMOS-POLLÁN,Raúl and MARTÍNEZ,Fabio. Ensemble-based approach for semisupervised learning in remote sensing. En: JOURNAL OF APPLIED REMOTE SENSING. vol. 15, no. 3, p. 1 – 16

#### 4.1 ENFOQUE DE ENSEMBLE: UNA REPRESENTACIÓN PROFUNDA

A pesar de su robustez reportada en la literatura, las representaciones visuales aprendidas de los enfoques de aprendizaje profundo dependen en gran medida de los datos de entrenamiento, e incluso la misma arquitectura puede variar según los parámetros utilizados durante la etapa de aprendizaje. En particular, estas arquitecturas no pueden garantizar la mejor hipótesis multiclase debido al requisito de un conjunto adecuado de imágenes satelitales correctamente etiquetadas. El SSL tiene como objetivo definir pseudo-etiquetas de alta confianza en  $\mathcal{D}_u$  usando el modelo entrenado en  $\mathcal{D}_\ell$ . Para lo cual, las imágenes recién etiquetadas son utilizadas para re-entrenar y ajustar la representación topológica de  $\mathcal{D}$ . Para lograr esto, se adoptó una metodología de ensemble desde  $\mathcal{M}$  redes convolucionales que en un enfoque de decisión tardía votan por una clase específica de detección remota de imágenes satelitales. Esto requirió tomar hipótesis de diferentes redes para tratar las representaciones visuales partiendo de una configuración de dominio general. Sin embargo, la representación inicial que utiliza estos pesos podría estar lejos del dominio de la detección remota de imágenes satelitales, y la frontera de decisión de las clases podría no delimitar correctamente aquellas clases con patrones de textura similares. Posteriormente, se abordó una representación ajustada al problema de la detección remota de imágenes satelitales utilizando el principio de transferencia de aprendizaje. La transferencia de aprendizaje (TL) es una técnica ampliamente conocida que permite ajustar representaciones de imágenes generales a un dominio específico (en nuestro caso, imágenes satelitales)<sup>55</sup>. Esta metodología de ensemble tiene como objetivo converger en una representación de detección remota de imágenes satelitales específica mediante la predicción de clases basada en la estrategia de decisión

---

<sup>55</sup> TAN,Chuanqi, et al. A Survey on Deep Transfer Learning. En: CoRR. vol. abs/1808.01974

tardía y el uso de muestras pseudo-etiquetadas de alta confianza en el refinamiento de las arquitecturas implicadas en el ensemble.

El método propuesto está compuesto por redes convolucionales robustas, entrenadas en etapas progresivas, para abordar el problema de la varianza por clase. Siguiendo un esquema de transferencia de aprendizaje, cada modelo  $m_k \in \mathcal{M}$  se adaptó primero al nuevo problema de clasificación. Específicamente, el enfoque propuesto utilizó tres redes neuronales convolucionales robustas: ResNet152, InceptionV3 e InceptionResNetV2, que presentaban capacidades de representación adecuadas en aplicaciones de detección remota<sup>56,57,58,59,60</sup>. Además, se adaptó una capa totalmente conectada de acuerdo con el número deseado de clases en nuestro dominio. En el paso de entrenamiento, las muestras  $D_{l_{train}}$  se utilizaron para ajustar el dominio de las imágenes naturales de ImageNet al dominio de la detección remota de imágenes satelitales de cada arquitectura  $m_k$  CNN. En el paso de etiquetado, las muestras  $D_{u_s}$  se utilizaron para ajustar la representación de las arquitecturas  $m_{k_s}^\dagger$  CNN en el dominio de la detección remota de imágenes satelitales. Para cada  $m_{k_s}^\dagger$  arquitectura CNN inferida de cada muestra no-etiquetada  $u_{k_s}$  una clase con una cierta probabilidad asociada  $p_{k_s}$ . A

---

<sup>56</sup> CHENG,Gong, et al. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 34, no. 1, p. 45-59

<sup>57</sup> MISHRA,Niti B. and CREWS,Kelley A. Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with Random Forest. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 35, no. 3, p. 1175-1198

<sup>58</sup> GÓMEZ-CHOVA,Luis, et al. Multimodal classification of remote sensing images: A review and future directions. En: PROCEEDINGS OF THE IEEE. vol. 103, no. 9, p. 1560-1584

<sup>59</sup> MA,Xinhua. Innovation-driven efficient development of the Longwangmiao Fm large-scale sulfur gas reservoir in Moxi block, Sichuan Basin. En: NATURAL GAS INDUSTRY B. vol. 3, no. 2, p. 99-107

<sup>60</sup> DALAL,Navneet and TRIGGS,Bill. Histograms of oriented gradients for human detection. 2005. p. 886-893

continuación, se llevó a cabo un proceso de decisión para determinar la clase asignada a cada muestra no etiquetada  $u_{k_s}$ . Luego, todos los modelos  $\mathcal{M}$  generaron clases predichas en subconjuntos no etiquetados, y para muestras con una confianza mayor que  $\tau$ , consideramos la predicción del modelo  $\widehat{y}_u^{(i)}$  como una nueva pseudo-etiqueta asociada con  $x_u^{(i)}$ . El proceso de decisión consistió en dos condiciones: La condición  $c_1$  verifica que la clase asignada por cada una de las arquitecturas  $m_{k_s}^\dagger$  CNN sea la misma. La condición  $c_2$  verifica que la probabilidad asociada con cada clase sea mayor que  $\tau$ . Si las condiciones  $c_1$  y  $c_2$  son ambas verdaderas, entonces la clase (pseudo-etiqueta) se asigna a la muestra  $u_{k_s}$ . Por lo tanto, la predicción se utiliza para complementar los datos de entrenamiento  $\mathcal{D}_\ell := \mathcal{D}_\ell \cup (x_u^{(i)}, \widehat{y}_u^{(i)})$ . Posteriormente, se utiliza el nuevo subconjunto  $\mathcal{D}_\ell$  para refinar el conjunto  $\mathcal{M}$  con imágenes de entrenamiento en toda la red. En cada iteración de entrenamiento, se consideraron estrategias como *dropout* y el aumento de datos para mejorar cada modelo y evitar el sobre-entrenamiento. A continuación, se expone una breve descripción de cada arquitectura:

**4.1.1 ResNet152:** se encuentra entre las arquitecturas de redes neuronales más utilizadas que abordan el problema del desvanecimiento del gradiente causado por la creciente profundidad de una CNN<sup>61</sup>. El modelo ResNet permite el paso de información relevante de una capa a la siguiente mediante el uso de conexiones de acceso directo (la transformación de una capa convolucional  $f(x)$ ), que omite una o más capas a la operación de combinación y formula la salida  $y$  como  $y = r(x + f(x))$ , donde  $r$  es una función de activación particular. Este bloque residual está construido para mejorar el proceso de aprendizaje al agregar la posibilidad de construir arquitecturas más profundas. Esta arquitectura se ha utilizado con éxito en muchas aplicaciones de detección remota de imágenes satelitales, como la

---

<sup>61</sup> HE, Kaiming, et al. Deep residual learning for image recognition. 2016. p. 770-778

segmentación semántica<sup>62</sup>, la clasificación de escenas<sup>63</sup> y la detección de objetos<sup>64</sup>.

**4.1.2 InceptionV3:**<sup>65</sup> es otra arquitectura popular que introduce el concepto de factorización en pequeñas convoluciones y usa ramas no solo en el módulo de *Inception* sino también en las propias ramas. Promueve representaciones de alta dimensionalidad con un proceso de entrenamiento más rápido en comparación con las arquitecturas clásicas. Tiene una gran aceptación en detección remota de imágenes satelitales debido al eficiente proceso de entrenamiento, incluida una representación más profunda, esencial para la clasificación.<sup>66,67,68</sup>.

---

<sup>62</sup> AUDEBERT,Nicolas; LE SAUX,Bertrand and LEFÈVRE,Sébastien. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 140, p. 20-32

<sup>63</sup> MAHDIANPARI,Masoud, et al. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 130, p. 13-31

<sup>64</sup> AZIMI,Seyed Majid, et al. Towards multi-class object detection in unconstrained remote sensing imagery. 2018. p. 150-165

<sup>65</sup> SZEGEDY,Christian, et al. Rethinking the inception architecture for computer vision. 2016. p. 2818-2826

<sup>66</sup> MAHDIANPARI,Masoud, et al. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. En: REMOTE SENSING. vol. 10, no. 7, p. 1119

<sup>67</sup> CAI,Z. L.; WENG,Qian and YE,S. Z. Research on Se-Inception in High-Resolution Remote Sensing Image Classification. En: THE INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. vol. 42, p. 539-545

<sup>68</sup> SHAFAYEY,Mayar A., et al. Comparison of CNNs for remote sensing scene classification. 2018. p. 27-32

**4.1.3 InceptionResNetV2:**<sup>69</sup> integra conceptos de vanguardia como ResNet, módulos de *Inception* y normalización por lotes para obtener con éxito una representación más profunda. Estos módulos se utilizan en detección remota de imágenes satelitales para lograr la fusión no-lineal de canales, con una evaluación eficaz en la clasificación.<sup>70,71</sup>.

## 4.2 ESQUEMA SEMI-SUPERVISADO

Los métodos SSL son útiles en dominios con datos etiquetados limitados. También han sido efectivos en problemas complejos con una sustancial similitud visual entre clases, incluso superando arquitecturas entrenadas bajo esquemas supervisados. Esto podría atribuirse a los enfoques SSL que, de forma natural, implementan el filtrado (selección) de muestras no-etiquetadas para una nueva representación del modelo.

En este trabajo, usamos una estrategia de etapas progresivas para la propagación de etiquetas. Cada etapa consistió en pasos de entrenamiento y etiquetado. En cada etapa  $S$ , se aplicó una estrategia de transferencia de aprendizaje para ajustar desde un dominio natural a un dominio de detección remota de imágenes satelitales utilizando un subconjunto  $D_{u_s}$  de los datos sin etiquetar para actualizar cada modelo  $m_k$  CNN de forma independiente (vea la Figura 7).

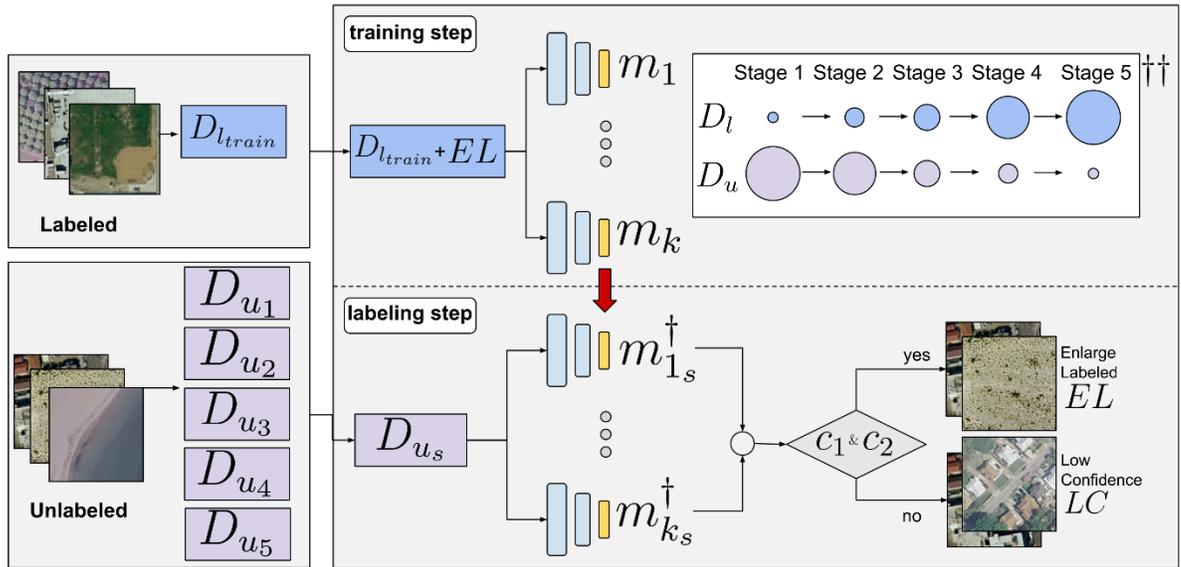
---

<sup>69</sup> SZEGEDY,Christian, et al. Rethinking the inception architecture for computer vision. 2016. p. 2818-2826

<sup>70</sup> MAHDIANPARI,Masoud, et al. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. En: REMOTE SENSING. vol. 10, no. 7, p. 1119

<sup>71</sup> HÄUFEL,Gisela, et al. Evaluation of CNNs for land cover classification in high-resolution airborne images. 2018. p. 1079003

Figura 7. Pipeline del enfoque propuesto



Fuente: elaboración propia.

El método SSL propuesto se resume en el Algoritmo 4 expuesto en la Figura 8. Las muestras etiquetadas  $D_l$  se dividieron en  $D_{l_{train}}$  y  $D_{l_{test}}$  (línea 1). Se inicializaron dos conjuntos vacíos: un conjunto de etiquetas ampliado ( $EL$ ) y muestras pseudo-etiquetadas de baja confianza (línea 2) ( $LC$ ). Las muestras no-etiquetadas  $D_u$  se dividieron en varios subconjuntos  $D_{u_s}$ , donde  $s$  es el número de etapas. Una etapa esta compuesta por un paso de entrenamiento y uno de etiquetado (línea 3).  $\mathcal{M}$  es el conjunto de  $k$  arquitecturas CNN inicializadas con los pesos de ImageNet (línea 4). Cada  $D_{u_s}$  subconjunto no-etiquetado  $D_u$  se usó en las etapas progresivas de la siguiente manera (líneas 5-24): Para cada modelo  $m_k$  que pertenece a  $\mathcal{M}$ , se realizó un paso de entrenamiento usando  $D_{l_{train}}$  junto con las muestras de  $EL$  aplicando una estrategia de transferencia de aprendizaje para adaptar desde el dominio natural a un dominio de detección remota de imágenes satelitales, lo que resultó en un modelo entrenado  $m_{k_s}^\dagger$ , que luego se guardó en  $\mathcal{M}_{k_s}$  (líneas 6-8). Por cada muestra no-etiquetada  $u$  de un subconjunto  $D_{u_s}$ , se realizó un paso de etiquetado utilizando cada modelo  $m_{k_s}^\dagger$  entrenado del conjunto

$M_{k_s}$  (líneas 9-15). De cada modelo  $m_{k_s}^\dagger$ , se obtuvo una pseudo-etiqueta  $u_{k_s}$ , y la probabilidad  $p_{k_s}$  para cada muestra no-etiquetada  $u$  usando una función de predicción (líneas 10-11). Si todas las pseudo-etiquetas  $u_{1_s}, \dots, u_{k_s}$  de la muestra son de la misma clase de detección remota de imágenes satelitales, y todas las probabilidades  $p_{1_s}, \dots, p_{k_s}$  son mayores que un umbral de alta confianza  $\tau$ , entonces las pseudo-etiquetas  $u_{k_s}$  se guardan en el conjunto  $EL$  (líneas 12-13). De lo contrario, las pseudo-etiquetas  $u_{k_s}$  se guardaron en el conjunto  $LC$  (líneas 14-15). Una vez completados los pasos de entrenamiento y etiquetado, el paso de evaluación final se llevó a cabo de la siguiente manera. Para cada  $l_{test}$  muestra del subconjunto  $D_{l_{test}}$ , se aplicó un paso de etiquetado utilizando cada  $m_{k_s}^\dagger$  modelo entrenado (líneas 16-24). Para cada modelo  $m_{k_s}^\dagger$ , se obtiene una pseudo-etiqueta  $l_{test_{k_s}}$  y su correspondiente probabilidad  $p_{test_{k_s}}$  (líneas 17-18). Si todos los  $l_{test_{1_s}}, \dots, l_{test_{k_s}}$  pertenecen a la misma clase entonces la muestra se almacena en  $l_{test_{pred}}$  (líneas 19-20). De lo contrario, la pseudo-etiqueta  $l_{test_{k_s}}$  con la máxima probabilidad o superior  $p_{test_{k_s}}$  se almacena en  $l_{test_{pred}}$  (líneas 21-22). Posteriormente, el  $l_{test_{pred}}$  se guardó en el conjunto  $D_{l_{test_{pred}}}$  (línea 23). Finalmente, se aplica una función de precisión usando los conjuntos  $D_{l_{test}}$  y  $D_{l_{test_{pred}}}$  para determinar el rendimiento del enfoque de conjunto (línea 24).

Figura 8. Algoritmo 4 Aprendizaje semi-supervisado de decisiones ajustado

---

Algoritmo 4 Aprendizaje Semi-supervisado de decisiones ajustado

---

```

1:  $D_l \rightarrow \{D_{l_{train}}, D_{l_{test}}\}$ 
2:  $EL, LC \leftarrow \{\}, \{\}$ 
3:  $D_u \rightarrow \{D_{u_1}, \dots, D_{u_s}\}$ 
4:  $\mathcal{M} \rightarrow \{m_1, \dots, m_k\}$ 
5: for  $D_{u_s} \in D_u$  do
6:   for  $m_k \in \mathcal{M}$  do
7:      $m_{k_s}^\dagger \leftarrow \text{train}(m_k, D_{l_{train}} + EL)$ 
8:      $\mathcal{M}_{k_s} \leftarrow m_{k_s}^\dagger$ 
9:   for  $u \in D_{u_s}$  do
10:    for  $m_{k_s}^\dagger \in \mathcal{M}_{k_s}$  do
11:       $u_{k_s}, p_{k_s} \leftarrow \text{label}(m_{k_s}^\dagger, u)$ 
12:      if  $(u_{1_s} == u_{k_s}) \& ((p_{1_s} \& \dots \& p_{k_s}) > \tau)$  then
13:         $EL \leftarrow u_{k_s}$ 
14:      else
15:         $LC \leftarrow u_{k_s}$ 
16:   for  $l_{test} \in D_{l_{test}}$  do
17:     for  $m_{k_s}^\dagger \in \mathcal{M}_{k_s}$  do
18:        $l_{test_{k_s}}, p_{test_{k_s}} \leftarrow \text{label}(m_{k_s}^\dagger, l_{test})$ 
19:       if  $(l_{test_{1_s}} == l_{test_{k_s}})$  then
20:          $l_{test_{pred}} \leftarrow l_{test_{k_s}}$ 
21:       else
22:          $l_{test_{pred}} \leftarrow \text{max}(l_{test_{k_s}} | p_{test_{k_s}})$ 
23:        $D_{l_{test_{pred}}} \leftarrow l_{test_{pred}}$ 
24:    $\text{accuracy}(D_{l_{test}}, D_{l_{test_{pred}}})$ 

```

---

Fuente: elaboración propia.

## 5. CONFIGURACIÓN EXPERIMENTAL

El método propuesto se validó en 4 conjuntos de datos de detección remota de imágenes satelitales de uso común (Ver muestras de algunas imágenes en la Figura 9), a continuación, una breve descripción de cada uno:

- El conjunto de datos **UC-Merced** se usa popularmente para la clasificación y recuperación de escenas de detección remota de imágenes satelitales<sup>72,73,74</sup>. Está compuesto por 2100 imágenes satelitales con 21 clases de escenas. Este conjunto de datos se extrajo de varias regiones del Mapa Nacional del Servicio Geológico de los Estados Unidos (USGS), que luego se recortaron en regiones de tamaño 256 x 256.
- El conjunto de datos **AID** se encuentra entre los más voluminosos utilizados para la clasificación de escenas aéreas. Comprende 30 clases y 10,000 imágenes aéreas de un tamaño de 600 x 600. Este conjunto de datos desequilibrado es un desafío debido a la alta variabilidad en las imágenes, ya que incluyen diferentes estaciones y son de diferentes países y regiones de todo el mundo.<sup>75</sup>
- El conjunto de datos **WHU-RS19** se extrajo de un conjunto de imágenes satelitales de Google Earth. El conjunto de datos comprende 1005 imágenes con un tamaño promedio de 600 x 600, distribuidas por igual en

---

<sup>72</sup> XIA,Gui-Song, et al. Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. En: REMOTE SENSING. vol. 7, no. 11, p. 15014-15045

<sup>73</sup> QI,Kunlun, et al. Land-use scene classification in high-resolution remote sensing images using improved correlatons. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 12, no. 12, p. 2403-2407

<sup>74</sup> MEKHALFI,Mohamed L., et al. Land-use classification with compressive sensing multifeature fusion. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 12, no. 10, p. 2155-2159

<sup>75</sup> XIA,Gui-Song, et al. AID: A benchmark data set for performance evaluation of aerial scene classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 55, no. 7, p. 3965-3981

19 clases. El principal desafío en este conjunto de datos es lidiar con las variaciones de la imagen relacionadas con la resolución, escala, orientación e iluminación.<sup>76,77,78</sup>

- El conjunto de datos **NWPU-RESISC45** comprende 31.500 imágenes satelitales divididas en 45 clases de escenas. El conjunto de datos (imágenes con un tamaño de 256 x 256) mantiene muchas variaciones en la traslación, la perspectiva, la postura de los objetos, la resolución espacial, la iluminación, el fondo y la oclusión. El objetivo principal de evaluar el enfoque propuesto con este conjunto de datos fue tratar la similitud visual en las muestras<sup>79</sup>.

---

<sup>76</sup> SHENG,Guofeng, et al. High-resolution satellite scene classification using a sparse coding based multiple feature combination. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 33, no. 8, p. 2395-2412

<sup>77</sup> MOUSTAKIDIS,Serafeim, et al. SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 50, no. 1, p. 149-169

<sup>78</sup> LONGBOTHAM,Nathan, et al. Very high resolution multiangle urban classification analysis. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 50, no. 4, p. 1155-1170

<sup>79</sup> CHENG,Gong; HAN,Junwei and LU,Xiaoqiang. Remote Sensing Image Scene Classification: Benchmark and State of the Art. En: PROCEEDINGS OF THE IEEE. vol. 105, no. 10, p. 1865-1883

Figura 9. Muestras de los conjuntos de datos



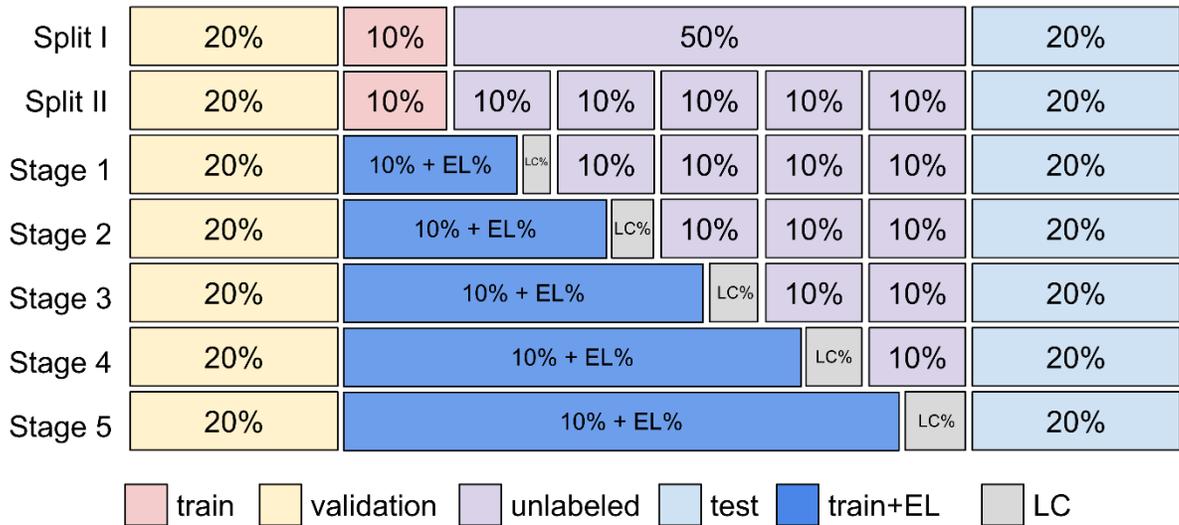
Fuente: elaboración propia

Se adoptó una estrategia de validación cruzada de *K-folds* ( $K=4$ ) para todos los conjuntos de datos con el objetivo de lograr el umbral de significación estadística en los resultados obtenidos. Por lo tanto, para emular el entorno semi-supervisado, cada *K-fold* se organizó varias veces en los siguientes subconjuntos: entrenamiento inicial, validación, no-etiquetado y evaluación. La primera división asignó el 10% para el entrenamiento inicial, el 20% para la validación, el 50% como muestras no-etiquetadas y el 20% restante para la evaluación. La segunda fue una división adoptada de Han et. al.<sup>80</sup> para comparar con el estado del arte

<sup>80</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

nuestra estrategia propuesta, lo cual requirió dividir en cinco conjuntos de lotes tomados del subconjunto de muestras no-etiquetadas. Cada lote no-etiquetado contenía un 10% de muestras no-etiquetadas. Después del paso de entrenamiento, en el paso de etiquetado, un subconjunto no-etiquetado se sometió a un proceso de etiquetado (explicado en el Algoritmo 4 expuesto en la Figura 8), y se utilizaron las muestras etiquetadas ampliadas (*EL*) junto con las muestras de entrenamiento iniciales originales para refinar la estrategia de ensemble en el enfoque propuesto. En particular, el número de muestras *EL* aumentó después de cada etapa. La Figura 10 ilustra el esquema de distribución de datos utilizado para validar el enfoque propuesto.

Figura 10. División de los datos



Fuente: elaboración propia

Aunque este enfoque puede adaptarse a cualquier cantidad de arquitecturas de CNN, debido a las limitaciones de memoria de la GPU en este trabajo, implementamos un experimento utilizando tres arquitecturas. Aquí, cada red se adaptó al dominio de la detección remota de imágenes satelitales, eliminando la última capa de softmax y colocando un vector embebido transicional (con 512 unidades y la capa utiliza una convolución de *Average Pooling global 2D*).

Además, se aplicó un *dropout* del 50% en esta capa de transición. Se utilizó un total de 30 épocas para cada red, con una tasa de aprendizaje de  $1e^{-4}$  y un tamaño de lote de 16.

El hiper-parámetro  $\tau$  de nuestra solución propuesta está relacionado con el filtrado de muestras de pseudo-etiquetas de alta confianza. Así mismo, al estar asociado con las probabilidades predichas por el modelo de ensemble, el rango de valores de  $\tau$  consiste entre 0 a 1. Para lo cual debido a exploración previa se observó que al seleccionar un umbral de alta confianza  $\tau$  cercano a 1 se obtiene mayor precisión en la asignación de pseudo-etiquetas, brindando consecuentemente un esquema semi-supervisado más robusto con un alto grado de control en la propagación del error. Experimentalmente, se fijó el valor de  $\tau = 0.9$ . Durante el entrenamiento, se implementó el aumento de datos para enriquecer la configuración del entrenamiento y evitar sesgos. Aumentamos las muestras en un factor de 2x, seguido por algunas transformaciones de la imagen, incluidos giros horizontales y verticales, y rotaciones de hasta  $90^\circ$ . Además, se implementó una estrategia de detención temprana para seguir automáticamente la pérdida de validación y detener si, después de cinco épocas, el cambio en esta métrica no se observaba una pérdida de  $1e^{-3}$ .

## 6. EVALUACIÓN Y RESULTADOS

Nuestro enfoque se validó inicialmente en el conjunto de datos UCMerced. Los Cuadros 1, 2 y 3 se resumen los resultados obtenidos en cada etapa del SSL. La evaluación incluyó arquitecturas independientes y el enfoque de ensemble, trabajando bajo el mismo principio SSL. En general, todas las arquitecturas reportan una mejora en su desempeño de este conjunto de datos, destacando el modelo InceptionV3 con una ganancia de  $\approx 3\%$  en precisión. El mejor rendimiento se logró con el enfoque de conjunto después de completar cinco etapas de entrenamiento, logrando una precisión promedio de  $\approx 96.15\%$ . A partir de este resultado, se puede observar que el enfoque de ensemble proporciona estabilidad en la representación, promovida por la regla de confianza. Esta arquitectura de ensemble logró una ganancia de  $\approx 2\%$  ( $> 42$  predicciones correctas adicionales), lo que resultó en una mejora significativa en el desempeño utilizando el conjunto de datos evaluado.

Cuadro 1. Precisión – UCMerced

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	92.25 $\pm$ 0,95	93.63 $\pm$ 1,55	93.81 $\pm$ 0,91	95.07 $\pm$ 1,08	95.37 $\pm$ 1,06
InceptionResNetV2	92.01 $\pm$ 0,95	92.13 $\pm$ 1,17	92.67 $\pm$ 1,31	92.91 $\pm$ 0,50	94.59 $\pm$ 0,93
ResNet152	91.95 $\pm$ 1,53	93.15 $\pm$ 0,75	92.55 $\pm$ 1,72	93.99 $\pm$ 0,83	94.29 $\pm$ 0,89
<b>Ensemble</b>	<b>94.53 <math>\pm</math> 0,36</b>	<b>95.19 <math>\pm</math> 0,20</b>	<b>95.25 <math>\pm</math> 1,03</b>	<b>95.61 <math>\pm</math> 0,36</b>	<b>96.15 <math>\pm</math> 0,39</b>

Fuente: elaboración propia.

Cuadro 2. F1- Score – UCMerced

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	92.15 $\pm$ 0,94	93.55 $\pm$ 1,59	93.71 $\pm$ 0,93	95.04 $\pm$ 1,10	95.29 $\pm$ 1,07
InceptionResNetV2	91.92 $\pm$ 0,99	92.07 $\pm$ 1,13	92.58 $\pm$ 1,37	92.85 $\pm$ 0,48	94.50 $\pm$ 0,98
ResNet152	91.88 $\pm$ 1,58	93.13 $\pm$ 0,73	92.46 $\pm$ 1,72	93.96 $\pm$ 0,88	94.21 $\pm$ 0,83
<b>Ensemble</b>	<b>94.50 <math>\pm</math> 0,37</b>	<b>95.14 <math>\pm</math> 0,24</b>	<b>95.16 <math>\pm</math> 1,05</b>	<b>95.58 <math>\pm</math> 0,38</b>	<b>96.10 <math>\pm</math> 0,39</b>

Fuente: elaboración propia.

Cuadro 3. Recall – UCMerced

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	92.25 ± 0,95	93.63 ± 1,55	93.81 ± 0,91	95.07 ± 1,08	95.37 ± 1,06
InceptionResNetV2	92.01 ± 0,95	92.13 ± 1,17	92.67 ± 1,31	92.91 ± 0,50	94.59 ± 0,93
ResNet152	91.95 ± 1,53	93.15 ± 0,75	92.55 ± 1,72	93.99 ± 0,83	94.29 ± 0,89
Ensemble	<b>94.53 ± 0,36</b>	<b>95.19 ± 0,20</b>	<b>95.25 ± 1,03</b>	<b>95.61 ± 0,36</b>	<b>96.15 ± 0,39</b>

Fuente: elaboración propia.

El enfoque propuesto se validó por segunda vez utilizando el conjunto de datos AID. Esta evaluación (reportada en el recuadro 4, 5 y 6) mostró una ganancia general de  $\approx 1\%$  ( $> 100$  predicciones correctas adicionales) después de ejecutar todos los esquemas SSL. Como en la evaluación anterior, ResNet152 (Etapa 5) e Inception V3 (Etapa 5) reportaron un desempeño comparable al enfoque de conjunto. Además, estas arquitecturas proporcionaron un rendimiento estable debido a una baja desviación estándar.

Cuadro 4. Precisión – AID

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	88.99 ± 0,42	89.84 ± 0,94	90.56 ± 0,08	90.97 ± 0,44	90.99 ± 0,26
InceptionResNetV2	87.26 ± 0,88	87.93 ± 0,42	88.84 ± 0,96	89.78 ± 0,62	89.63 ± 0,50
ResNet152	89.10 ± 1,04	90.10 ± 1,38	90.93 ± 0,75	91.66 ± 1,17	91.51 ± 0,15
Ensemble	<b>91.26 ± 0,83</b>	<b>92.02 ± 0,58</b>	<b>92.26 ± 0,23</b>	<b>92.78 ± 0,67</b>	<b>92.45 ± 0,17</b>

Fuente: elaboración propia.

Cuadro 5. F1-Score - AID

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	88.84 ± 0,38	89.71 ± 0,94	90.46 ± 0,05	90.86 ± 0,41	90.84 ± 0,21
InceptionResNetV2	87.18 ± 0,86	87.81 ± 0,40	88.74 ± 0,94	89.64 ± 0,65	89.50 ± 0,57
ResNet152	88.97 ± 1,05	89.99 ± 1,37	90.78 ± 0,77	91.50 ± 1,18	91.39 ± 0,20
Ensemble	<b>91.12 ± 0,85</b>	<b>91.89 ± 0,57</b>	<b>92.15 ± 0,23</b>	<b>92,65 ± 0,68</b>	<b>92.31 ± 0,19</b>

Fuente: elaboración propia.

Cuadro 6. Recall - AID

Arquitectura	Etapas 1	Etapas 2	Etapas 3	Etapas 4	Etapas 5
<b>InceptionV3</b>	88.99 ± 0,42	89.84 ± 0,94	90.56 ± 0,08	90.97 ± 0,44	90.99 ± 0,26
<b>InceptionResNetV2</b>	87.26 ± 0,88	87.93 ± 0,42	88.84 ± 0,96	89.78 ± 0,62	89.63 ± 0,50
<b>ResNet152</b>	89.10 ± 1,04	90.10 ± 1,38	90.93 ± 0,75	91.66 ± 1,17	91.51 ± 0,15
<b>Ensemble</b>	<b>91.26 ± 0,83</b>	<b>92.02 ± 0,58</b>	<b>92.26 ± 0,23</b>	<b>92.78 ± 0,67</b>	<b>92.45 ± 0,17</b>

Fuente: elaboración propia.

La tercera validación del enfoque propuesto utilizó el conjunto de datos WHU-RS19. Esta evaluación (reportada en los Cuadros 7, 8 y 9) mostró una ganancia general de  $\approx 3\%$  ( $> 30$  predicciones adicionales corregidas) después de ejecutar el esquema SSL. Como en la evaluación anterior, ResNet152 (Etapas 5) también reportó un desempeño similar al enfoque de ensemble. Estas arquitecturas también proporcionaron un rendimiento estable debido a una baja desviación estándar.

Cuadro 7. Precisión – WHU-RS19

Arquitectura	Etapas 1	Etapas 2	Etapas 3	Etapas 4	Etapas 5
<b>InceptionV3</b>	90.36 ± 1,31	92.58 ± 0,89	92.32 ± 0,66	93.75 ± 1,53	93.36 ± 2,10
<b>InceptionResNetV2</b>	87.11 ± 2,10	90.62 ± 1,13	90.62 ± 0,95	91.80 ± 2,14	92.06 ± 0,78
<b>ResNet152</b>	88.80 ± 3,36	92.84 ± 2,80	91.54 ± 0,66	94.01 ± 1,93	94.40 ± 2,22
<b>Ensemble</b>	<b>92.97 ± 2,42</b>	<b>94.53 ± 1,56</b>	<b>93.62 ± 0,50</b>	<b>94.27 ± 2,13</b>	<b>95.18 ± 1,72</b>

Fuente: elaboración propia.

Cuadro 8. F1-Score – WHU-RS19

Arquitectura	Etapas 1	Etapas 2	Etapas 3	Etapas 4	Etapas 5
<b>InceptionV3</b>	90.26 ± 1,26	92.53 ± 0,79	92.22 ± 0,68	93.68 ± 1,60	93.26 ± 2,15
<b>InceptionResNetV2</b>	87.16 ± 1,99	90.55 ± 1,10	90.55 ± 1,10	91.73 ± 2,00	92.13 ± 0,85
<b>ResNet152</b>	88.72 ± 3,27	92.81 ± 2,87	91.32 ± 0,77	93.99 ± 1,97	94.40 ± 2,23
<b>Ensemble</b>	<b>92.89 ± 2,31</b>	<b>94.46 ± 1,59</b>	<b>93.51 ± 0,47</b>	<b>94.19 ± 2,18</b>	<b>95.19 ± 1,66</b>

Fuente: elaboración propia.

Cuadro 9. Recall – WHU-RS19

Arquitectura	Etapas 1	Etapas 2	Etapas 3	Etapas 4	Etapas 5
InceptionV3	90.36 ± 1,31	92.58 ± 0,89	92.32 ± 0,66	93.75 ± 1,53	93.36 ± 2,10
InceptionResNetV2	87.11 ± 2,10	90.62 ± 1,13	90.62 ± 0,95	91.80 ± 2,14	92.06 ± 0,78
ResNet152	88.80 ± 3,36	92.84 ± 2,80	91.54 ± 0,66	94.01 ± 1,93	94.40 ± 2,22
<b>Ensemble</b>	<b>92.97 ± 2,42</b>	<b>94.53 ± 1,56</b>	<b>93.62 ± 0,50</b>	<b>94.27 ± 2,13</b>	<b>95.18 ± 1,72</b>

Fuente: elaboración propia.

En los Cuadros 10, 11 y 12 se reportan los resultados obtenidos utilizando el enfoque propuesto sobre el conjunto de datos NWPU-RESISC45. Se logró una ganancia general de  $\approx 1\%$  después de todas las etapas utilizando el enfoque propuesto. InceptionV3 y ResNet152 mostraron el rendimiento más estable en todas las etapas. Teniendo en cuenta que en este caso el 1% de ganancia significó 350 muestras adicionales, este fue el conjunto de datos más desafiante en nuestro esquema de validación propuesto. Otra observación relevante fue que, para los enfoques individuales, InceptionV3 y ResNet152 mostraron una mejora de  $\approx 1\%$ , destacando la complejidad en la representación de este conjunto de datos. Sin embargo, es relevante mencionar que, en cada etapa, el enfoque propuesto mostró el mejor desempeño para cada arquitectura.

Cuadro 10. Precisión – NWPU-RESISC45

Arquitectura	Etapas 1	Etapas 2	Etapas 3	Etapas 4	Etapas 5
InceptionV3	90.33 ± 0,15	90.68 ± 0,15	90.07 ± 0,21	90.92 ± 0,20	91.12 ± 0,29
InceptionResNetV2	87.00 ± 0,97	86.15 ± 1,99	87.12 ± 1,18	88.03 ± 0,83	86.87 ± 3,00
ResNet152	90.19 ± 0,20	90.12 ± 1,07	89.64 ± 1,54	90.57 ± 0,67	90.97 ± 0,29
<b>Ensemble</b>	<b>91.42 ± 0,44</b>	<b>91.67 ± 0,44</b>	<b>91.26 ± 0,56</b>	<b>91.95 ± 0,64</b>	<b>91.98 ± 0,55</b>

Fuente: elaboración propia.

Cuadro 11. F1-Score – NWPU-RESISC45

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	90.33 ± 0,16	90.65 ± 0,16	90.02 ± 0,20	90.87 ± 0,22	91.05 ± 0,30
InceptionResNetV2	87.01 ± 0,96	86.16 ± 2,02	87.07 ± 1,26	88.07 ± 0,74	86.68 ± 3,31
ResNet152	90.17 ± 0,21	90.09 ± 1,08	89.60 ± 1,56	90.54 ± 0,67	90.93 ± 0,29
Ensemble	<b>91.38 ± 0,43</b>	<b>91.64 ± 0,43</b>	<b>91.21 ± 0,60</b>	<b>91.92 ± 0,63</b>	<b>91.92 ± 0,55</b>

Fuente: elaboración propia.

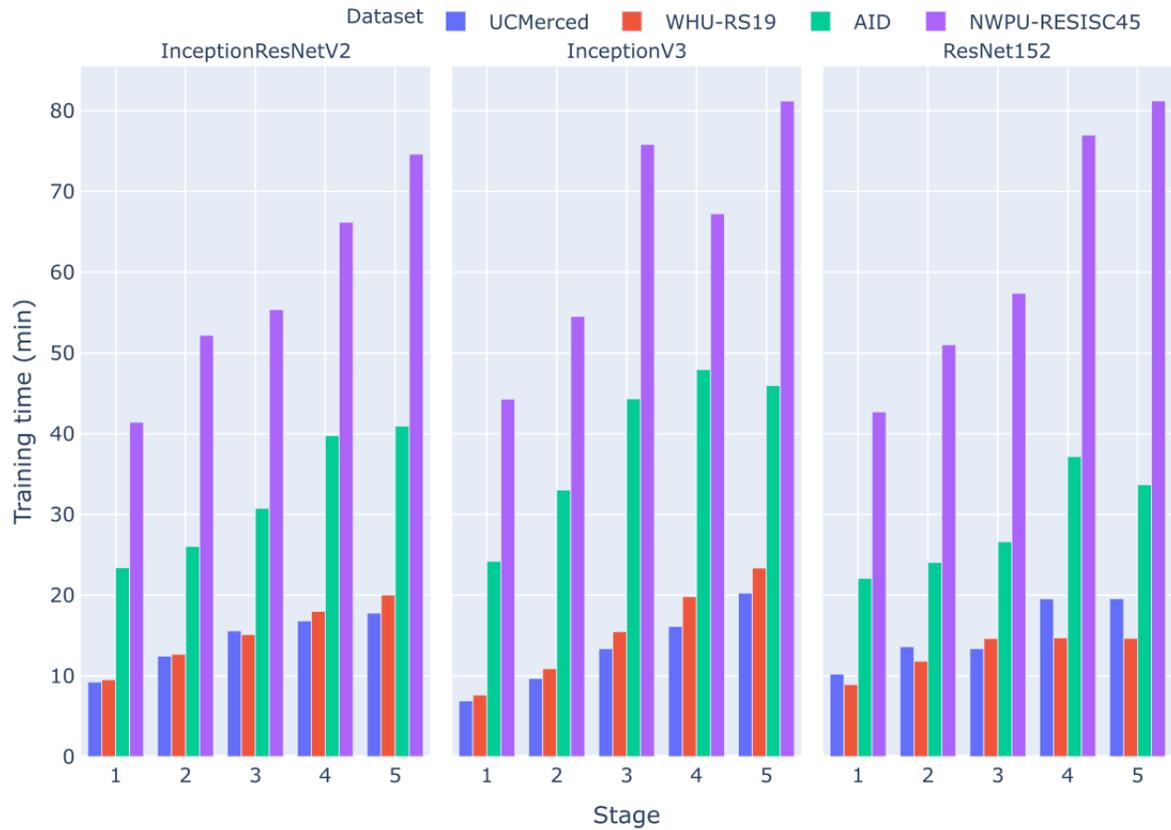
Cuadro 12. Recall – NWPU-RESISC45

Arquitectura	Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
InceptionV3	90.33 ± 0,15	90.68 ± 0,15	90.07 ± 0,21	90.92 ± 0,20	91.12 ± 0,29
InceptionResNetV2	87.00 ± 0,97	86.15 ± 1,99	87.12 ± 1,18	88.03 ± 0,83	86.87 ± 3,00
ResNet152	90.19 ± 0,20	90.12 ± 1,07	89.64 ± 1,54	90.57 ± 0,67	90.97 ± 0,29
Ensemble	<b>91.42 ± 0,44</b>	<b>91.67 ± 0,44</b>	<b>91.26 ± 0,56</b>	<b>91.95 ± 0,64</b>	<b>91.98 ± 0,55</b>

Fuente: elaboración propia.

El tiempo medio de entrenamiento (en minutos) requerido por cada arquitectura involucrada en el enfoque propuesto validado en los conjuntos de datos explorados se describe en la Figura 11. Los resultados se calcularon utilizando un clúster que comprende: 2x Intel Xeon Gold 5118 de 16 núcleos, 512 GB de RAM y una RTX TITAN con 24 GB de VRAM. Cabe señalar que, en cada etapa, el tiempo requerido para entrenar cada modelo aumentó hasta casi duplicar el tiempo de entrenamiento de la línea de base. Este comportamiento podría explicarse por el factor de aumento, una estrategia ampliamente utilizada para manejar el problema del sobre-entrenamiento de los modelos. Además, el tiempo de entrenamiento requerido por el enfoque propuesto sobre cada conjunto de datos difiere según el tamaño de cada conjunto de datos. El tiempo medio de inferencia que requirió nuestro enfoque propuesto para etiquetar una muestra no-etiquetada fue de  $\approx 15$  [ms].

Figura 11. Tiempo de entrenamiento del método propuesto

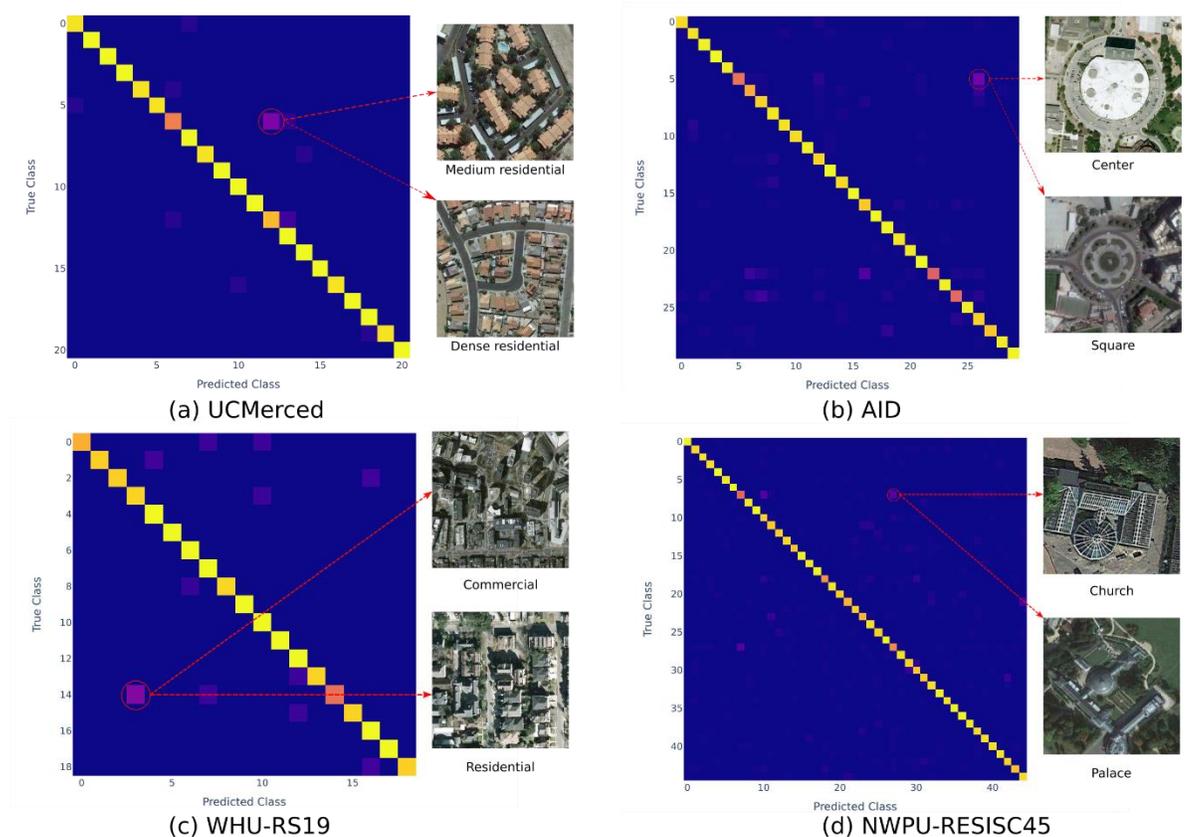


Fuente: elaboración propia.

La Figura 12 presenta un análisis detallado por clase de los resultados obtenidos en todos los conjuntos de datos considerados. En general, el enfoque propuesto evidencia una tendencia en los resultados por clase, con varianza baja. Por ejemplo, para UCMerced, el escenario de clasificación más desafiante se encuentra entre Residencial medio y denso, con un error de  $\approx 25\%$ . Esto podría explicarse porque los objetos de cada clase son similares en apariencia visual. Para el conjunto de datos WHU-RS19, el modelo de ensemble propuesto confundió entre clases comerciales y residenciales con un error de  $\approx 27\%$ , que podría atribuirse a las estrechas similitudes visuales estructurales en estas clases. Además, para el conjunto de datos AID, el modelo confundió las clases Centro y Cuadra con un error de  $\approx 23\%$ . Esto puede haber ocurrido porque la clase Centro

poseía una vía circular en el centro que era difícil de diferenciar en la mayoría de los casos, en contraste con la clase Cuadrada, que incluía vías cuadradas y circulares. Para el conjunto de datos NWPU-RESISC45, el modelo de ensemble confundió las clases Palacio e Iglesia con un error de  $\approx 14\%$  que es aceptable debido a las estrechas similitudes estructurales entre estas clases en particular.

Figura 12. Muestras de matrices de confusión



Fuente: elaboración propia.

Finalmente, se llevó a cabo una comparación de rendimiento con los métodos más avanzados para todos los conjuntos de datos considerados y considerando estrategias cercanas. El Cuadro 13 resume los resultados obtenidos para todos

los conjuntos de datos con respecto a las estrategias propuestas por Dai *et. al*<sup>81</sup> y Han *et. al.*<sup>82</sup>. En casi todos los casos, nuestro enfoque propuesto superó las estrategias de referencia, mostrando flexibilidad para adaptarse a diferentes conjuntos de datos y solidez en la representación de los datos. En particular, en los conjuntos de datos AID y UCMerced, el enfoque propuesto mejoró los resultados al usar lotes (5%) y (10%) para comenzar a entrenar. Para UCMerced, con solo el 10% de los datos, el enfoque de ensemble también superó de manera similar a los enfoques de última generación.

En contraste con las mismas condiciones iniciales de sólo el 5% de las muestras etiquetadas contra Dai *et. al*<sup>83</sup> (5%), el enfoque propuesto mejoró con una ganancia del 6%. De forma similar, el enfoque propuesto superó los resultados de referencia informados para todas las etapas de entrenamiento para el conjunto de datos de AID. Para el WHU-RS19 en la configuración del 5% de los datos etiquetados el enfoque propuesto muestra limitaciones al compararse con el trabajo de Dai *et. al*<sup>84</sup>, sin embargo, en la configuración del 10% nuestro método mostró un rendimiento comparable con el trabajo de Han *et. al.*<sup>85</sup> Para el conjunto

---

<sup>81</sup> DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

<sup>82</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

<sup>83</sup> DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

<sup>84</sup> DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

<sup>85</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

de datos NWPU-RESISC45, nuestro enfoque propuesto superó el enfoque de vanguardia con una mejora de  $\approx 4\%$  con una desviación estándar baja.

En comparación con otros enfoques semi-supervisados en la detección remota de imágenes satelitales, encontramos que nuestro enfoque propuesto tiene diferencias significativas con los métodos de última generación. Zhou *et al.*<sup>86</sup>, Yan *et al.*<sup>87</sup>, y Wang *et al.*<sup>88</sup> en el conjunto de datos NWPU-RESISC45 usando el 20% del conjunto de entrenamiento inicial reportó precisiones de  $93.47 \pm 0.60$ ,  $83.12 \pm 0.26$  y  $88.413$ , respectivamente. En contraste, nuestro enfoque propuesto reportó  $94.41 \pm 0.77$  en la puntuación de precisión, que excede levemente la puntuación del método de última generación. Además, Wang *et al.*<sup>89</sup> utilizando el 10% del conjunto de entrenamiento inicial reportó una precisión de  $83.127$ , mientras que nuestro método indicó  $91.98 \pm 0.55$  para uno de los conjuntos de datos más complejos, mostrando una diferencia de aproximadamente 9% (3150 predicciones correctas adicionales). Para darle mayor solidez a la evaluación del método propuesto, analizamos la diferencia estadística del método propuesto en contraste con el estado del arte aplicando un análisis de varianza clásico denominado como ANOVA. Comprobamos que había diferencias estadísticas significativas ( $p < 0.05$ ) entre Dai *et al.*<sup>90</sup> frente a nuestro enfoque propuesto para los conjuntos de

---

<sup>86</sup> ZHOU,Zhao, et al. Satellite image scene classification via ConvNet with context aggregation. 2018. p. 329-339

<sup>87</sup> YAN,Peiyao, et al. Semi-supervised representation learning for remote sensing image classification based on generative adversarial networks. En: IEEE ACCESS. vol. 8, p. 54135-54144

<sup>88</sup> WANG,Chen, et al. Semi-Supervised Learning-Based Remote Sensing Image Scene Classification Via Adaptive Perturbation Training. p. 541-544

<sup>89</sup> WANG,Chen, et al. Semi-Supervised Learning-Based Remote Sensing Image Scene Classification Via Adaptive Perturbation Training. p. 541-544

<sup>90</sup> DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

datos AID y UCMerced. Además, Han *et al.*<sup>91</sup> frente a nuestro enfoque propuesto en los conjuntos de datos AID, UCMerced y NWPU-RESISC45 también mostró diferencias estadísticamente significativas ( $p < 0.05$ ). Estos resultados sugieren un desempeño satisfactorio del método de ensemble sobre los conjuntos de datos validados con la capacidad de adaptar representaciones en diferentes dominios.

Cuadro 13. Rendimiento del enfoque propuesto vs Estado del Arte

Conjunto de datos	UCMerced		NWPU-RESISC45		AID		WHU-RS19	
	5%	10%	10%	20%	5%	10%	5%	10%
Zhou et. al	-	-	-	93.47 ± 0,60	-	-	-	-
Yan et. al	-	-	-	83.12 ± 0,26	-	-	-	-
Wang et. al	-	-	83.12	88.41	-	-	-	-
Dai et. al	85.14	-	-	-	83.72 ± 0,22	-	94.73	-
Han et. al	-	94.52 ± 1,38	88.60 ± 0,31	-	-	91.35 ± 0,83	-	96.84 ± 1,33
<b>Nuestro</b>	<b>91.42 ± 0,28</b>	<b>96.15 ± 0,39</b>	<b>91.98 ± 0,55</b>	<b>94.41 ± 0,77</b>	<b>89.61 ± 0,62</b>	<b>92.45 ± 0,17</b>	<b>86.22 ± 3,22</b>	<b>95.18 ± 1,72</b>

Fuente: elaboración propia.

<sup>91</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

## 7. CONCLUSIONES

El enfoque de ensemble semi-supervisado propuesto aprovecha las representaciones profundas convolucionales complementarias y permite el etiquetado de nuevas muestras con alta confianza de acuerdo con el consenso entre varias redes. Además, cada nuevo lote etiquetado se utiliza para fortalecer la representación visual a través de una estrategia de transferencia de aprendizaje que da como resultado clasificaciones más precisas. Esta estrategia se evaluó exhaustivamente en cuatro conjuntos de datos de referencia. El enfoque propuesto logró resultados apreciables con respecto a la comparación de referencia, utilizando solo 10% de datos en la primera iteración y, posteriormente, aumentando el conocimiento de transferencia a través de lotes del 10% de datos no-etiquetados. En general, el enfoque propuesto logró una precisión promedio de  $> 90\%$  después de la Etapa Cinco, mostrando mejoras significativas entre etapas en algunos conjuntos de datos. Cabe destacar que para conjuntos de datos más grandes (con  $> 31,500$  muestras), el método logró una ganancia de  $\approx 4\%$  sobre algunos enfoques propuestos en la literatura.<sup>92,93</sup>

Por ejemplo, Han et. al.<sup>94</sup> usó vectores embebidos extraídos de capas convolucionales profundas para entrenar una máquina de soporte vectorial (SVM) en una estrategia semi-supervisada de co-entrenamiento. Estos vectores fueron

---

<sup>92</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

<sup>93</sup> DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

<sup>94</sup> HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

capturados de un dominio general, lo cual limita la clasificación correcta entre clases con una similitud visual cercana. Por el contrario, el enfoque propuesto logró una mejor descripción visual debido a la estrategia de transferencia de aprendizaje implementada en cada etapa del entrenamiento semi-supervisado, que abordó nuevas características visuales para ajustar y enriquecer la representación. Esta ventaja fue evidente usando la misma configuración experimental, logrando un mejor rendimiento sobre el conjunto de datos NWPU-RESISC45 con una precisión incremental y una desviación estándar más baja de  $91.42 \pm 0.44$  (primera etapa) a  $91.98 \pm 0.55$  (quinta etapa). Además, Dai *et. al* propuso el uso de vectores de incrustación para entrenar métodos clásicos de aprendizaje automático y una estrategia de propagación que maximizara la distancia mutua de las muestras no-etiquetadas. En este enfoque, se formuló una representación intermedia para representar clases desde vectores embebidos. Esta estrategia admitió nuevas observaciones no-etiquetadas de representaciones intermedias, pero se restringió a un dominio específico dadas las muestras de entrenamiento. Nuestro enfoque propuesto aplicó la transferencia de aprendizaje, la cual ajustó progresivamente la representación para manejar estas limitaciones, pero permaneció abierto a la variabilidad de un dominio natural inicial. Con respecto a este trabajo en particular, el enfoque propuesto logró una mejora de 6% en precisión sobre el conjunto de datos de UCMerced.

El enfoque propuesto logró resultados significativos en cuatro conjuntos de datos utilizando tres redes neuronales convolucionales y reglas de decisión. En general, este enfoque comenzó a procesarse a partir de un conjunto de datos de entrenamiento débil (10% del total del conjunto de datos) con resultados estables, a pesar de las limitadas muestras etiquetadas. Atribuimos esto a la sólida representación obtenida de ImageNet. Posteriormente, en cada etapa, el enfoque procedía a etiquetar nuevas muestras con mayor precisión y reutilizarlas para volver a entrenar los modelos utilizando una estrategia de transferencia de aprendizaje. Una limitación principal del enfoque propuesto fue el tiempo

necesario para ajustar las representaciones. Experimentalmente, encontramos que, en promedio, el enfoque propuesto tomó 40 minutos en la primera etapa de entrenamiento y 80 minutos en la última etapa. También es digno de mención que desde la Etapa uno a la Etapa cinco, la precisión no necesariamente mejoró de forma monótona (ver los Cuadros 1, 4, 7 y 10), lo que podría ser asociado con la propagación del error en el paso de etiquetado.

Como trabajo futuro proponemos investigar estrategias para reducir dinámicamente dicha propagación de errores, por ejemplo, incluyendo más arquitecturas o aprendiendo reglas más sofisticadas para filtrar las muestras no-supervisadas en la etapa de entrenamiento. Planteamos explorar la calidad de clasificación de los parches visualizando de forma global la ubicación de estos y comparar con el estado del arte. Además, la exploración de estrategias de extremo a extremo permitiría compartir redes en arquitecturas de ensemble, lo que podría generar un mejor rendimiento.

## BIBLIOGRAFÍA

ALHICHRI,Haikel, et al. Tile-based semisupervised classification of large-scale VHR remote sensing images. En: JOURNAL OF SENSORS. vol. 2018,

ANUSE,Alwin and VYAS,Vibha. A novel training algorithm for convolutional neural network. En: COMPLEX & INTELLIGENT SYSTEMS. vol. 2, no. 3, p. 221-234

ARVANITI,Eirini, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. En: SCIENTIFIC REPORTS. vol. 8, no. 1, p. 1-11

AUDEBERT,Nicolas; LE SAUX,Bertrand and LEFÈVRE,Sébastien. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 140, p. 20-32

AZIMI,Seyed Majid, et al. Towards multi-class object detection in unconstrained remote sensing imagery. 2018. p. 150-165

BAI,Wenjia, et al. Semi-supervised learning for network-based cardiac MR image segmentation. 2017. p. 253-260

BLUM,Avrim and MITCHELL,Tom. Combining labeled and unlabeled data with co-training. 1998. p. 92-100

BROSCH,Tom and TAM,Roger. Efficient training of convolutional deep belief networks in the frequency domain for application to high-resolution 2D and 3D images. En: NEURAL COMPUTATION. vol. 27, no. 1, p. 211-227

CAI,Z. L.; WENG,Qian and YE,S. Z. Research on Se-Inception in High-Resolution Remote Sensing Image Classification. En: THE INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES. vol. 42, p. 539-545

CASTELLUCCIO,Marco, et al. Land use classification in remote sensing images by convolutional neural networks. En: arXiv PREPRINT arXiv:1508.00092.

CHAPELLE,Olivier; SCHOLKOPF,Bernhard and ZIEN,Alexander. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. En: IEEE TRANSACTIONS ON NEURAL NETWORKS. vol. 20, no. 3, p. 542

CHENG,Gong, et al. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 34, no. 1, p. 45-59

CHENG,Gong; HAN,Junwei and LU,Xiaoqiang. Remote sensing image scene classification: Benchmark and state of the art. En: PROCEEDINGS OF THE IEEE. vol. 105, no. 10, p. 1865-1883

CHENG,Gong; HAN,Junwei and LU,Xiaoqiang. Remote Sensing Image Scene Classification: Benchmark and State of the Art. En: PROCEEDINGS OF THE IEEE. vol. 105, no. 10, p. 1865-1883

DAI,Xueyuan, et al. Semisupervised Scene Classification for Remote Sensing Images: A Method Based on Convolutional Neural Networks and Ensemble Learning. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. June.vol. 16, no. 6, p. 869-873

DALAL,Navneet and TRIGGS,Bill. Histograms of oriented gradients for human detection. 2005. p. 886-893

FENG,Ziyi; HUANG,Guanhua and CHI,Daocai. Classification of the Complex Agricultural Planting Structure with a Semi-Supervised Extreme Learning Machine Framework. En: REMOTE SENSING. vol. 12, no. 22, p. 3708

GÓMEZ-CHOVA,Luis, et al. Multimodal classification of remote sensing images: A review and future directions. En: PROCEEDINGS OF THE IEEE. vol. 103, no. 9, p. 1560-1584

GU,Xiaowei and ANGELOV,Plamen P. A semi-supervised deep rule-based approach for remote sensing scene classification. 2019. p. 257-266

HAN,Wei, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 145, p. 23-43

HARALICK,Robert M.; SHANMUGAM,Karthikeyan and DINSTEN,Its' Hak. Textural features for image classification. En: IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. no. 6, p. 610-621

HÄUFEL,Gisela, et al. Evaluation of CNNs for land cover classification in high-resolution airborne images. 2018. p. 1079003

HE,Kaiming, et al. Deep residual learning for image recognition. 2016. p. 770-778

HUANG,Yanbo, et al. Agricultural remote sensing big data: Management and applications. En: JOURNAL OF INTEGRATIVE AGRICULTURE. vol. 17, no. 9, p. 1915-1931

JAIN,Anil K.; RATHA,Nalini K. and LAKSHMANAN,Sridhar. Object detection using Gabor filters. En: PATTERN RECOGNITION. vol. 30, no. 2, p. 295-309

KALKOWSKI,Sebastian, et al. Real-time analysis and visualization of the YFCC100M dataset. 2015. p. 25-30

KIM,Minho; MADDEN,Marguerite and WARNER,Timothy A. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery. En: PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING. vol. 75, no. 7, p. 819-829

LEE,Dong-Hyun. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. 2013. p. 2

LI,Xiaoxiao and SHAO,Guofan. Object-based urban vegetation mapping with high-resolution aerial photography as a single data source. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 34, no. 3, p. 771-789

LI,Youyou, et al. Semi-Supervised Remote Sensing Classification Via Associative Transfer. 2018. p. 2475-2478

LIU,Jiantao, et al. Urban green plastic cover mapping based on VHR remote sensing images and a deep semi-supervised learning framework. En: ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION. vol. 9, no. 9, p. 527

LIU,Yanfei; ZHONG,Yanfei and QIN,Qianqing. Scene classification based on multiscale convolutional neural network. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 56, no. 12, p. 7109-7121

LONGBOTHAM,Nathan, et al. Very high resolution multiangle urban classification analysis. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 50, no. 4, p. 1155-1170

LOWE,David G. Distinctive image features from scale-invariant keypoints. En: INTERNATIONAL JOURNAL OF COMPUTER VISION. vol. 60, no. 2, p. 91-110

MA,Lei, et al. Deep learning in remote sensing applications: A meta-analysis and review. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 152, p. 166-177

MA,Xinhua. Innovation-driven efficient development of the Longwangmiao Fm large-scale sulfur gas reservoir in Moxi block, Sichuan Basin. En: NATURAL GAS INDUSTRY B. vol. 3, no. 2, p. 99-107

MAHDIANPARI,Masoud, et al. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. En: ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING. vol. 130, p. 13-31

MAHDIANPARI,Masoud, et al. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. En: REMOTE SENSING. vol. 10, no. 7, p. 1119

MARTHA,Tapas Ranjan, et al. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 49, no. 12, p. 4928-4943

MCCLOSKEY,David; CHARNIAK,Eugene and JOHNSON,Mark. Effective self-training for parsing. 2006. p. 152-159

MEIER,Raphael, et al. Patient-specific semi-supervised learning for postoperative brain tumor segmentation. 2014. p. 714-721

MEKHALFI,Mohamed L., et al. Land-use classification with compressive sensing multifeature fusion. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 12, no. 10, p. 2155-2159

MISHRA,Niti B. and CREWS,Kelley A. Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with Random Forest. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 35, no. 3, p. 1175-1198

MOUSTAKIDIS,Serafeim, et al. SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 50, no. 1, p. 149-169

OJALA,Timo; PIETIKAINEN,Matti and MAENPAA,Topi. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. En: IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. vol. 24, no. 7, p. 971-987

PLAZAS,Miguel; RAMOS-POLLÁN,Raúl and MARTÍNEZ,Fabio. Ensemble-based approach for semisupervised learning in remote sensing. En: JOURNAL OF APPLIED REMOTE SENSING. vol. 15, no. 3, p. 1 – 16

QI,Kunlun, et al. Land-use scene classification in high-resolution remote sensing images using improved correlatons. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 12, no. 12, p. 2403-2407

SAJJADI,Mehdi; JAVANMARDI,Mehran and TASDIZEN,Tolga. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. 2016. p. 1163-1171

SCOTT,Grant J., et al. Training deep convolutional neural networks for land–cover classification of high-resolution imagery. En: IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. vol. 14, no. 4, p. 549-553

SHAFAEY,Mayar A., et al. Comparison of CNNs for remote sensing scene classification. 2018. p. 27-32

SHENG,Guofeng, et al. High-resolution satellite scene classification using a sparse coding based multiple feature combination. En: INTERNATIONAL JOURNAL OF REMOTE SENSING. vol. 33, no. 8, p. 2395-2412

SUBRAMANYA,A. and TALUKDAR,P. P. Graph-Based Semi-Supervised Learning. Morgan & Claypool, 2014.

SWAIN,Michael J. and BALLARD,Dana H. Color indexing. En: INTERNATIONAL JOURNAL OF COMPUTER VISION. vol. 7, no. 1, p. 11-32

SZEGEDY,Christian, et al. Rethinking the inception architecture for computer vision. 2016. p. 2818-2826

TAN,Chuanqi, et al. A Survey on Deep Transfer Learning. En: CoRR. vol. abs/1808.01974,

TRIGUERO,Isaac; GARCIA,Salvador and HERRERA,Francisco. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. En: KNOWLEDGE AND INFORMATION SYSTEMS. vol. 42, no. 2, p. 245-284

WANG,Chen, et al. Semi-Supervised Learning-Based Remote Sensing Image Scene Classification Via Adaptive Perturbation Training. p. 541-544

XIA,Gui-Song, et al. AID: A benchmark data set for performance evaluation of aerial scene classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 55, no. 7, p. 3965-3981

XIA,Gui-Song, et al. Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. En: REMOTE SENSING. vol. 7, no. 11, p. 15014-15045

YALNIZ,I. Zeki, et al. Billion-scale semi-supervised learning for image classification. En: arXiv PREPRINT arXiv:1905.00546.

YAN,Peiyao, et al. Semi-supervised representation learning for remote sensing image classification based on generative adversarial networks. En: IEEE ACCESS. vol. 8, p. 54135-54144

YAROWSKY,David. Unsupervised word sense disambiguation rivaling supervised methods. 1995. p. 189-196

YUAN,Qiangqiang, et al. Deep learning in environmental remote sensing: Achievements and challenges. En: REMOTE SENSING OF ENVIRONMENT. vol. 241, p. 111716

ZHANG,Fan; DU,Bo and ZHANG,Liangpei. Saliency-guided unsupervised feature learning for scene classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. vol. 53, no. 4, p. 2175-2184

ZHANG,Jun, et al. Domain Adaptation Based on Correlation Subspace Dynamic Distribution Alignment for Remote Sensing Image Scene Classification. En: IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

ZHOU,Dengyong, et al. Learning with local and global consistency. 2004. p. 321-328

ZHOU,Zhao, et al. Satellite image scene classification via ConvNet with context aggregation. 2018. p. 329-339

ZHU,Qiqi, et al. A deep-local-global feature fusion framework for high spatial resolution imagery scene classification. En: REMOTE SENSING. vol. 10, no. 4, p. 568

ZHU,Xiaojin. Semi-Supervised Learning: From Gaussian Fields to Gaussian Processes Xiaojin Zhu John Lafferty Zoubin Ghahramani August 18, 2003.