Explicación de la segmentación semántica no supervisada para la detección de trastornos hematológicos.

por

Jorge Rodolfo Gómez Arreola

Tesis sometida como requisito parcial para obtener el grado de

MAESTRÍA EN CIENCIAS Y TECNOLOGÍAS BIOMÉDICAS

en el

Instituto Nacional de Astrofísica, Óptica y Electrónica
Septiembre 2023
Santa María Tonantzintla, San Andrés Cholula, Puebla, México

Bajo la supervisión de:

Dra. Raquel Díaz Hernández
Investigadora Titular INAOE

Agradecimientos

Deseo expresar mi más sincero agradecimiento al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) por la invaluable beca otorgada a lo largo de estos dos años, la cual ha sido fundamental para respaldar mis estudios y mi crecimiento académico.

No puedo dejar de agradecer de manera especial a mi asesora, la Dra. Raquel Díaz Hernández, cuya paciencia, dedicación y constante apoyo han sido una guía fundamental desde el inicio de mi travesía en el programa de maestría. Su sabiduría y consejos han sido una fuente inestimable de aprendizaje.

No puedo pasar por alto reconocer al Dr. Leopoldo Altamirano Robles, cuyo apoyo incansable y orientación experta han sido esenciales en la elaboración de este trabajo de tesis.

Por último, pero no menos importante, extiendo mi gratitud a mi familia y amigos. Su amor, aliento y constante cariño han sido mi inspiración constante. Sin su apoyo incondicional, este logro tan significativo no hubiera sido posible.

Cada uno de estos agradecimientos es un reflejo de la red de apoyo y aliento que me ha impulsado a alcanzar este importante hito en mi carrera académica.

Resumen

La segmentación semántica no supervisada es una técnica en el campo de la visión por computadora que se utiliza para asignar automáticamente etiquetas o clases a los píxeles de una imagen. En el contexto de las imágenes sanguíneas, esta técnica desempeña un papel crucial en el diagnóstico, la automatización, la investigación y la toma de decisiones clínicas en el campo de los trastornos hematológicos. Su importancia radica en mejorar la precisión, la eficiencia y la comprensión de estas enfermedades, lo cual constituye la motivación principal de esta tesis. Para llevar a cabo este estudio, se utilizan imágenes provenientes de la base de datos PKG - C-NMC_Leukemia, cuya composición se describe detalladamente en el presente documento. El enfoque utilizado se basa en una red neuronal convolucional (CNN) que se entrena con imágenes en formato RGB. Además, se emplea el método de inteligencia artificial explicativa RISE (Randomized Input Sampling for Explanation) para generar mapas de calor y así obtener una comprensión más nítida de las decisiones efectuadas por el modelo; en el proceso de generación de mapas de calor mediante el método RISE, se utilizaron conjuntos de 3000, 5000 y 7000 máscaras aleatorias. Cada una de estas máscaras contribuyó a resaltar regiones particulares en las imágenes sanguíneas, subrayando así su importancia para la CNN. Este enfoque, al emplear diferentes cantidades de máscaras, permitió una exploración detallada de las áreas críticas que influyen en las decisiones del modelo. Es importante destacar que este enfoque se basa en el uso de datos no etiquetados y utiliza métodos de aprendizaje automático e inteligencia artificial para identificar y agrupar regiones en la imagen en función de sus características visuales, como color, textura y forma. La combinación de la segmentación semántica no supervisada y la inteligencia artificial explicativa permite identificar y delimitar las regiones de interés en las imágenes sanguíneas, mejorando así el diagnóstico de los trastornos hematológicos.

Abstract

Unsupervised semantic segmentation is a technique in the field of computer vision that is used to automatically assign labels or classes to pixels in an image. In the context of blood images, this technique plays a crucial role in the diagnosis, automation, research, and clinical decision-making in the field of hematological disorders. Its significance lies in improving the accuracy, efficiency, and understanding of these diseases, which is the main motivation of this thesis. To carry out this study, images from the PKG - C-NMC_Leukemia database are used, whose composition is described in detail in this document. The approach used is based on a convolutional neural network (CNN) that is trained with RGB format images. Additionally, the explanatory artificial intelligence method RISE (Randomized Input Sampling for Explanation) is employed to generate heatmaps that visualize the pixels relevant to the decisions made by the neural network. It is important to note that this approach is based on the use of unlabeled data and utilizes machine learning and artificial intelligence methods to identify and group regions in the image based on their visual characteristics such as color, texture, and shape. The combination of unsupervised semantic segmentation and explanatory artificial intelligence allows for the identification and delineation of regions of interest in blood images, thereby improving the diagnosis of hematological disorders. Furthermore, this approach facilitates task automation, accelerates scientific research in the field, and provides relevant information for informed clinical decision-making.

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

La segmentación semántica es un algoritmo de aprendizaje profundo que asigna etiquetas o categorías a cada píxel de una imagen, permitiendo reconocer y segmentar objetos en diversas aplicaciones, como la conducción autónoma, la generación de imágenes médicas y la inspección industrial [1]. En el contexto específico del análisis de imágenes de frotis de sangre periférica, la segmentación semántica desempeña un papel crucial en el diagnóstico temprano de enfermedades como la leucemia y la anemia, así como en otros trastornos sanguíneos. Estas imágenes contienen tres tipos principales de células sanguíneas: glóbulos rojos (GR), glóbulos blancos (GB) y plaquetas. Para evaluar el estado de estas células, se realiza un hemograma completo (CBC) que mide el número y la calidad de cada componente en las imágenes del frotis de sangre. La identificación precisa de los diferentes tipos de células es fundamental para el tratamiento y la recuperación de los pacientes. Sin embargo, el recuento manual de células sanguíneas en las imágenes del frotis de sangre puede ser agotador para los especialistas, lo que destaca la necesidad de un sistema asistido por computadora [2]. La segmentación semántica no supervisada se presenta como una técnica prometedora en este contexto, ya que permite identificar y segmentar automáticamente las células sanguíneas sin la necesidad de datos de entrenamiento previamente etiquetados. Sin embargo, uno de los desafíos de esta técnica radica en la interpretación de los resultados, debido a la naturaleza de caja negra de los modelos de aprendizaje profundo, lo cual dificulta comprender cómo y por qué se toman decisiones específicas.

La Inteligencia Artificial Explicable (XAI) desempeña un papel fundamental en este contexto. XAI se refiere a un conjunto de técnicas y herramientas que permiten a los seres humanos comprender y explicar las decisiones tomadas por los sistemas de inteligencia artificial [3]. Con XAI, es posible obtener explicaciones claras y detalladas sobre los resultados generados por los modelos de aprendizaje profundo, lo que a su vez aumenta la transparencia y la confianza en dichos resultados.

En resumen, la combinación de la segmentación semántica no supervisada y la XAI se convierte en una poderosa herramienta para la identificación y segmentación de

diversos tipos de células en muestras médicas, particularmente en el diagnóstico de trastornos hematológicos. Esta técnica puede contribuir significativamente a mejorar la precisión y eficacia del diagnóstico.

1.1. Problema de investigación

Los avances más recientes en el campo de la segmentación semántica enfrentan desafíos relacionados con la disponibilidad de datos etiquetados y la falta de transparencia en el proceso de toma de decisiones de los modelos. En particular, la falta de explicabilidad dificulta la comprensión de los factores clave considerados por la red para llegar a una decisión específica.

Sin embargo, al combinar la segmentación semántica no supervisada con técnicas de XAI, se pueden superar estas limitaciones [4], con el presente trabajo se busca obtener una visión más clara y detallada de los procesos de identificación y segmentación de células hematológicas. Al eliminar la necesidad de datos etiquetados, se abre la posibilidad de utilizar una mayor cantidad de datos e imágenes.

Además, la aplicación de técnicas de XAI proporciona una interpretación más granular de las decisiones tomadas por los algoritmos. Al visualizar y explicar los aspectos relevantes considerados por la red, se promueve la transparencia y el entendimiento del modelo. Esta combinación de técnicas tiene un potencial significativo para mejorar la precisión y la eficacia del diagnóstico hematológico.

1.2. Justificación

La segmentación semántica de imágenes de sangre desempeña un papel crucial en el ámbito de la medicina y la investigación biomédica al permitir la identificación y diferenciación precisa de diversas estructuras y elementos presentes en las muestras sanguíneas, como células sanguíneas, glóbulos blancos, glóbulos rojos y otros componentes relevantes para el diagnóstico y seguimiento de enfermedades. Sin embargo, los métodos tradicionales de segmentación semántica en imágenes médicas de cualquier tipo, requieren de un etiquetado manual intensivo y costoso, lo cual limita su aplicabilidad en grandes volúmenes de imágenes y puede conducir a resultados subjetivos y variables[5, 6, 7]. En respuesta a estas limitaciones, resulta imperativo desarrollar métodos de segmentación semántica no supervisada en este contexto. Al utilizar un enfoque no supervisado, se elimina la necesidad de un etiquetado manual, lo cual ahorra costos y tiempo, además de permitir el uso de grandes volúmenes de datos para el entrenamiento del modelo. Asimismo, al emplear algoritmos de inteligencia artificial explicativa, es posible superar las limitaciones mencionadas y obtener una visión más

clara y detallada de los procesos de identificación y segmentación de células hematológicas.

En resumen, la justificación de esta investigación radica en la necesidad de desarrollar métodos de segmentación semántica no supervisada mediante el uso de algoritmos de inteligencia artificial explicativa para imágenes de sangre. Estos métodos permiten superar las limitaciones de los enfoques tradicionales, objetivos y eficientes en términos de costo y tiempo.

1.3. Objetivos

1.3.1. Objetivo General

Investigar y aplica la técnica de Inteligencia Artificial Explicable (XAI) Randomized Input Sampling for Explanation (RISE) a un modelo de aprendizaje profundo en segmentación semántica no supervisada enfocado a anomalías en imágenes sanguíneas.

1.3.2. Objetivos Específicos

- Obtener un conjunto de datos en imágenes de célula sanguíneas. La base de datos servirá para entrenamiento y pruebas de las redes neuronales.
- Seleccionar y entrenar un modelo de aprendizaje profundo enfocado a segmentación no supervisada.
- Complementar la red neuronal convolucional con un método de inteligencia artificial explicable para obtener las características más relevantes en las imágenes.
- Generar mapas de calor con inteligencia artificial explicativa de los resultados obtenidos con la segmentación semántica no supervisada.

1.4. Alcances y Limitaciones

La segmentación semántica utilizando algoritmos de XAI (inteligencia artificial explicada explícitamente) es una técnica de procesamiento de imágenes que puede producir resultados precisos y útiles en ciertas situaciones. Sin embargo, también tiene sus limitaciones.

1.4.1. Alcances

- La segmentación semántica no supervisada usando algoritmos de XAI, puede ser útil para tareas de segmentación de imágenes en las que no hay suficientes datos etiquetados para un aprendizaje supervisado.

- Los algoritmos o modelos de Deep Learning para la segmentación semántica no supervisada pueden identificar automáticamente patrones y características en las imágenes que son relevantes para la tarea de segmentación.
- El uso de XAI (Inteligencia Artificial Explicable) es de gran importancia, ya que permite interpretar y explicar los resultados obtenidos por el modelo. Esto es especialmente útil en aplicaciones donde la comprensión humana es necesaria.

1.4.2. Limitaciones

- La segmentación semántica no supervisada puede tener una precisión limitada en comparación con los enfoques supervisados que utilizan datos etiquetados. Esto se debe a que los algoritmos no tienen una guía explícita sobre qué objetos o regiones deben ser segmentados.
- Estos algoritmos pueden requerir grandes cantidades de datos y poder computacional para obtener resultados precisos y útiles. Esto puede ser un desafío en aplicaciones prácticas con limitaciones de recursos.
- La segmentación semántica no supervisada puede ser sensible a la variación en la iluminación, la perspectiva y la escala, lo que puede afectar su precisión y confiabilidad en diferentes condiciones.
- La interpretación de los resultados de los algoritmos de XAI puede ser compleja y requerir conocimientos especializados en el campo. Además, los resultados pueden ser subjetivos.

1.5. Contribución

La contribución de este trabajo radica en nuestra capacidad para comprender mejor los modelos de caja negra mediante el uso de un método de inteligencia artificial explicativa, sin necesidad de modificar la arquitectura original de la red neuronal. Además, presentamos los primeros mapas de calor para redes neuronales convolucionales en segmentación semántica no supervisada utilizando el método explicativo RISE. Nos centramos específicamente en la detección de leucemia linfoblástica aguda, ya que uno de los principales desafíos durante el desarrollo de este trabajo fue la escasez de información disponible sobre el tema.

1.6. Estructura de la tesis

Con el fin de dar una descripción detallada de los conceptos necesarios para la realización de esta tesis, se utiliza la siguiente estructura:

Capítulo 1:Introducción. En esta sección, se proporciona una descripción detallada del problema abordado en esta tesis, así como su justificación, exposición clara de la problemática que lleva a investigar y trabajar en este tema, objetivos que se pretenden lograr con este trabajo. Asimismo, se identifican y destacan los alcances y limitaciones existentes, de la misma manera se establece la contribución final.

Capítulo 2: Marco teórico. Se exponen los fundamentos teóricos que respaldan el desarrollo de esta tesis. Se brindan conceptos generales de las técnicas utilizadas, proporcionando una base sólida para comprender los enfoques y metodologías empleadas en el estudio. Se presentan de manera clara y concisa los conceptos teóricos relevantes relacionados con las técnicas aplicadas. Permitiendo establecer una comprensión adecuada de los modelos teóricos y los fundamentos metodológicos utilizados en el desarrollo de la investigación.

Capítulo 3:Trabajo Relacionado. Se ofrece una descripción de los trabajos más relevantes relacionados con el tema de la presente tesis. En primer lugar, se abordan los temas relacionados con la segmentación semántica no supervisada, donde se exploran los avances y las investigaciones más destacadas en este campo. Posteriormente, se presentan los trabajos relevantes que se centran en la aplicación de técnicas de inteligencia artificial explicativa a la segmentación de imágenes.

Capítulo 4: Trabajo Desarrollado. En este capítulo, se detalla el proceso utilizado para realizar la segmentación semántica no supervisada en imágenes de frotis sanguíneo. Se comienza presentando la base de datos utilizada y, en caso necesario, los métodos de preprocesamiento aplicados a las imágenes. Luego, se describen las técnicas basadas en redes neuronales convolucionales utilizadas en el proceso. Además, se aplica el método explicativo RISE para mejorar la comprensión de los resultados obtenidos. Finalmente, se presentan las conclusiones obtenidas en este capítulo.

Capítulo 5: Resultados y Discusiones: En este apartado se presentan los resultados principales obtenidos mediante el uso de un modelo de aprendizaje profundo seleccionado, así como de los diferentes métodos de inteligencia artificial explicativa descritos en el marco teórico. Se proporciona una descripción detallada de los resultados obtenidos, respaldada por los fundamentos teóricos correspondientes.

3**Capítulo 6: Conclusiones y trabajo futuro:** Capítulo final en donde se presentan las conclusiones, observaciones, dificultades, etc., encontradas durante el desarrollo del trabajo y se proporciona una descripción de las acciones que se pueden llevar a cabo en el corto, mediano y largo plazo, basadas en los resultados obtenidos.

Al final de este documento se encuentran listadas las referencias bibliográficas más relevantes que respaldan la realización de esta tesis.

Capítulo 2

Marco Teórico

2.1. Introducción

En este capítulo se hace un análisis sobre el estudio e importancia de la detección de trastornos hematológicos. Posteriormente, se describe la importancia de la técnica de segmentación no supervisada usando diferentes métodos tales como algoritmos o redes neuronales. Finalizando con la sección de Inteligencia Artificial Explicable XAI aplicada a la segmentación semántica no supervisada, describiendo los métodos explicativos LIME, SHAP Y RISE.

2.2. Trastornos hematológicos

La leucemia es un tipo de cáncer que afecta los tejidos que producen la sangre, como la médula ósea y los glóbulos blancos. Se caracteriza por una producción descontrolada de células sanguíneas anormales, que reemplazan a las células sanas y dificultan el funcionamiento normal del sistema inmunológico.

Existen varios tipos de leucemia, siendo los principales la leucemia mieloide aguda (LMA), la leucemia mieloide crónica (LMC), la leucemia linfocítica aguda (LLA) y la leucemia linfocítica crónica (LLC). Cada tipo se diferencia por el tipo de células sanguíneas afectadas y su velocidad de progresión. Los síntomas de la leucemia pueden variar, pero algunos de los más comunes incluyen fatiga, debilidad, pérdida de peso, fiebre, sudoración nocturna, sangrado fácil, moretones frecuentes y aumento de los ganglios linfáticos.

2.3. Clasificación de leucemias

El sistema actual utilizado para clasificar la leucemia se basa en la clasificación de neoplasias hematopoyéticas de la Organización Mundial de la Salud (OMS) en su versión

de 2016. Esta clasificación se establece en función de una combinación de características clínicas, morfológicas, inmunofenotípicas y genéticas. [9]

Las leucemias también se clasifican comúnmente como:

- **Agudas o crónicas:** Según el porcentaje de blastos o células de leucemia en la médula ósea o la sangre.

- **Mieloide o linfoide:** Según la línea predominante de las células malignas.

Para el año 2022, la Sociedad Americana del Cáncer estima la distribución de nuevos casos de leucemia en los Estados Unidos por tipo de la siguiente manera [10]: Leucemia mieloide aguda (AML): 33 %, Leucemia linfoblástica aguda (ALL): 11 %, Leucemia mieloide crónica (CML): 15 %, Leucemia linfocítica crónica (CLL): 33 %, Otras leucemias: 8 %. Las leucemias más comunes y sus características distintivas se resumen en el cuadro 2.1.

Características	**Linfoblástica Aguda**	**Mieloide Aguda**	**Linfocítica Crónica**	**Mieloide Crónica**
Edad de mayor incidencia	Infancia	Cualquier edad	Edad media y avanzada	Edad adulta
Recuento glóbulos blancos	Arriba del 50 %	Arriba del 60 %	Arriba del 98 %	Arriba del 100 %
Recuento diferencial glóbulos blancos	Muchos linfoblastos	Muchos mieloblastos	Pequeños linfocitos	Serie mieloide completa
Anemia	Grave en >90 %	Grave en >90 %	Leve en aproximadamente el 50 %	Leve en el 80 %
Plaquetas	Bajas en >80 %	Grave en >90 %	Bajas entre 20 y 30 %	Altas en 60 % y bajas en 10 %
Linfadenopatía	Común	Ocasional	Común	Inexistente
Esplenomegalia	50 %	60 %	Moderado	Grave
Otras características	Sistema nervioso no se afecta	presencia cuerpos de auer	Ocasionalmente anemia	Cromosoma Filadelfia positivo en >90 %

Cuadro 2.1: Hallazgos al momento del diagnóstico en las leucemias más comunes. [10]

2.4. Segmentación de imágenes

La segmentación de imágenes ha recibido mucha atención en el campo de investigación de la visión por computadora en los últimos tiempos. Algunas de las aplicaciones más importantes incluyen la detección de objetos, el reconocimiento de texturas, entre otras. Existen diversas técnicas para llevar a cabo la segmentación, y en el enfoque supervisado, se utiliza un conjunto de datos que consiste en pares de imágenes y etiquetas semánticas a nivel de píxel para el entrenamiento. El objetivo principal es entrenar un sistema que pueda clasificar las etiquetas de un conjunto conocido de categorías para los píxeles de la imagen.[12]

En contraste, la segmentación no supervisada se centra más en predecir las etiquetas de píxeles de manera más global. Una vez que se obtiene la representación de características a nivel de píxel, los segmentos de la imagen se pueden obtener agrupando los vectores de características. Sin embargo, el diseño de la representación de características sigue siendo un desafío. La representación de características deseada depende en gran medida del contenido de la imagen objetivo.[13]

En los últimos tiempos, las redes neuronales convolucionales (CNN) se han aplicado con éxito a la segmentación semántica de imágenes en escenarios de aprendizaje supervisado, como la conducción autónoma y los juegos de realidad aumentada. Si bien las CNN no se utilizan con frecuencia en escenarios completamente no supervisados, tienen un gran potencial para extraer características detalladas de los píxeles de la imagen, lo cual es necesario para la segmentación de imágenes sin supervisión. [14]

2.4.1. Métodos de segmentación

Desde la década de 1970, la segmentación de imágenes ha recibido atención continua por parte de los investigadores en visión por computadora. Los métodos clásicos de segmentación se centran principalmente en resaltar y obtener la información contenida en una sola imagen, lo cual a menudo requiere conocimientos profesionales e intervención humana. Sin embargo, es difícil obtener información semántica de alto nivel a partir de las imágenes. Los métodos de co-segmentación implican identificar objetos comunes en un conjunto de imágenes, lo cual requiere la adquisición de ciertos conocimientos previos. En particular, la segmentación semántica aún presenta desafíos debido a la anotación limitada o dispersa, el desequilibrio de clases, el sobreajuste, el largo tiempo de entrenamiento y la desaparición del gradiente. Por lo tanto, es necesario resumir de manera sistemática los métodos de segmentación existentes.

A continuación, realizamos un análisis y una lista de los métodos existentes de segmentación de imágenes, y presentamos de manera sistemática las técnicas esenciales de segmentación semántica basadas en redes neuronales profundas, tal como se muestra en la Figura 2.1.

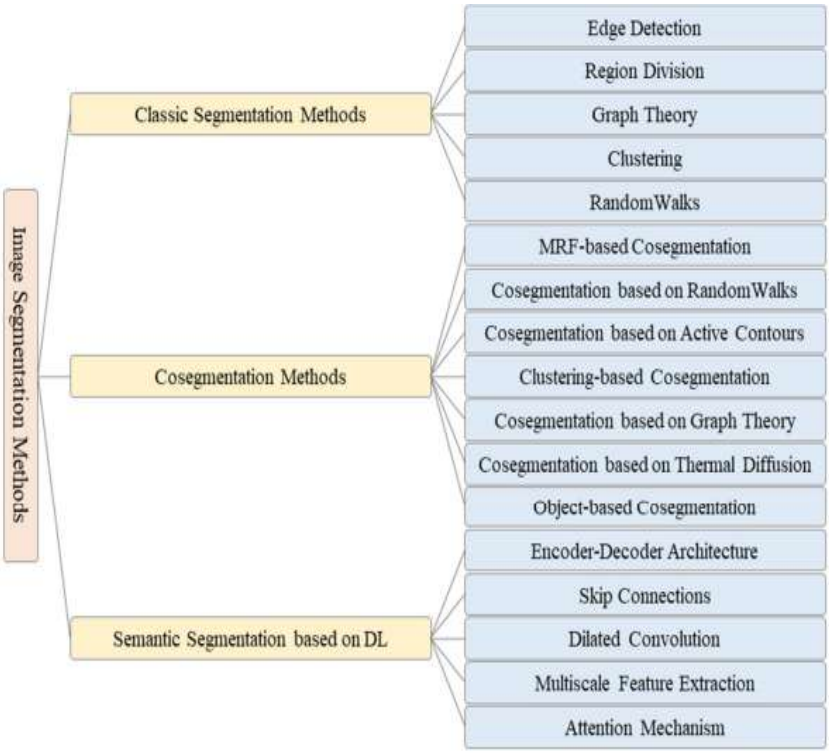


Figura 2.1: Categorías de los diferentes métodos de segmentación. [15]

2.4.2. Métodos basados en detección de bordes

La detección de bordes tiene como objetivo identificar los puntos de cambio brusco en el nivel de gris de una imagen, los cuales suelen representar los límites entre diferentes regiones. Este método, también conocido como técnica de bordes paralelos, es uno de los métodos más antiguos de segmentación. Hay varios métodos disponibles para llevar a cabo la detección de bordes. [15]

Método de Canny

Yu. Y et al. [15] resume que el método de Canny es altamente efectivo en la eliminación de ruido y la segmentación de líneas con continuidad, finura y rectitud. Sin embargo, este operador es más complejo y tarda más tiempo en ejecutarse. En la producción industrial real, se utiliza generalmente un umbral de gradiente para el procesamiento en tiempo real de alta velocidad, mientras que para obtener resultados de alta calidad

se opta por el operador de Canny más avanzado. Aunque los operadores diferenciales pueden localizar eficientemente los límites de diferentes regiones, no pueden garantizar la continuidad y el cierre de los bordes debido a la presencia de numerosos puntos y líneas discontinuas en regiones de alta resolución. Por ende, se selecciona el operador de Canny para mayor resolución.

Método de detección de bordes en serie

Asimismo, Yu. Y. et. al. [15] menciona que la técnica de bordes en serie, que consiste en la concatenación de puntos de bordes para formar un límite cerrado. Las técnicas de bordes en serie incluyen principalmente algoritmos de búsqueda en gráficos y algoritmos de programación dinámica. En los algoritmos de búsqueda en gráficos, los puntos de los bordes se representan mediante una estructura de grafo, y se busca el camino con el costo mínimo en el grafo para determinar los límites cerrados. Sin embargo, este enfoque suele requerir un alto costo computacional. Por otro lado, el algoritmo de programación dinámica utiliza reglas heurísticas para reducir la carga computacional durante la búsqueda.

Método de contornos activos

En la reseña llevada a cabo por Yu [15], se destaca el último enfoque basado en bordes, conocido como contornos activos, serpientes o snakes. Consiste en ajustar una curva cerrada inicial, basada en el gradiente de la imagen, a las características locales de la misma. El objetivo es encontrar la curva cerrada que minimice una función de energía, logrando así la segmentación precisa de la imagen.

Este método es altamente sensible a la ubicación del contorno inicial, por lo que es crucial que la inicialización esté cerca del contorno objetivo para obtener resultados precisos. Mediante la iteración y adaptación de la curva, los contornos activos pueden ajustarse y adaptarse a los detalles de la imagen, logrando así una segmentación más precisa de los objetos de interés

2.4.3. Métodos basados en la generación de superpíxeles

Los superpíxeles son una serie de pequeñas áreas irregulares compuestas por píxeles con posiciones y características similares (como brillo, color y textura). El uso de superpíxeles en lugar de píxeles individuales para representar características puede reducir la complejidad del procesamiento de imágenes, por lo que se utiliza frecuentemente en la etapa de preprocesamiento de la segmentación de imágenes. Los métodos de segmentación de imágenes basados en la generación de superpíxeles incluyen principalmente técnicas de agrupamiento y teoría de grafos.[15]

Teoría de grafos

Los métodos de segmentación de imágenes basados en teoría de grafos utilizan una representación de la imagen como un grafo, donde los nodos representan los píxeles o regiones de la imagen, y las aristas representan las relaciones entre ellos.

Utilizando las propiedades y algoritmos de la teoría de grafos, se han desarrollado varios métodos de segmentación. El algoritmo GrabCut [16], por ejemplo, emplea una técnica iterativa que combina datos de color y posición para separar una imagen en áreas claras y oscuras. Otra técnica posible es Graph Cut [17], que implementa la teoría de grafos para identificar la partición óptima de una imagen en regiones y luego utiliza un algoritmo iterativo destinado a disminuir la función de energía. Para una segmentación precisa, es útil implementar métodos que puedan captar la estructura general y la coherencia de una imagen, lo que les permite capturar tanto la similitud local como las relaciones globales entre regiones o píxeles. Aunque se deben usar algoritmos eficientes para lograr resultados dentro de un marco de tiempo razonable, la complejidad computacional de estos métodos puede ser alta, particularmente para imágenes grandes.

2.4.4. Métodos basados en agrupamiento

Los métodos de segmentación de imágenes basados en agrupamiento se enfocan en agrupar píxeles o regiones de una imagen en conjuntos homogéneos que comparten características similares.

Estos métodos utilizan técnicas de agrupación para dividir las imágenes en segmentos o regiones importantes. Uno de estos algoritmos es el agrupamiento de K-means, se basa en el algoritmo de Lloyd y funciona de la siguiente manera: (i) se inicializan K puntos como centros de conglomerados; (ii) se calculan las distancias entre cada punto i de la imagen y los K centros de conglomerados, elija el que tenga la menor distancia como clasificador k_i; (iii) promedie los puntos en cada categoría (centroide) y mueva el centro del clúster hacia el centroide; (iv) repita los pasos (ii) y (iii) hasta que el algoritmo converja. En resumen, K-means es un proceso iterativo de centros de clúster de computación. Aunque el algoritmo tiene una velocidad de convergencia rápida y es resistente al ruido, no es adecuado para tratar áreas no adyacentes y solo puede converger a una solución óptima local, pero no puede converger a una solución óptima global.

2.4.5. Redes neuronales y visión por computadora

La visión por computadora es una disciplina en la que los modelos de aprendizaje profundo han contribuido significativamente a mejorar la precisión de los sistemas [18, 21]. En el campo de la salud, la aplicación de la inteligencia artificial para la interpretación de imágenes médicas es una área de investigación ampliamente estudiada. Existen nu-

merosos ejemplos concretos de cómo el aprendizaje profundo y la visión por computadora han sido aplicados en este contexto. En este trabajo en particular, nos centraremos en un tipo específico de red neuronal convolucional (CNN, por sus siglas en inglés), que es uno de los enfoques más avanzados en el campo de la visión por computadora. Las CNN han demostrado un rendimiento comparable al humano en muchas tareas de diagnóstico basadas en imágenes [18, 21]. Estas redes están compuestas por capas convolucionales, de agrupamiento y completamente conectadas. La capa convolucional es especialmente relevante para la identificación de patrones, líneas y bordes en las imágenes [20, 21]. Por otro lado, las capas de agrupamiento reducen el número de características mediante la agregación de características similares. En general, las CNN capturan diferentes representaciones a lo largo de sus capas, aprendiendo características específicas de la imagen [21].

2.5. Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son un tipo de arquitectura de redes neuronales especializadas en el procesamiento de datos estructurados, como imágenes y señales. A diferencia de las redes neuronales tradicionales, las CNN están diseñadas específicamente para aprovechar la estructura espacial presente en los datos de entrada. Las redes neuronales convolucionales han demostrado ser muy eficaces en tareas de visión por computadora, como clasificación de imágenes, detección de objetos y segmentación semántica. Su capacidad para aprender y reconocer patrones visuales complejos las ha convertido en una herramienta fundamental en campos como la inteligencia artificial, la medicina, la conducción autónoma y muchos otros.

2.5.1. Arquitectura SegNet

SegNet es una arquitectura de red desarrollada específicamente para la tarea de segmentación semántica de imágenes. Fue propuesta por Badrinarayanan et. al [22] y ha sido ampliamente utilizada en aplicaciones de segmentación en el campo de la visión por computadora. Utiliza capas convolucionales en su estructura para aprender características y realizar la segmentación de manera efectiva.

SegNet se basa en el concepto de codificador-decodificador y utiliza capas convolucionales para aprender características y realizar la segmentación de manera eficiente. La arquitectura de SegNet consta de un codificador que extrae características de la imagen de entrada utilizando capas convolucionales y capas de agrupación (pooling). Luego, utiliza una estructura de decodificador que realiza la reconstrucción de la imagen segmentada a partir de las características extraídas.[22]

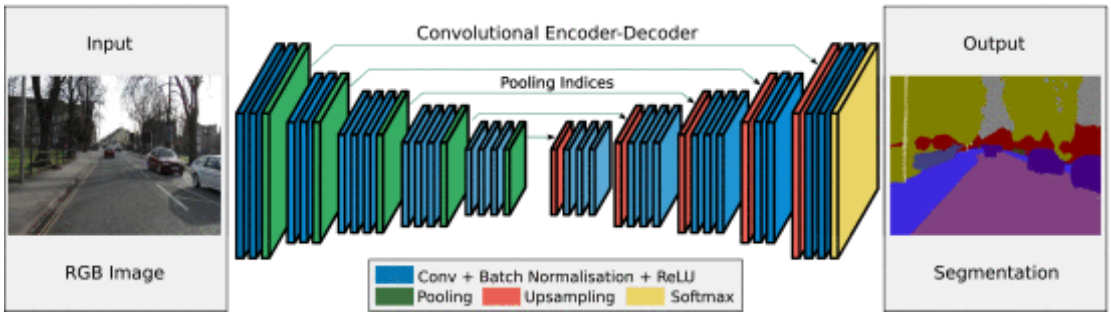


Figura 2.2: Ilustración de la arquitectura SegNet. No hay capas completamente conectadas, por lo que es únicamente convolucional [22].

2.5.2. Arquitectura U-Net

La arquitectura U-Net fue propuesta Ronnebeger et.al [23] y ha sido ampliamente utilizada en tareas de segmentación de imágenes biomédicas. La red U-Net es especialmente conocida por su capacidad para realizar segmentación semántica precisa, y se ha utilizado en aplicaciones como la segmentación de tejidos en imágenes médicas. La arquitectura de U-Net consta de un codificador, que captura características de la imagen de entrada utilizando capas convolucionales y capas de agrupación (pooling), y un decodificador, que realiza la reconstrucción de la imagen segmentada a partir de las características extraídas [23].

Lo que distingue a U-Net de otras arquitecturas de CNN es su estructura de U simétrica, donde las características extraídas en el codificador se concatenan con las características en el decodificador, lo que permite la fusión de detalles de alta resolución en la etapa de decodificación [23].

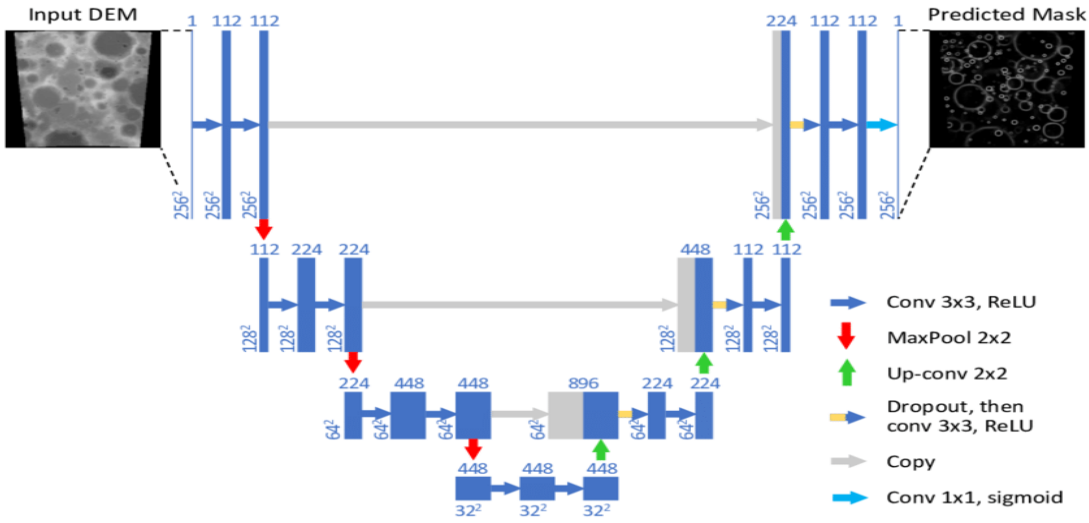


Figura 2.3: Arquitectura U-net (ejemplo para 32x32 píxeles en la resolución más baja). [24].

2.6. Métodos de inteligencia artificial explicable.

La Inteligencia Artificial Explicable (XAI, por sus siglas en inglés) se refiere al desarrollo de enfoques y técnicas que permiten comprender y explicar cómo los modelos de Inteligencia Artificial (IA) toman decisiones. A medida que los modelos de IA se vuelven más complejos, como las redes neuronales profundas, entender cómo llegan a sus predicciones se vuelve fundamental para asegurar la transparencia, la confianza y la responsabilidad en su aplicación. Entre los métodos utilizados en XAI se encuentran LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations) y RISE (Randomized Input Sampling for Explanation). Estos métodos buscan proporcionar explicaciones locales y comprensibles de las decisiones de los modelos de IA.

2.6.1. Local Interpretable Model-Agnostic Explanations (LIME)

El principal objetivo de LIME [25] es proporcionar un método explicativo que capture el comportamiento de nuestro modelo de inteligencia artificial basado en su instancia principal. Ya sea un clasificador, una red de segmentación u otro tipo de modelo, este método es capaz de ofrecer explicaciones sin necesidad de examinar su estructura interna, lo que significa que no se requiere modificar la arquitectura original de la red. En otras palabras, se adapta al modelo en cuestión.

En un primer momento, definimos una explicación como un modelo $\mathbf{g} \in \mathbf{G}$, donde $\mathbf{G}$ comprende modelos interpretables como árboles de decisión, listas de reglas, modelos lineales, entre otros. El dominio de $\mathbf{g}$ está compuesto por $\{0,1\}^{d'}$, lo cual puede indicar la presencia o ausencia de componentes importantes. Dado que no todos los elementos son interpretables, se permite una medida de complejidad en lugar de la interpretabilidad, que se denota como $\Omega(g)$ para los modelos.

Finalmente, denotamos al modelo que deseamos explicar como $f : \mathbb{R}^d \rightarrow \mathbb{R}$. En el caso de la clasificación, $f(x)$ representa la probabilidad de que x pertenezca a una determinada clase. Al utilizar $\Pi_x(z)$, medimos la proximidad de z a la instancia de datos x en términos de características o atributos, lo cual define una región local alrededor de x. Esta región local es importante porque nos permite concentrarnos en explicar únicamente una porción relevante para el atributo que deseamos explicar, en lugar de tratar de explicar toda la instancia de datos de una sola vez.

Por último, se define $L = (f, g, \Pi_x)$+$\Omega(g)$ para describir qué tan bien el modelo g se ajusta a la función f. Por lo tanto, la ecuación que representa la explicación generada por LIME se define de la siguiente manera:

$$\xi = argmin_{g \in G} L(f, g, \Pi_x) + \Omega(g) \tag{2.1}$$

La ecuación (2.1) busca encontrar el modelo g, dentro de la clase G, que minimiza una combinación de dos términos: $L = (f, g, \Pi_x)$, que mide qué tan bien se ajusta g a la función f en la región local Π_x y $\Omega(g)$, que representa una medida de complejidad o interpretabilidad del modelo g.

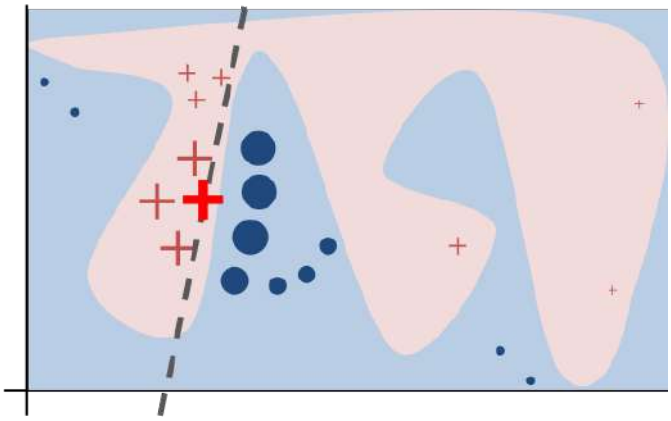


Figura 2.4: Ejemplo ilustrativo que presenta la intuición de LIME al proporcionar una explicación del modelo. LIME muestrea instancias, obtiene predicciones utilizando f y las pondera por la proximidad a la instancia que se está explicando. La línea punteada es la explicación aprendida que es localmente fidedigna. [25].

2.6.2. SHapley Additive exPlanations (SHAP)

El objetivo de SHAP [26] es explicar la predicción de una instancia x al calcular la contribución de cada característica en dicha predicción. Este método de explicación utiliza los valores de Shapley, que provienen de la teoría de juegos cooperativos. Estos valores de Shapley nos indican cómo distribuir de manera justa el pago (es decir, la predicción) entre las características. Al explicar una imagen, los píxeles pueden agruparse en superpíxeles y la predicción puede distribuirse entre ellos. La explicación SHAP se denota de la siguiente manera:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{2.2}$$

Donde g es el modelo a explicar, $z' \in \{0,1\}^M$ es el vector de coalición, es decir, un vector binario que indica ausencia o presencia de las características, M es el tamaño máximo de la coalición y ϕ_j es la atribución para una característica, por lo que la Ec. (2.2) se puede reescribir de la siguiente manera cuando tenemos únicamente 1's que indican presencia de las características.

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{2.3}$$

El método SHAP se basa en tres propiedades elementales que deben cumplir los valores SHAP:

- **Precisión Local:** La explicación debe ser precisa al explicar la salida del modelo para una instancia de datos específica.
- **Ausencia de características:** Las características que no están presentes en una instancia no deberían contribuir a la salida del modelo.
- **Consitencia:** Al intercambiar características en un modelo, la explicación de cada característica no debe cambiar.

Al comprender las ecuaciones (2.2) y (2.3), se puede apreciar qué características son relevantes y cómo influyen en la salida del modelo. Al utilizar métodos explicativos como SHAP, se logra proporcionar explicaciones claras y comprensibles, lo que contribuye a mejorar la transparencia y la confianza en los modelos de aprendizaje automático.

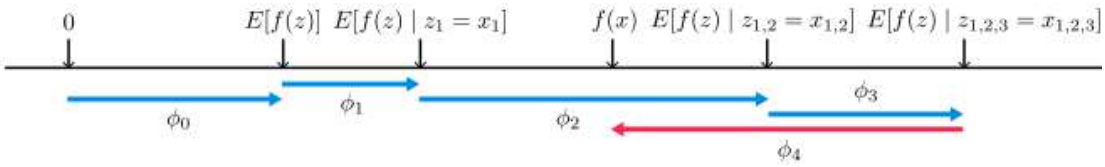


Figura 2.5: Valores SHAP atribuyen a cada característica el cambio en la predicción esperada del modelo. [26]

2.6.3. Randomized Input Sampling for Explanation (RISE)

Es un método de inteligencia artificial explicativo (XAI) similar a LIME o SHAP. Proporciona explicaciones interpretables sobre cómo un modelo de aprendizaje automático o aprendizaje profundo (ML o DL) llega a una decisión específica. A diferencia de LIME o SHAP, RISE utiliza un muestreo aleatorio de la imagen de entrada y aplica una máscara a la imagen antes de pasarla a través del modelo [27].

Una forma de medir la importancia de una región específica en una imagen es perturbarla y observar cómo afecta al modelo de caja negra. Por ejemplo, se pueden establecer las intensidades de los píxeles en cero, difuminar las regiones o aplicar ruido. Estos cambios permiten evaluar la sensibilidad del modelo a las perturbaciones en diferentes partes de la imagen.

El proceso de generación de máscaras en RISE implica realizar un muestreo aleatorio de combinaciones de píxeles para crear una máscara única en cada muestra. Estas máscaras se utilizan posteriormente para atenuar las regiones correspondientes en la imagen original antes de que sea ingresada al modelo. Al repetir este proceso varias veces, se obtiene una distribución de importancia para cada parte de la imagen, lo que permite identificar las regiones más relevantes para la predicción del modelo. Matemáticamente, el proceso es de la siguiente manera:

Sea $f : X \to \mathbb{R}$ un modelo de caja negra y sea $X = \{I | I : \Lambda \to \mathbb{R}^3\}$ el espacio de imágenes en color con tamaño H x W donde cada imagen I es la asignación de coordenadas a tres valores de color. Asi hacemos $M : \Delta \to \{0, 1\}$ sea una máscara bianria con distribución D. Considerando como la variable aleatoria $f(I \odot M)$ donde $\odot$ denota el producto elemento por elemento. Primero enmascarando la imagen conservando solo un subconjunto de píxeles. Luego, calculando el puntaje de confianza para la imagen enmascarada. Definiendo la importancia del pixel $\lambda \in \Lambda$ como el puntaje esperado sobre todas las posibles máscaras M condicionados al evento de que el pixel λ sea observado, $M(\lambda) = 1$.

$$S_{I,f}(\lambda) = \mathbb{E}_M[f(I \odot M) | M(\lambda) = 1] \tag{2.4}$$

El término $f(I \odot M)$ es alto cuando los píxeles conservados por la máscara M son importantes, por lo que la ecuación (2.4) puede ser reescrita de la siguiente manera:

$$\begin{aligned} S_{I,f}(\lambda) &= \sum_M f(I \odot M) P[M = m | M(\lambda) = 1] \\ &= \frac{1}{M(\lambda) = 1} \sum_m f(I \odot m) P[M = m, M(\lambda) = 1] \end{aligned} \tag{2.5}$$

Por lo que

$$P[M = m, M(\lambda) = 1] = \begin{cases} 0 & \text{si } m(\lambda) = 0 \\ P[M = m] & \text{si } m(\lambda) = 1 \end{cases} = m(\lambda) P[M = m] \tag{2.6}$$

Sustituyendo $P[M = m, M(\lambda) = 1]$ de la ecuación (2.6) en la ecuación (2.5):

$$S_{I,f}(\lambda) = \frac{1}{P[M(\lambda) = 1]} \sum_m f(I \odot m) \cdot m(\lambda) \cdot P[M = m] \tag{2.7}$$

Finalmente reescribimos en forma matricial, combinando el hecho de que $P[M(\lambda) = 1] = \mathbb{E}[M(\lambda)]$

$$S_{I,f} = \frac{1}{\mathbb{E}[M]} \sum_m f(I \odot m) \cdot m \cdot P[M = m] \tag{2.8}$$

Por lo tanto, el mapa de relevancia que explique la decisión del modelo f en la imagen I será calculado mediante la ecuación (2.8) como una suma ponderada de máscaras aleatorias. Luego, tomamos el promedio ponderado de las máscaras, donde los pesos son

las puntuaciones de confianza, y lo normalizamos por la esperanza dé M. El método RISE no utiliza ninguna información interna del modelo, por lo que es adecuado para explicar modelos de caja negra.

RISE se enfoca en proporcionar explicaciones a modelos de aprendizaje automático y aprendizaje profundo existentes. Dado que la segmentación semántica no supervisada no se basa en modelos predefinidos, sino en técnicas de agrupación, el método explicativo RISE no puede ser aplicado directamente. Sin embargo, es permitido utilizar técnicas basadas en RISE para la interpretación y visualización de resultados enfocados a la segmentación semántica no supervisada. Destacando el uso para resaltar las regiones más importantes en imágenes segmentadas una vez que alimentan al modelo. Esto podría ayudar a una mejor comprensión y proporciona una explicación visual más clara de los resultados de la segmentación.

Si bien RISE es una técnica útil para la explicabilidad de modelos de aprendizaje automático, es importante tener en cuenta sus limitaciones. Estas incluyen la necesidad de un alto poder computacional debido a la generación de múltiples muestras y la falta de garantía de que las explicaciones sean completamente precisas o abarquen todas las posibles justificaciones para una predicción. Sin embargo, sigue siendo un método valioso para comprender y analizar el comportamiento de los modelos de aprendizaje automático en el contexto de la visión por computadora.

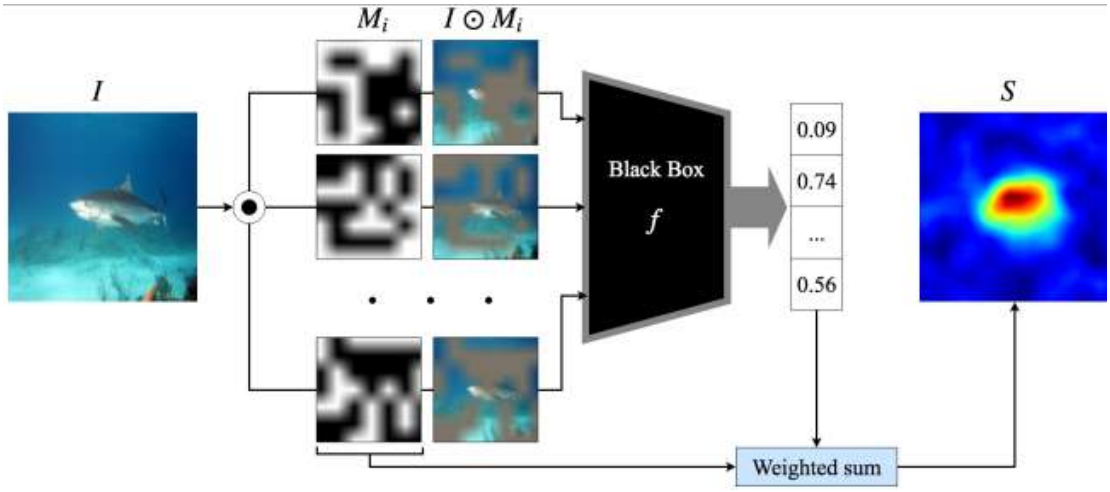


Figura 2.6: Resumen del método RISE para la obtención de los mapas de calor. [27]

Capítulo 3

Trabajo Relacionado

En este capítulo, se realiza una exhaustiva revisión de los trabajos relacionados que respaldan y contribuyen al desarrollo de la presente tesis. El enfoque principal se centra en la segmentación semántica utilizando métodos de aprendizaje no supervisado basados en redes neuronales, así como en la aplicación de métodos explicativos agnósticos para modelos de caja negra, es decir, aquellos en los que no es necesario realizar modificaciones adicionales a la arquitectura original de la red. Se lleva a cabo un análisis detallado de diversos trabajos relacionados con el campo de la segmentación semántica, abarcando tanto enfoques supervisados como no supervisados. Además, se examina el uso de redes neuronales convolucionales en esta área y se exploran los métodos explicativos para mejorar la comprensión de las decisiones tomadas por estas redes. Con el fin de proporcionar una visión general, se presenta en el cuadro 3.1 una comparativa de los trabajos relacionados con la temática abordada en esta tesis.

Autor y Año	**Segmentación Semántica**	**Segmentación Semántica No Supervisada**	**Segmentación Explicativa**	**Segmentación Semántica No Supervisada Explicativa**
Tran, T. 2018	✓			
Cho. J. 2021	✓	✓		
Alharbi, A. 2022	✓			
M. Hamilton. 2022	✓	✓		
Dardouillet. P. 2022	✓		✓	
J. Arreola 2023	✓	✓	✓	✓

Cuadro 3.1: Comparativa de los trabajos relacionados con la presente tesis.

3.1. Segmentación semántica en imágenes sanguíneas

En su trabajo, Tran. T. [2] aplica la tecnología de segmentación semántica de aprendizaje profundo para identificar y segmentar los glóbulos rojos y blancos en imágenes de extensiones de sangre utilizando la arquitectura SegNet [22]. Este enfoque clasifica todos los píxeles de una imagen y produce una imagen resultante segmentada por clases. Los resultados experimentales muestran que el modelo alcanza una precisión global del 89.45 %. Además, se logra una alta precisión en la segmentación de glóbulos blancos, glóbulos rojos y el fondo de las imágenes de extensiones de sangre, con valores de 94.93 %, 91.11 % y 87.32 %, respectivamente. Estos resultados resaltan la efectividad del enfoque propuesto y su potencial para aplicaciones clínicas en el campo de la hematología.

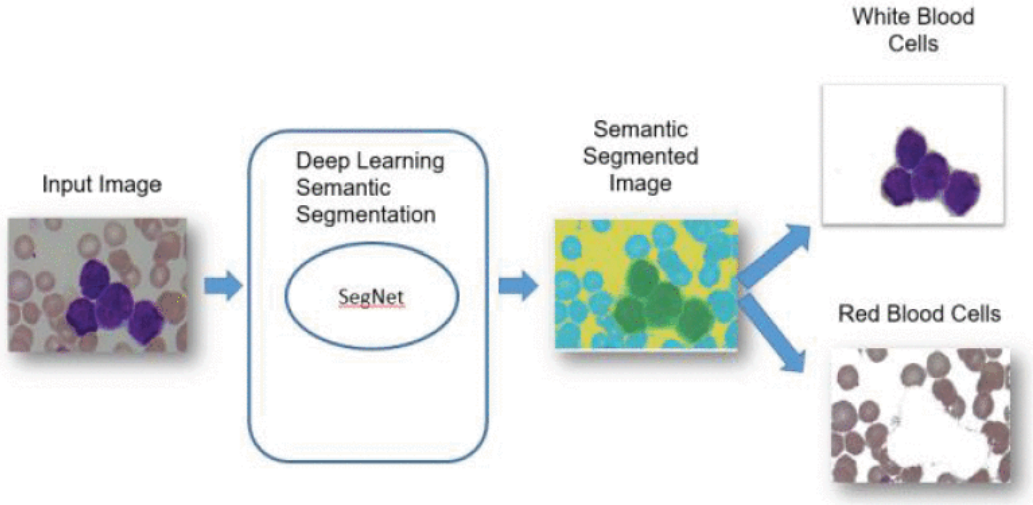


Figura 3.1: Representación del metolodgía de segmentación propuesta en [2].

De igual manera en su artículo de investigación Alharbi, A. [28] propone un modelo innovador que combina las redes ResNet y UNet para extraer características y segmentar los leucocitos en muestras de sangre. Los resultados experimentales muestran un desempeño sólido del modelo, lo que sugiere que es una herramienta adecuada para el análisis de datos en el campo de la hematología. Para evaluar el modelo, se utilizó un enfoque de validación cruzada en tres conjuntos de datos que contenían diferentes tipos de glóbulos blancos, utilizando conjuntos de datos públicos disponibles. La precisión general de segmentación del modelo propuesto alcanzó aproximadamente el 96 %, superando en rendimiento a enfoques anteriores.

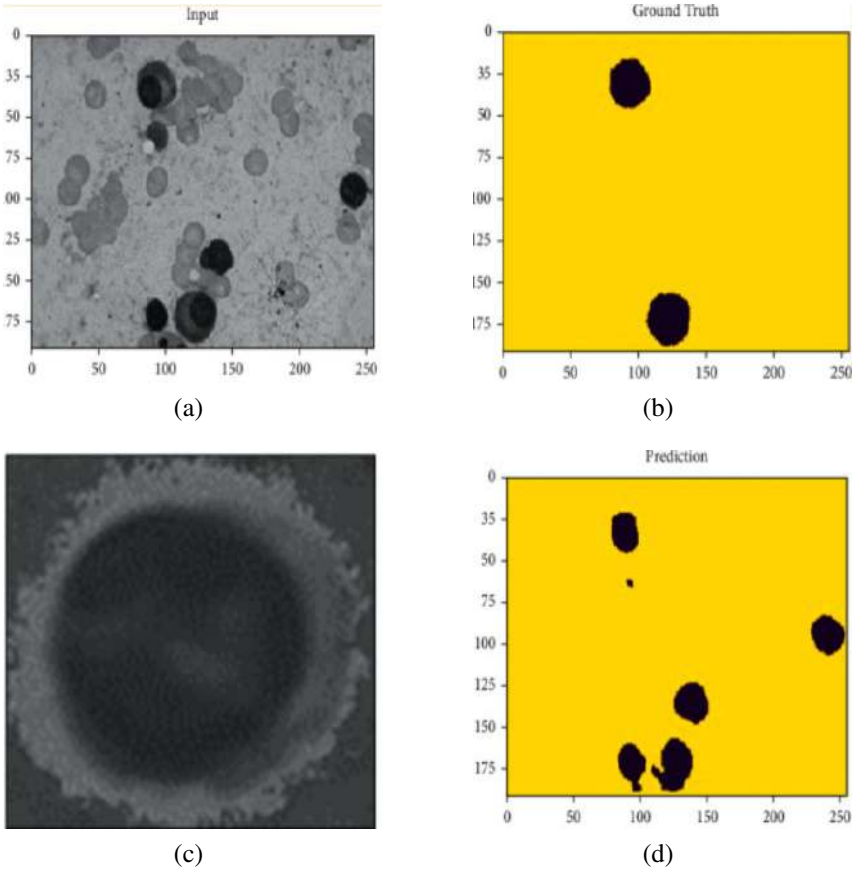


Figura 3.2: Muestras predichas del conjunto de datos. (a) Imagen de entrada. (b) Imagen de verdad de referencia. (c) Imagen de UNet. (d) Imagen del modelo propuesto. [28].

3.2. Segmentación semántica no supervisada

Inicialmente, se realiza una exhaustiva revisión de los trabajos relacionados con el uso de redes neuronales en el contexto de la segmentación semántica mediante aprendizaje no supervisado. Esta aproximación revolucionaria tiene como objetivo detectar áreas de interés en imágenes sin la necesidad de contar con etiquetas o anotaciones previas. Una de las técnicas más destacadas utilizadas en este campo es el empleo de redes neuronales convolucionales (CNN). Estas poderosas redes son entrenadas en grandes conjuntos de datos sin etiquetas, empleando sofisticadas técnicas como autoencoders o generación adversarial, con el fin de adquirir representaciones de alto nivel de las imágenes. Posteriormente, se aplican técnicas de clustering o segmentación para identificar meticulosamente las regiones de interés basadas en las características aprendidas por la red.

En su trabajo, Hamilton [29] plantea que la segmentación semántica no supervisada tiene como objetivo descubrir y localizar categorías semánticamente significativas dentro

de un conjunto de imágenes sin anotaciones previas. Para abordar este desafío, Hamilton presenta STEGO (Self-supervised Transformer with Energy-based Graph Optimization), un marco de trabajo innovador que extrae características no supervisadas y las convierte en etiquetas semánticas discretas de alta calidad. En el núcleo de STEGO se encuentra una función de pérdida contrastiva que impulsa la formación de grupos compactos entre características mientras se preservan sus relaciones en el conjunto de imágenes. Esta metodología ha logrado una mejora significativa con respecto al estado del arte previo en los desafíos de segmentación semántica de CocoStuff y Cityscapes.

Con STEGO, se logra obtener una representación semántica precisa y de alta calidad, sin depender de anotaciones previas, lo que permite una segmentación efectiva de objetos y categorías en imágenes sin supervisión. Este avance ofrece nuevas posibilidades en el campo de la segmentación semántica y contribuye al desarrollo de técnicas más avanzadas y precisas en el procesamiento de imágenes.

Figura 3.3: Predicciones de la segmentación semántica no supervisada usando la red STEGO [29].

Cho, J [30] presenta un nuevo marco de trabajo para la segmentación semántica sin anotaciones a través de técnicas de agrupamiento. Los métodos de agrupamiento convencionales están limitados a imágenes seleccionadas, de una sola etiqueta y centradas en objetos, mientras que los datos del mundo real son predominantemente no seleccionados, multietiqueta y centrados en escenas.

Se propone un método que utiliza la consistencia geométrica como una guía para aprender a reconocer y adaptarse a diferentes variaciones en la iluminación y en la forma de los objetos en una imagen. En otras palabras, se enseña al modelo ser insensible a cambios de brillo y a diferentes perspectivas o transformaciones geométricas en las

imágenes que procesa. Esto permite obtener resultados más consistentes y precisos en la segmentación semántica.

En resumen Cho.J [30] propone un enfoque innovador para la segmentación semántica sin anotaciones mediante el uso de técnicas de agrupamiento a nivel de píxel y la incorporación de consistencia geométrica. Con Pixel-level feature Clustering using Invariance and Equivariance (PiCIE), la segmentación precisa de categorías tanto de objetos, como de escenas, sin la necesidad de ajustes manuales o preparación específica para la tarea es alcanzable. Esto representa un avance significativo en el campo de la segmentación semántica y tiene el potencial de mejorar la precisión y eficiencia de las aplicaciones relacionadas con el procesamiento de imágenes.

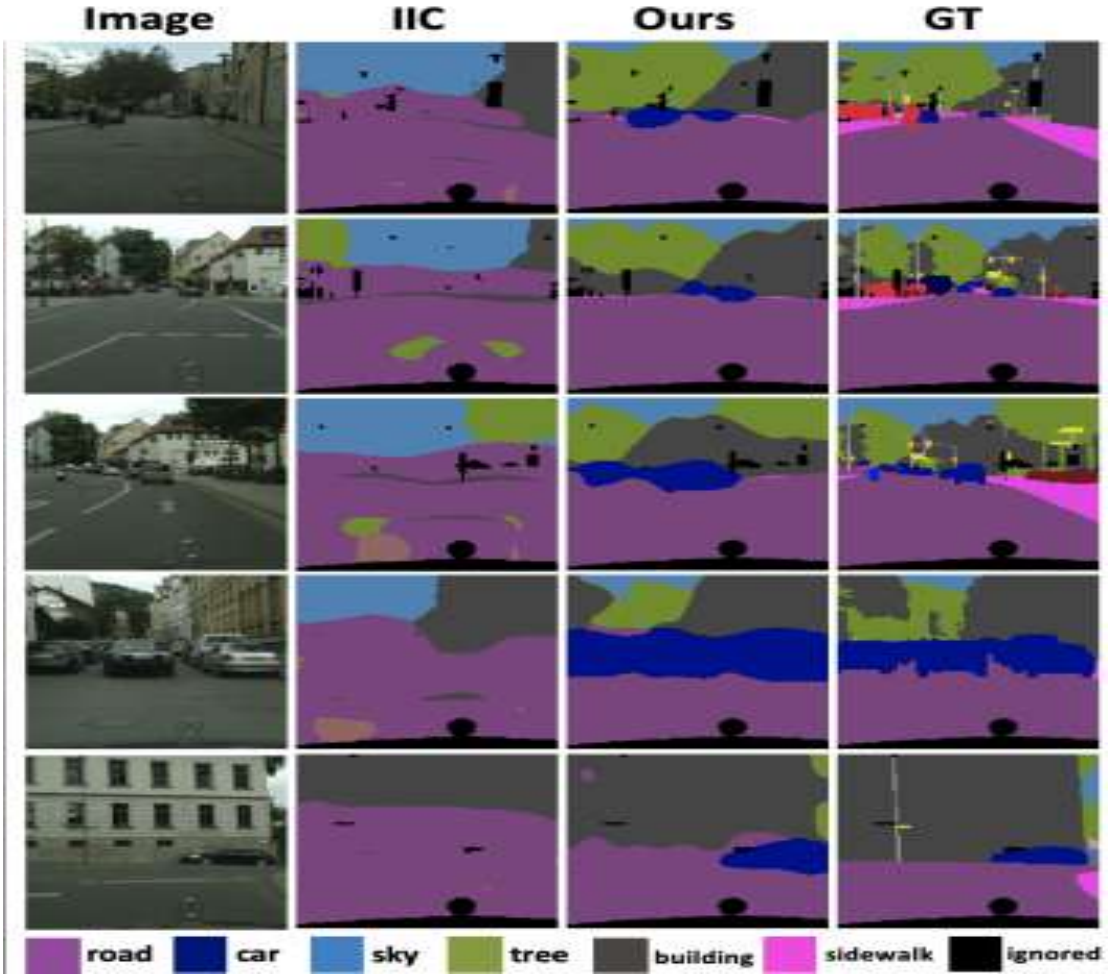


Figura 3.4: Resultados cualitativos generales de segmentación semántica no supervisada con el método propuesto por Cho. J. et. al en [30].

3.3. Inteligencia artifical explicativa en la segmentación.

Dardoulliet. J. en [31] plantea la problemática de comprender las decisiones por parte de los modelos de caja negra, tarea que para los humanos no es sencilla y por lo que su uso puede limitarse en varias situaciones. Ante esta situación, se propone un flujo de trabajo general que permite adaptar métodos de explicabilidad de última generación, especialmente SHAP y RISE a tareas de segmentación de imágenes, permitiendo explicar tanto píxeles como áreas en las imágenes. Demostrando la relevancia de este enfoque en una aplicación crítica como la detección de derrames de petróleo en la superficie del mar.

Aparte de la detección de derrames de petróleo, estos métodos explicativos no se limitan exclusivamente a esa aplicación. En el presente trabajo proponemos un enfoque aplicado a imágenes de frotis sanguíneo como una de las tareas que pueden abordarse utilizando estos métodos de inteligencia artificial explicable.

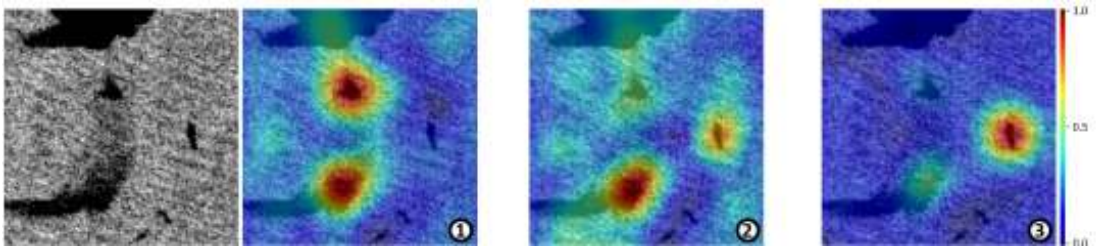


Figura 3.5: Mapas de explicación sobre la decisión del modelo para algunos píxeles de interés (círculos enumerados en rojo) con RISE [31].

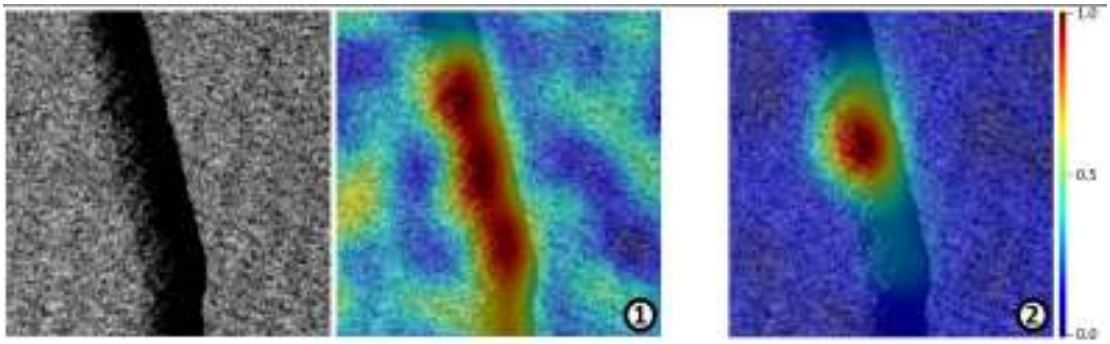


Figura 3.6: Mapas de explicación generados utilizando el método RISE [31] para destacar la influencia del modelo en la toma de decisiones para ciertos píxeles de interés.

Capítulo 4

Metodología

En la metodología, se presenta en detalle el desarrollo del presente trabajo, abarcando desde la obtención de la base de datos utilizada para el entrenamiento de la red neuronal convolucional, hasta la aplicación del método explicativo RISE para obtener los mapas de calor que indican la relevancia de los píxeles en nuestro modelo de caja negra. Este proceso permite comprender y visualizar qué áreas de la imagen son consideradas más importantes por el modelo en sus decisiones, lo cual resulta fundamental para una interpretación clara y confiable de los resultados obtenidos.

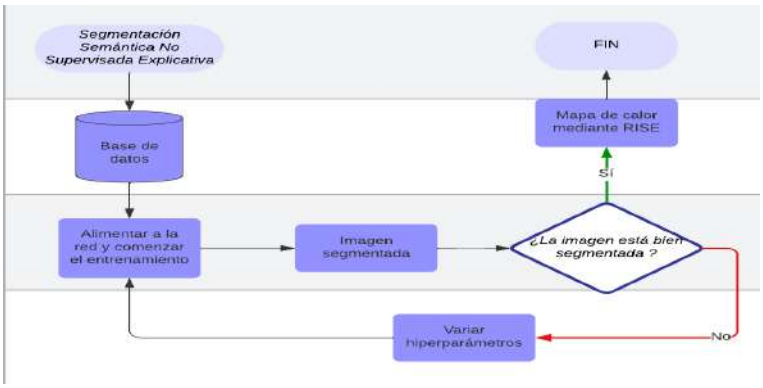


Figura 4.1: Diagrama de flujo de los pasos a seguir para la segmentación semántica no supervisada explicativa.

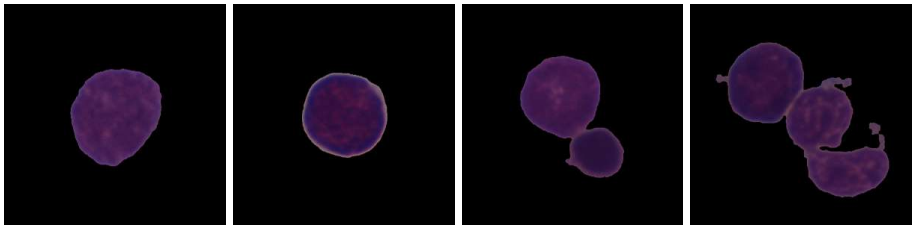
4.1. Base de datos PKG - C-NMC_Leukemia

La base de datos utilizada en este estudio es la PKG - C-NMC_Leukemia [32], la cual se centra en la obtención de células relacionadas con la leucemia linfoblástica aguda. Las imágenes obtenidas se dividen en dos categorías: células cancerosas y células sanas. Las imágenes tienen un tamaño de 420 x 420 píxeles y se dividen en tres conjuntos distintos que se detallan a continuación.

La composición del conjunto de entrenamiento es la siguiente: hay un total de 73 sujetos, de los cuales 47 tienen leucemia linfoblástica aguda (LLA) y 26 son sujetos normales. El conjunto de entrenamiento contiene un total de 10,661 imágenes de células, con 7,272 imágenes de células con LLA y 3,389 imágenes de células normales.

El conjunto de prueba preliminar está compuesto por 28 sujetos, de los cuales 13 tienen LLA y 15 son sujetos normales. Este conjunto contiene 1,867 imágenes de células, con 1,219 imágenes de células con LLA y 648 imágenes de células normales.

Por último, el conjunto de prueba final está compuesto por 17 sujetos, de los cuales 9 tienen LLA y 8 son sujetos normales. Este conjunto contiene 2,586 imágenes de células en total. En la figura 4.2 se muestra un ejemplo del conjunto de datos.



(a) Hem1 (b) Hem2 (c) Hem3 (d) Hem4

Figura 4.2: Ejemplos de imágenes de Leucemia linfoblástica aguda (LLA) proporcionadas por la base de datos.[32]

4.2. Entrenamiento de la red

Las CNN no se utilizan con frecuencia en escenarios completamente no supervisados; sin embargo, tienen un gran potencial para extraer características detalladas de los píxeles de las imágenes, lo cual es necesario para la segmentación de imágenes no supervisada. [33] El presente trabajo en esencia se basa en una CNN con un enfoque de aprendizaje conjunto que predice, para una imagen de entrada arbitraria, etiquetas de clúster desconocidas y aprende los parámetros óptimos para el agrupamiento de píxeles de la imagen.

Se asume que una buena solución de segmentación de imágenes coincide con la solución que proporcionaría un humano. La CNN presentada en el trabajo [12] se basa en tres criterios elementales:

- Los píxeles con características similares deben asignarse la misma etiqueta
- Los píxeles espacialmente continuos deben asignarse a la misma etiqueta

- El número de etiquetas de clúster únicas debe ser grande. (La CNN requiere que se asigne un número mínimo de etiquetas para comenzar el entrenamiento.)

La CNN utilizada en el presente trabajo optimiza conjuntamente las funciones de extracción de características y las funciones de agrupamiento para satisfacer estos criterios y permitir el aprendizaje de extremo a extremo [12]

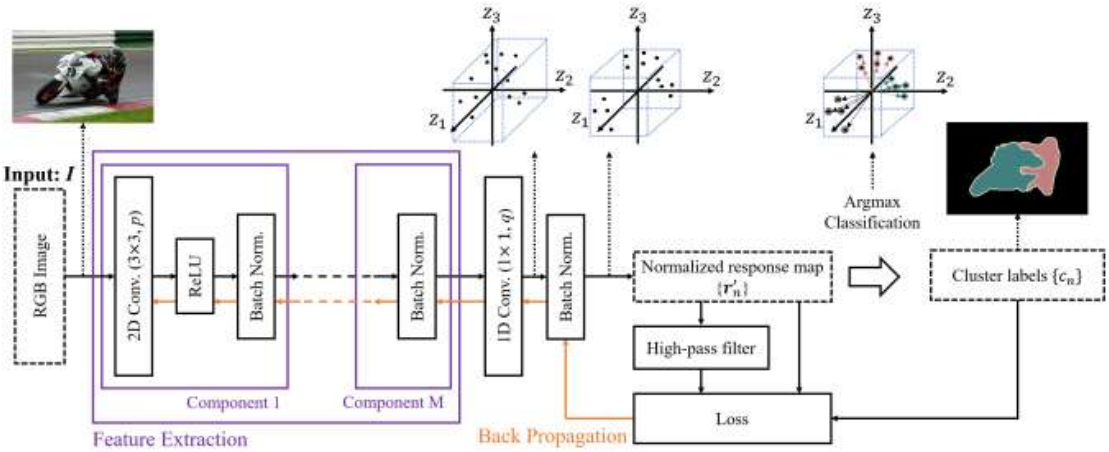


Figura 4.3: Arquitectura de la CNN propuesta para segmentación semántica no supervisda [12].

4.2.1. Funcionamiento

El problema que se resuelve para la segmentación de imágenes se describe de la siguiente manera. Sea $\{\}$ un conjunto de $\{\}_{n=1}^{N}$ donde N representa el número de píxeles en una imagen de color de entrada $X = v_n \in \mathbb{R}^3$.

Sea $f : \mathbb{R}_3 \rightarrow \mathbb{R}_p$ una función de extracción de características y $\{x_n \in \mathbb{R}_p\}$ un conjunto de vectores de características de dimensión p de los píxeles de la imagen. Se asignan etiquetas de clúster $\{c_n \in \mathbb{Z}\}$ a todos los píxeles mediante $c_n = g(x_n)$, donde $g : \mathbb{R}_p \rightarrow \mathbb{Z}$ denota una función de mapeo. [12]

Por lo tanto, g es una función de asignación que devuelve la etiqueta del centroide de clúster más cercano a x_n. En el caso en el que f y g sean fijos, c_n se obtienen utilizando la ecuación mencionada anteriormente. Por el contrario, si f y g son entrenables mientras que c_n se queda fijo, entonces la ecuación mencionada anteriormente se puede considerar como un problema de clasificación supervisada estándar. El presente trabajo de tesis consiste primero en predecir c_n desconocidos mientras se entrenan los parámetros de f y g de manera completamente no supervisada [12].

Se trabaja con un algoritmo basado en CNN para satisfacer los tres criterios elementales mencionados en la sección 4.2. Las funciones de extracción de características para x_n

y c_n se optimizan conjuntamente de una manera que satisface todos los criterios mencionados anteriormente. Para permitir el aprendizaje de extremo a extremo de la CNN, proponiendo un enfoque iterativo para predecir c_n utilizando funciones diferenciables.

Arquitectura de la red

1. **Restricción en la similitud de características:** Se asigna la misma etiqueta a los píxeles que tienen características similares en la segmentación semántica. Para lograr esto, se utiliza un clasificador lineal que agrupa las características de cada píxel en q clases. La entrada es una imagen RGB $X = v_n \in \mathbb{R}^3$, donde cada valor de píxel se normaliza a [0, 1]. Para obtener un mapa de características x_n de p dimensiones a partir de la imagen v_n, se utilizan M componentes convolucionales. Cada componente consiste en una convolución 2D, una función de activación ReLU y una función de normalización de lotes. Aquí, se configuran p filtros de tamaño 3 × 3 para todos los componentes M. Posteriormente, se obtiene un mapa de respuesta $r_n = W_c xn$ al aplicar un clasificador lineal, donde $W_c \in \mathbb{R}^{qxp}$. Luego, se normaliza el mapa de respuesta a $r_n^{'}$ de modo que tenga una media cero y una varianza unitaria.

2. **Restricción en el número de etiquetas de clúster únicas:** En la segmentación de imágenes no supervisada, no se sabe previamente cuántos segmentos se deben generar en una imagen. Por lo tanto, el número de etiquetas de clúster únicas debe adaptarse al contenido de la imagen. El enfoque utilizado implica clasificar los píxeles en un número arbitrario de etiquetas $q^{'}$ $q^{'}(1 \leq q^{'} \leq q)$, siendo q el número máximo de etiquetas. Un valor alto de $q^{'}$ índica una sobresegmentación, mientras que un valor bajo indica una subsegmentación.

   En el proceso de entrenamiento de la red neuronal, se establece un valor grande como el número inicial (máximo) de etiquetas de clúster q. Luego, durante la actualización iterativa, los píxeles que son similares o están cerca espacialmente se fusionan considerando las restricciones de similitud de características. Esto lleva a reducir el número de etiquetas de clúster únicas $q^{'}$, aunque no haya una restricción explícita en q.

4.2.2. Función de pérdida

La función de pérdida propuesta $\mathcal{L}$ consiste en una restricción de similitud de características y una restricción de continuidad espacial, que se denota de la siguiente manera:

$$\mathcal{L} = \underbrace{\mathcal{L}_{sim}(\{r'_n, c_n\}}_{Caractersticas\ similares} + \underbrace{\mu\mathcal{L}_{con}(\{r'_n\}))}_{Continuidad\ espacial} \tag{4.1}$$

donde μ representa el peso para equilibrar las dos restricciones.

Características similares

Las etiquetas de los grupos c_n se obtienen aplicando la función argmax al mapa de respuestas normalizado r'_n. En esta sección la pérdida se calcula de la siguiente manera:

$$\mathcal{L}_{sim}\left\{r'_n, c_n\right\} = \sum_{n=1}^{N}\sum_{i=1}^{q} -\delta(i - c_n)ln(r'_n) \tag{4.2}$$

Donde

$$\delta(t) = \begin{cases} 1 & \text{si t = 0} \\ 0 & \text{De otra forma} \end{cases} \tag{4.3}$$

La función de pérdida tiene como objetivo mejorar la similitud de las características similares en la segmentación. Después de agrupar los píxeles de la imagen según sus características, es deseable que los vectores de características dentro del mismo grupo sean similares entre sí, mientras que los vectores de características de grupos diferentes sean distintos. Al minimizar esta función de pérdida, los pesos de la red neuronal se actualizan para facilitar la extracción de características más efectivas para el agrupamiento.

Restricción en la continuidad espacial

En la segmentación de imágenes, es deseable que los grupos de píxeles estén espacialmente conectados. Esto significa que los píxeles similares que se agrupan en un segmento deben estar adyacentes o cercanos en la imagen. La continuidad espacial es importante porque refleja la estructura y coherencia visual de los objetos presentes en la imagen. Para fomentar esta continuidad espacial, se introduce una restricción adicional que favorece que los píxeles vecinos tengan las mismas etiquetas de grupo. La pérdida de continuidad espacial, representada como $\mathcal{L}_{con}$, se define de la siguiente manera:

$$\mathcal{L}_{con}(\{r'_n\}) = \sum_{\xi=1}^{W-1}\sum_{\eta=1}^{H-1} \|r'_{\xi+1,\eta} - r'_{\xi,\eta}\|_1 + \|r'_{\xi,\eta+1} - r'_{\xi,\eta}\|_1 \tag{4.4}$$

Donde W y H representan el ancho y alto de una imagen de entrada, y $r'_{\xi,\eta}$ representa el valor del píxel en las coordenadas (ξ, η) en el mapa de respuestas r'_n. Al utilizar la pérdida de continuidad espacial $\mathcal{L}_{con}$, se puede reducir el número excesivo de etiquetas causado por patrones o texturas complejas. Esta pérdida ayuda a preservar la coherencia visual y a evitar segmentaciones fragmentadas.

4.2.3. Aprendizaje backpropagation.

En esta sección, se describe el método de entrenamiento utilizado para la segmentación semántica de imágenes mediante aprendizaje no supervisado. Cuando la imagen se in-

troduce en la red, se realiza una predicción de las etiquetas de clúster con los parámetros de la red para segmentación fijos, siguiendo el proceso ilustrado en la sección 4.2 donde se presenta la arquitectura. Luego, se procede con el entrenamiento de los parámetros de la red utilizando las etiquetas predichas mediante el proceso de retroceso basado en el descenso del gradiente. Se calcula y retropropaga la función de pérdida L, como se explica en la sección 4.2.2, para actualizar los filtros de las capas convolucionales. Este proceso se repite *T* veces hasta obtener la predicción final de las etiquetas de clúster.

La figura 4.3 muestra algunas funciones utilizadas en la red, como la función de activación ReLU y las capas convolucionales de 3x3, entre otras. Se destaca la capa de normalización por lotes entre la última capa convolucional y la función de clasificación argmax en la red propuesta. La normalización por lotes juega un papel importante en la obtención de las etiquetas c_n, ya que, al tratarse de un escenario no supervisado, las etiquetas para la predicción no son fijas, y puede haber múltiples etiquetas de salida c_n que logren una función de pérdida cercana a cero. Se configura la tasa de aprendizaje con valores de 0.1 y 0.01. Asimismo, se varía el número de etiquetas a predecir en nuestro modelo de segmentación semántica, siendo 3 el valor mínimo asignado para iniciar el proceso de segmentación.

Figura 4.4: Comparación de los resultados de segmentación no supervisada, donde los distintos segmentos se muestran en colores diferentes. [33]

4.3. Método explicativo RISE

Una vez que se ha entrenado nuestra red de segmentación semántica no supervisada, aplicamos el método explicativo XAI RISE, el cual se explica en la sección 2.6.3 del documento, respaldado por sus fundamentos teóricos. A continuación, se enumeran los pasos necesarios para aplicar el método explicativo:

1. Preparación de los datos:
    - Obtener un conjunto de imágenes de entrada para la segmentación semántica no supervisada.
    - Realizar cualquier preprocesamiento necesario en las imágenes, como redimensionamiento, normalización, etc. Para el presente trabajo se tuvo que realizar un redimensionado de las imágenes a un tamaño de 30 x 30, posteriormente convertirla a un tensor y normalizarla

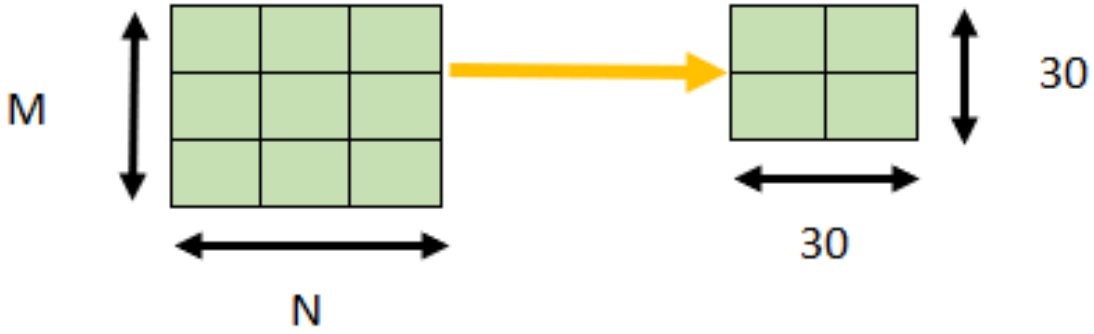
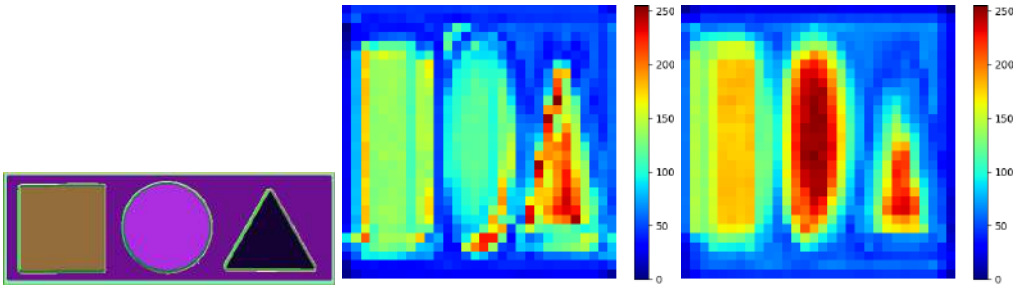


Figura 4.5: Redimensionamiento de las imágenes para la aplicación del método explicativo RISE.

2. Cargar el modelo de segmentación semántica:
    - Utilizar un modelo preentrenado de segmentación semántica o entrenar uno desde cero con el uso de la GPU según las necesidades. Para el presente trabajo no se utiliza un modelo preentrenado.
3. Generación y aplicación de máscaras:
    - Generar un conjunto de máscaras aleatorias binarias que se utilizarán para enmascarar la imagen original.
    - Aplicar las máscaras generadas a la imagen original para obtener imágenes enmascaradas.
    - Pasar las imágenes enmascaradas a través del modelo y obtener sus salidas.

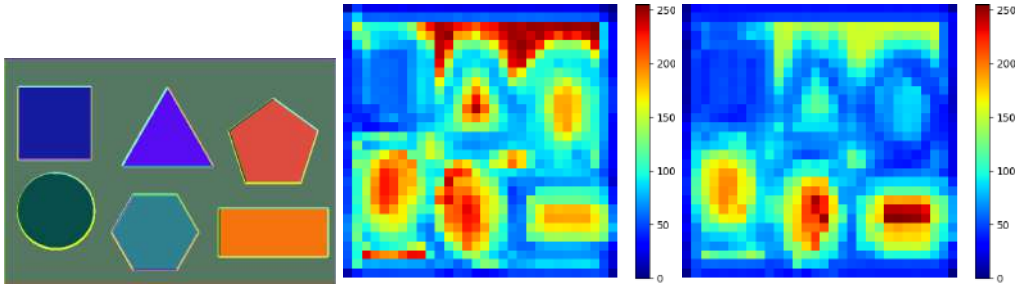
4. Generar el mapa de importancia:
    - Calcular los puntajes de importancia comparando las salidas de las imágenes enmascaradas con la salida de la imagen original para generar los mapas de importancia para la segmentación semántica.
    - Procesar adecuadamente los mapas de importancia para obtener una representación visualmente significativa, como normalizar los valores de importancia, aplicar mapas de colores, etc.
5. Visualizar los resultados:
    - Utilizar bibliotecas de visualización para mostrar los mapas de importancia y los resultados de la segmentación semántica.
    - Se pueden superponer los mapas de importancia en las imágenes originales para resaltar las regiones importantes.

Basándonos en las ecuaciones que rigen el método explicativo RISE [27] y en los fundamentos teóricos que respaldan la red de segmentación semántica no supervisada [12], se realizaron pruebas iniciales para visualizar los píxeles de mayor importancia en la toma de decisiones para la segmentación de imágenes sin anotaciones previas. En primer lugar, se probaron imágenes no relacionadas con trastornos hematológicos, utilizando una imagen que contenía un conjunto de figuras geométricas. Esto se hizo para verificar la validez del método de segmentación semántica no supervisada y del enfoque de explicación RISE.



(a) Imagen segmentada por la red (b) Mapa de calor 1 (c) Mapa de calor 2

Figura 4.6: Mapas de calor generados con el método RISE para una imagen artificial.



(a) Imagen segmentada por la red (b) Mapa de calor 1 (c) Mapa de calor 2

Figura 4.7: Mapas de calor generados con el método RISE para una imagen artificial.

El mapa de calor o heatmap se genera como una imagen en formato RGB. Como se mencionará más adelante en el documento, los píxeles más relevantes se resaltan en un color rojo intenso, lo que significa que su importancia se refleja en el canal rojo de la imagen RGB.

En las Figuras 4.6 y 4.7 se presentan los mapas de calor generados mediante el método RISE, los cuales brindan una visualización de los píxeles más relevantes para la segmentación semántica no supervisada utilizando una red neuronal convolucional. Estos mapas fueron obtenidos a través de diversas pruebas y ajustando los hiperparámetros de la red de segmentación semántica no supervisada, así como del método explicativo RISE. Específicamente, se destaca el mapa de calor (c) en la figura 4.6, el cual fue generado utilizando un rango de aprendizaje reducido, así como el mapa de calor (B) de la figura 4.7 .

4.4. Conclusión del capítulo

En este capítulo, se ha presentado la arquitectura de una red neuronal convolucional para la segmentación semántica no supervisada, tomando como referencia el artículo de Kanezaki et al. [12]. El enfoque del algoritmo se centra en la detección y clasificación de objetos en una imagen sin la necesidad de utilizar información previa o etiquetas.

El proceso de entrenamiento se basa en la optimización iterativa de los parámetros de la red mediante el descenso de gradiente, lo que permite que la red aprenda a agrupar píxeles similares en clústeres coherentes para realizar la segmentación. Esta aproximación sin supervisión es especialmente valiosa cuando no se dispone de un conjunto de datos etiquetado y permite una mayor flexibilidad en la detección de objetos.

Posteriormente, se ha complementado la implementación con el método explicativo de RISE, la cual se utiliza para calcular la importancia de cada píxel en el proceso de

segmentación. RISE utiliza máscaras aleatorias generadas para enmascarar diferentes regiones de la imagen y así evaluar cómo afectan estas máscaras al resultado de la segmentación. Esta herramienta proporciona una valiosa visualización de las áreas más relevantes para la clasificación y ayuda a comprender mejor el proceso de toma de decisiones del modelo.

La combinación de la segmentación no supervisada con RISE añade un componente adicional de interpretabilidad al modelo, permitiendo una mayor transparencia en su funcionamiento y facilitando la identificación de áreas que requieran mejoras o ajustes. Los primeros resultados obtenidos mediante el método explicativo se presentan en las figuras 4.6 y 4.7, donde se resaltan en rojo los perímetros correspondientes a las figuras, lo que nos brinda una idea inicial de cómo funciona el algoritmo explicativo.

En resumen, la arquitectura de la red neuronal convolucional para segmentación semántica no supervisada, combinada con la técnica explicativa de RISE, representa un enfoque prometedor para abordar la segmentación de imágenes sin la necesidad de etiquetas y al mismo tiempo proporcionar una mayor comprensión de las decisiones del modelo. Este enfoque tiene un amplio potencial de aplicación en el campo del procesamiento de imágenes médicas, la clasificación de objetos y otras áreas donde la interpretabilidad del modelo es fundamental para la toma de decisiones precisas y confiables.

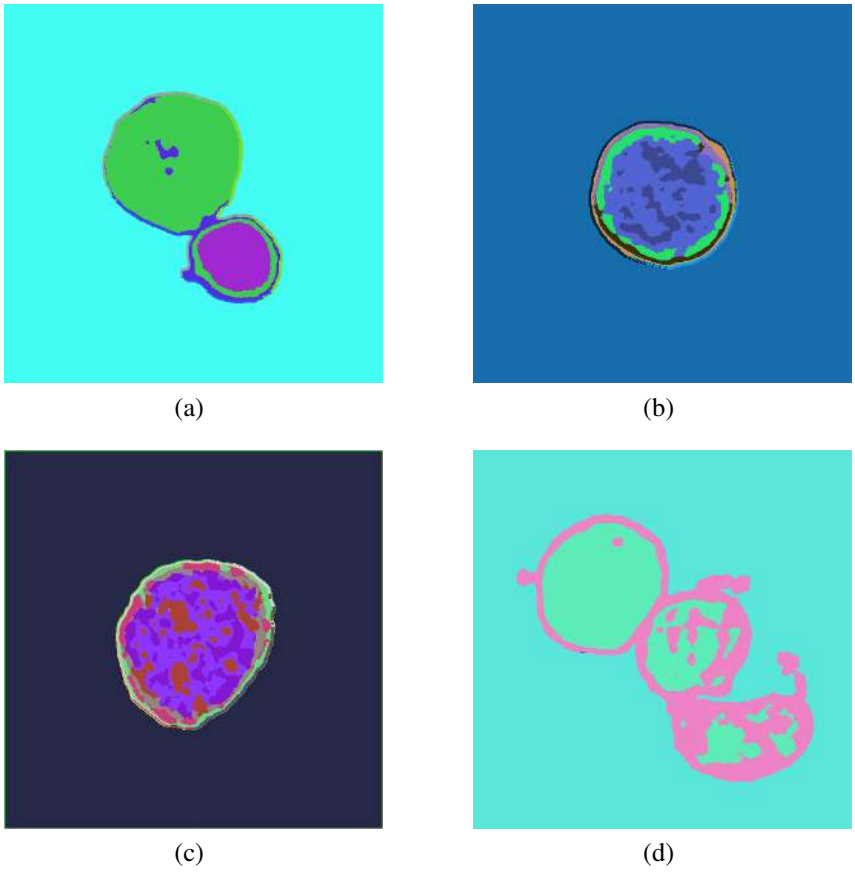
Capítulo 5

Resultados

En este capítulo, se exponen los resultados del proceso de entrenamiento de la red neuronal convolucional para la segmentación semántica no supervisada. En primer lugar, se presentan ejemplos de los mapas de calor obtenidos mediante los métodos LIME y SHAP. A continuación, se describe el entrenamiento adicional realizado utilizando el método RISE, que ha permitido obtener una visión más clara y detallada de los procesos de identificación y segmentación de células hematológicas. Finalmente, se presentan los mapas de calor resultantes que resaltan las regiones de mayor importancia, lo que brinda una comprensión más profunda de los aspectos cruciales para la clasificación precisa de las células de leucemia linfoblástica. Estos resultados representan las principales contribuciones de esta tesis, ya que permiten una interpretación más transparente y confiable de los procesos de segmentación, mejorando la precisión y eficacia del diagnóstico hematológico.

5.1. Resultados del entrenamiento red de segmentación semántica no supervisada

Para llevar a cabo la segmentación semántica de imágenes de células sobre leucemia linfoblástica mediante aprendizaje no supervisado, se emplea una red neuronal convolucional (CNN) que puede ser iniciada con una sola imagen en formato RGB. Sin embargo, es esencial ajustar cuidadosamente los hiperparámetros clave para obtener resultados óptimos. Algunos de estos hiperparámetros cruciales incluyen el rango de aprendizaje, que determina la magnitud de los ajustes de los pesos de la red durante el proceso de entrenamiento; el número de etiquetas que se desean obtener para segmentar las células, lo cual puede variar según la complejidad de las imágenes y la cantidad de clases que se pretenden identificar; y el número de épocas de entrenamiento, que determina cuántas veces la red recorrerá todo el conjunto de datos durante el aprendizaje.



(a) (b) (c) (d)

Figura 5.1: Ejemplos de segmentación semántica no supervisada mediante la CNN empleada en [12].

Los resultados experimentales han demostrado que la segmentación no supervisada mediante CNN puede alcanzar una alta precisión en la identificación y segmentación de células de leucemia linfoblástica en imágenes. La CNN es capaz de aprender patrones y características visuales complejas específicas de este tipo de células, lo que le permite distinguir con gran precisión entre los diferentes tipos de células presentes en las imágenes de leucemia linfoblástica.

Para lograr estos resultados, se realizaron experimentos ajustando cuidadosamente los hiperparámetros de la CNN. La configuración del rango de aprendizaje, el número de etiquetas a obtener y el número de épocas fueron esenciales para optimizar el rendimiento de la segmentación. Al explorar diferentes combinaciones de hiperparámetros y realizar pruebas a prueba y error, se encontraron los valores óptimos que permitieron a la CNN obtener un alto grado de precisión en la tarea de identificación y segmentación de células de leucemia linfoblástica. La capacidad de adaptar la CNN a estas características específicas de las imágenes de leucemia linfoblástica resalta su versatilidad y potencial para abordar problemas de segmentación en el campo de la hematología de manera precisa y eficiente.

5.2. Mapas de calor LIME y SHAP

SHAP son aplicados en primera instancia para una red de clasificación. En el capítulo dos de este trabajo, se presentan los fundamentos teóricos de los métodos explicativos LIME y SHAP, así como su aplicación en modelos de caja negra para redes neuronales convolucionales. Además, se proporcionarán fundamentos que respalden el uso del método explicativo RISE, en comparación con los modelos mencionados en esta sección. A continuación, se presentan ejemplos de cómo se aplican inicialmente los métodos LIME y SHAP en una red de clasificación. Estos ejemplos sirven para ilustrar el proceso y la interpretación proporcionada por dichos métodos.

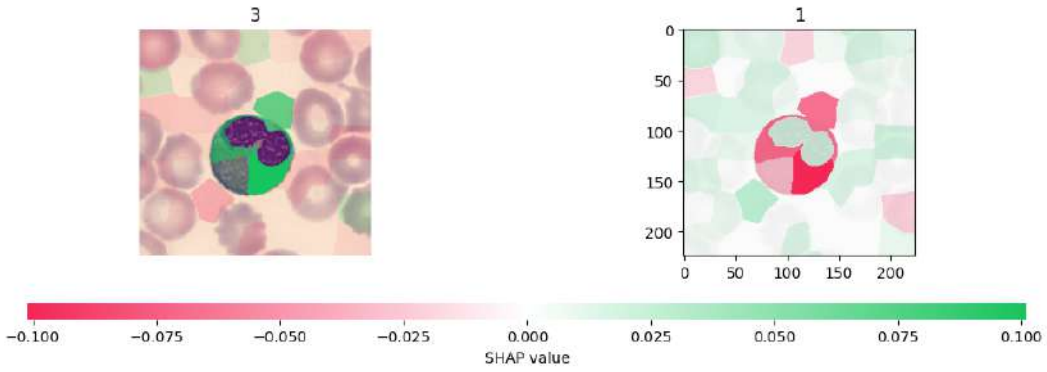
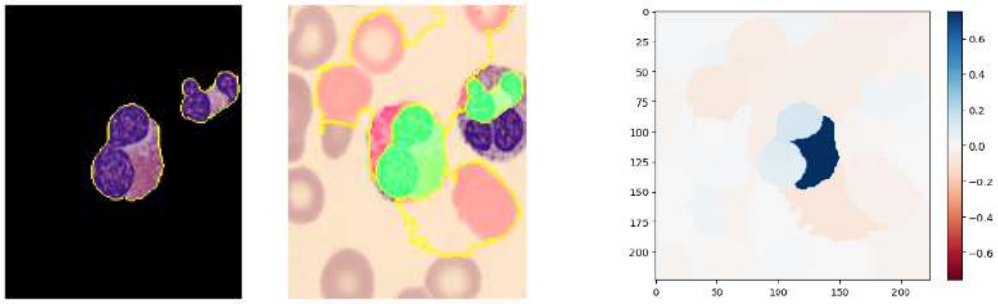


Figura 5.2: Mapa de calor del método explicativo SHAP. [26]

El mapa de calor presentado en la figura 5.2 resalta de manera visual cómo cada característica contribuye al resultado final de la predicción. Estos mapas de calor generados por SHAP proporcionan una herramienta eficaz para visualizar y comprender con mayor claridad qué partes específicas de una imagen o conjunto de características tienen una mayor influencia en las decisiones tomadas por el modelo. Las contribuciones positivas se destacan en color verde, mientras que las contribuciones negativas se representan en rojo.

El mapa de calor presentado en la figura 5.3 ofrece una representación gráfica del cálculo de los pesos de los píxeles, que señalan su contribución a las decisiones del modelo. Esta visualización contribuye significativamente a aumentar la confianza en los resultados proporcionados por el modelo, al brindar una comprensión más clara de qué áreas de la imagen están influyendo en las decisiones tomadas.



(a) Imagen con las región de interés asilada y su máscara

(b) Mapa de calor.

Figura 5.3: Segmentación de la región de interés y mapa de calor generado con el método explicativo LIME. [25]

5.3. Mapas de calor RISE

La efectividad del método RISE se basa en el principio de que al generar una gran cantidad de máscaras aleatorias y promediar sus predicciones acumulativas, se puede obtener una representación más robusta y confiable de los píxeles importantes para las predicciones del modelo. Aunque la elección de la máscara es aleatoria, la agregación de varias máscaras le permite reducir el sesgo y capturar patrones significativos en la imagen. Cabe señalar que la confiabilidad del método RISE no solo se basa en la selección aleatoria de máscaras, sino que también depende de la arquitectura y el rendimiento del modelo de segmentación semántica. Es más probable que el método RISE genere mapas de calor útiles y confiables si el modelo captura adecuadamente los patrones y características relevantes en los datos de entrada.

Si bien no existe una combinación general de hiperparámetros del método RISE que funcione de manera universal para las redes de segmentación semántica no supervisada, es importante considerar los siguientes aspectos:

- **Tamaño de las máscaras:** El tamaño de las máscaras influye en cómo se abordan las características y las relaciones espaciales en la imagen. Máscaras más grandes podrían capturar características más amplias, mientras que máscaras más pequeñas podrían enfocarse en detalles finos. Sin embargo, usar máscaras muy grandes puede perder detalles importantes y hacer que la red no se concentre en regiones específicas.

- **Probabilidad de activación:** La probabilidad de activación controla la densidad de píxeles activados en cada máscara. Valores más altos de probabilidad de activación (como 0.9) significan que hay una mayor probabilidad de que se active cada píxel en la máscara, lo que podría resultar en una mayor atención a características específicas en la imagen. Valores más bajos (como 0.5) generan máscaras

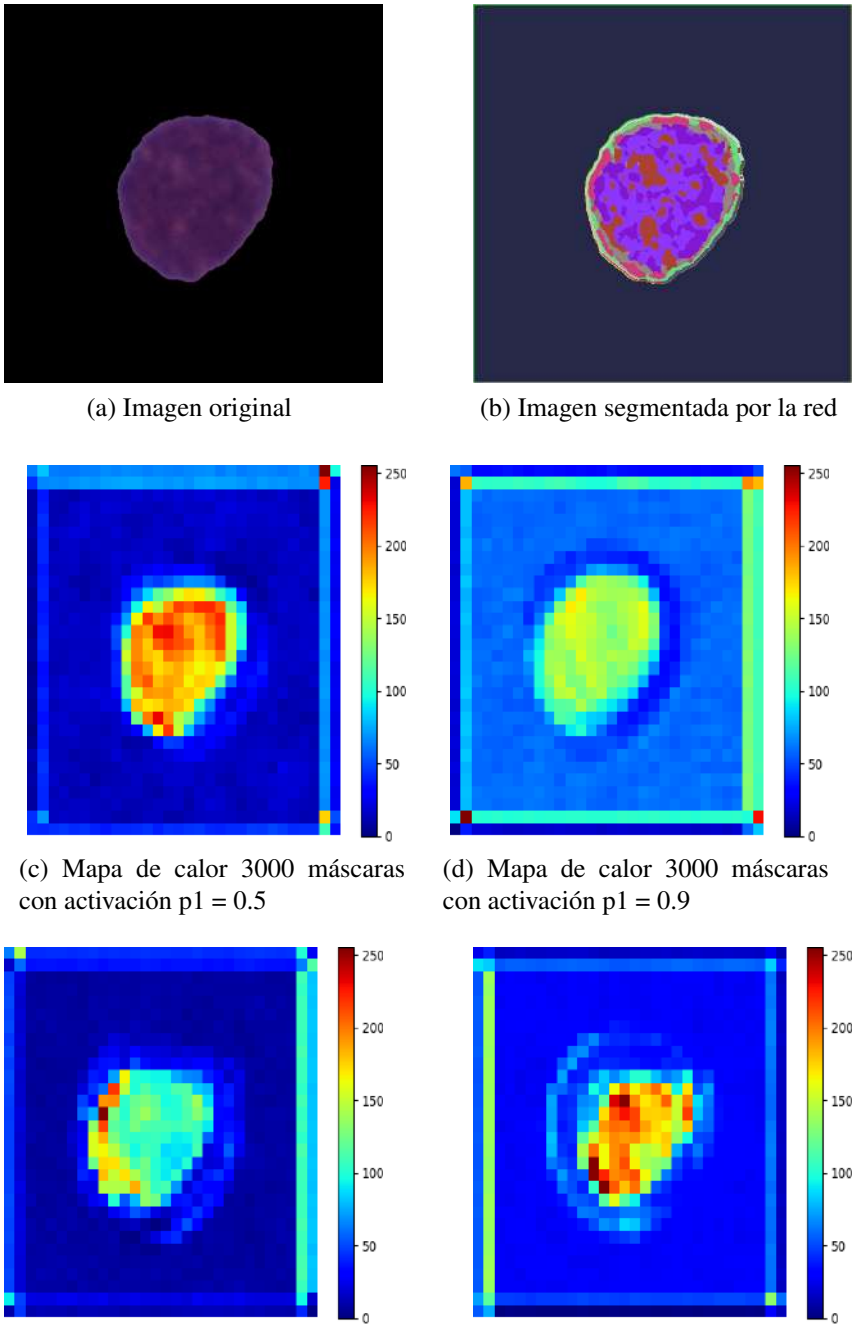
más dispersas, lo que podría permitir una exploración más amplia de las características.

- **Número de máscaras:** El número de máscaras determina cuántas iteraciones del proceso se realizarán para calcular los mapas de calor. Más máscaras permiten una exploración más exhaustiva de diferentes regiones y variaciones de la imagen, pero también puede aumentar el tiempo de cálculo.

Para la CNN enfocada en la tarea de segmentación semántica no supervisadas [12], elegimos empíricamente diferentes números de máscaras, en particular, para el conjunto de imágenes mostradas en el presente trabajo usamos 3000 y 5000 máscaras, empleando un tamaño de imagen de 30 x 30 con un tamaño de máscara de 10 y probabilidades de activación p_1 = 0.5 y p_1 = 0.9 en los canales RGB. (Si no se puede obtener el mapa de calor deseado, se deben modificar los valores de tamaño de máscara y probabilidad de activación). Para este trabajo, utilizamos un conjunto de datos obtenido de [32], donde no se proporcionan posiciones de referencia reales ni máscaras de segmentación, por lo que no se puede cuantificar el rendimiento de la interpretabilidad. Se generaron mapas de calor aumentando el número de máscaras a 7000 los cuales se muestran en el Apéndice B

Cada mapa de índices de relevancia obtenido de la red neuronal se representa como una matriz de valores en el rango de 0 a 255. Estos valores se visualizan utilizando una escala de colores que va desde el azul hasta el rojo, pasando por tonos como el amarillo y el naranja. Al analizar una imagen, cada píxel se le asigna un color en función de su relevancia, como se muestra en la Figura 5.1 (utilizando el mapa de color 'jet'). Los píxeles menos importantes se representan en tonos de azul, mientras que los más relevantes se representan en tonos de rojo. De esta manera, se puede identificar de forma visual los datos más significativos para la red neuronal.

Figura 5.4: Mapa de colores jet. [34]



(a) Imagen original

(b) Imagen segmentada por la red

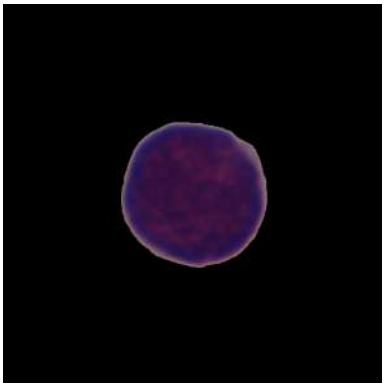
(c) Mapa de calor 3000 máscaras con activación p1 = 0.5

(d) Mapa de calor 3000 máscaras con activación p1 = 0.9

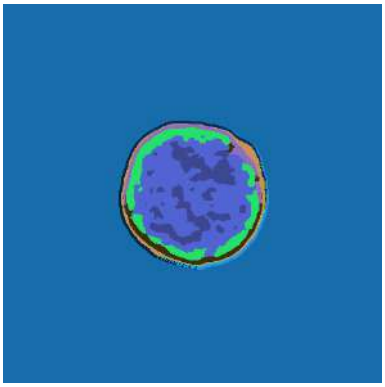
(e) Mapa de calor 5000 máscaras con activación p1 = 0.5

(f) Mapa de calor 5000 máscaras con activación p1 = 0.9

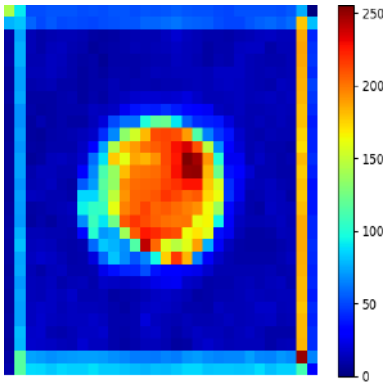
Figura 5.5: Mapas de calor generados con el método explicativo RISE para una célula enferma del dataset PKG - C-NMC_Leukemia [32].



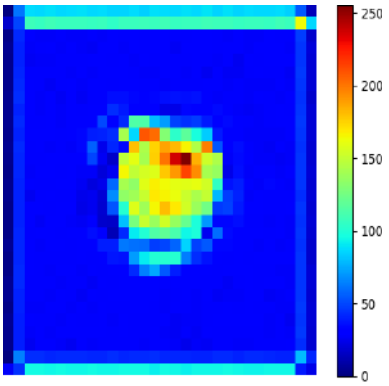
(a) Imagen original



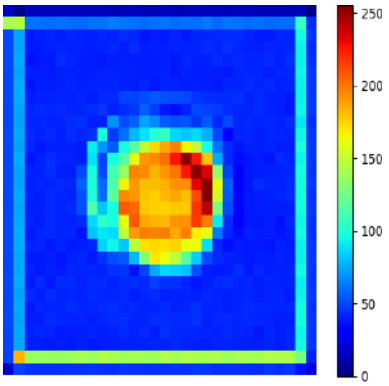
(b) Imagen segmentada por la red



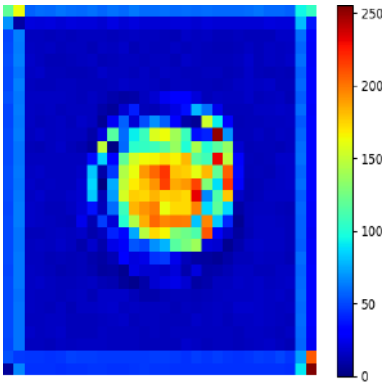
(c) Mapa de calor 3000 máscaras con activación p1 = 0.5



(d) Mapa de calor 3000 máscaras con activación p1 = 0.9

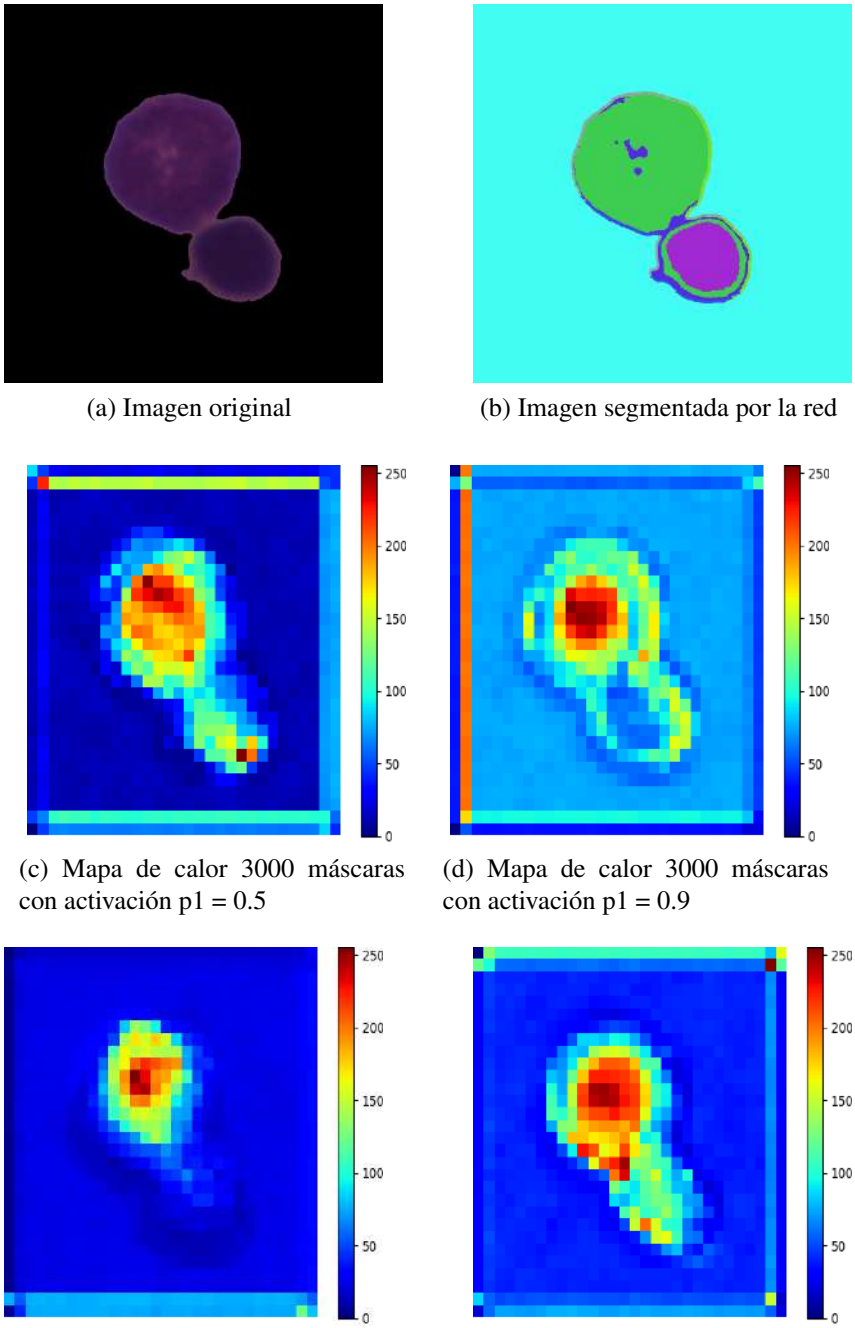


(e) Mapa de calor 5000 máscaras con activación p1 = 0.5



(f) Mapa de calor 5000 máscaras con activación p1 = 0.9

Figura 5.6: Mapas de calor generados con el método explicativo RISE para una célula enferma del dataset PKG - C-NMC_Leukemia [32].



(a) Imagen original

(b) Imagen segmentada por la red

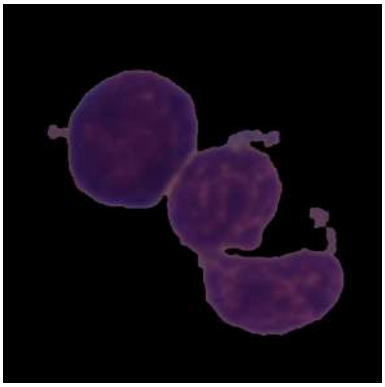
(c) Mapa de calor 3000 máscaras con activación p1 = 0.5

(d) Mapa de calor 3000 máscaras con activación p1 = 0.9

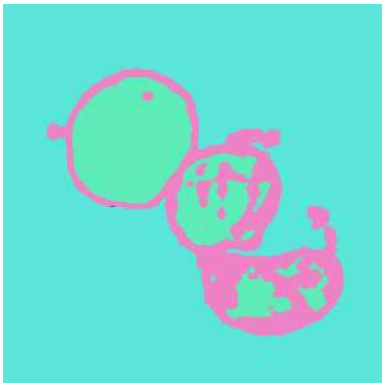
(e) Mapa de calor 5000 máscaras con activación p1 = 0.5

(f) Mapa de calor 5000 máscaras con activación p1 = 0.9

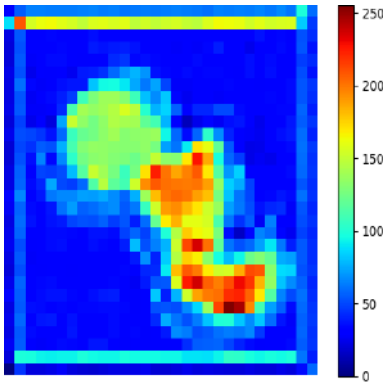
Figura 5.7: Mapas de calor generados con el método explicativo RISE para una célula enferma del dataset PKG - C-NMC_Leukemia [32].



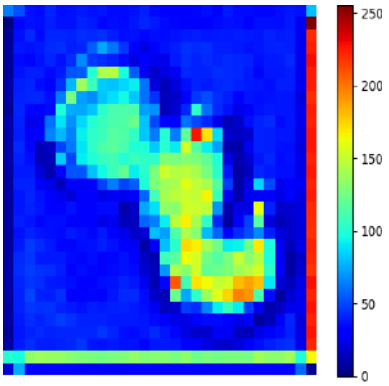
(a) Imagen original



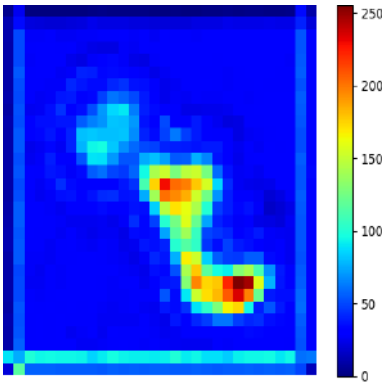
(b) Imagen segmentada por la red



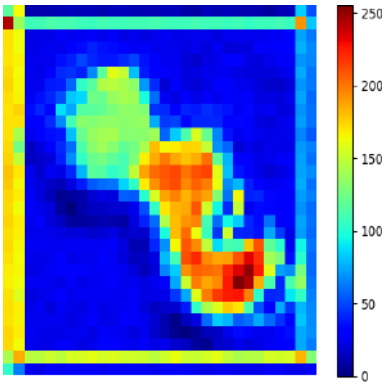
(c) Mapa de calor 3000 máscaras con activación p1 = 0.5



(d) Mapa de calor 3000 máscaras con activación p1 = 0.9



(e) Mapa de calor 5000 máscaras con activación p1 = 0.5



(f) Mapa de calor 5000 máscaras con activación p1 = 0.9

Figura 5.8: Mapas de calor generados con el método explicativo RISE para una célula enferma del dataset PKG - C-NMC_Leukemia [32].

5.4. Discusión de resultados

En las Figuras 5.2 y 5.3 se presentan los mapas de calor generados utilizando los métodos explicativos LIME y SHAP. Estos mapas de calor fueron aplicados a un clasificador de imágenes sanguíneas y se incluyen en esta sección para brindar un apoyo visual que facilite la comprensión de su generación. Estos mapas de calor se fundamentan en los conceptos matemáticos presentados en el marco teórico del presente trabajo, lo que contribuye a su validez y robustez.

A diferencia de otros métodos explicativos estudiados en el capítulo dos de este trabajo, como LIME [25] y SHAP [26], RISE [27] ofrece varias ventajas distintivas en el contexto de segmentación semántica no supervisada. RISE se enfoca en localizar áreas importantes en la imagen, lo que resulta esencial para lograr una segmentación precisa de estructuras y elementos en imágenes de células sanguíneas. Al generar múltiples máscaras aleatorias y promediar los resultados, RISE proporciona explicaciones estables y consistentes, lo que contribuye a una mayor confianza en las predicciones del modelo. Otra de las características destacadas de RISE es su interpretabilidad a nivel de píxel. Al proporcionar mapas de calor que resaltan las regiones de la imagen que son más relevantes para la predicción del modelo, RISE permite una interpretación más precisa y detallada a nivel de píxel. Esta información es especialmente beneficiosa para nuestra tarea de segmentación semántica, ya que proporciona una explicación clara de las áreas que influyen en las decisiones del modelo y ayuda a comprender mejor el proceso de segmentación. Finalmente, una ventaja adicional de RISE es que no requiere acceso a la arquitectura original de la red, lo que simplifica su implementación y permite una mayor flexibilidad en el uso de diferentes modelos de segmentación. Esto facilita la incorporación de RISE en nuestro enfoque de segmentación semántica no supervisada sin necesidad de modificar la estructura de la red original.

Debido al extenso número de iteraciones requeridas para generar las máscaras aleatorias y calcular las contribuciones de cada región de la imagen, el método RISE consume una cantidad considerable de recursos computacionales. Esta alta demanda de recursos puede generar un tiempo de procesamiento prolongado y requerir el uso de hardware potente para obtener resultados eficientes. Los mapas de calor se generaron usando un equipo ASUS-TUF GAMING F15 con un procesador 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz 2.69 GHz, RAM 16.0 GB (15.7 GB usable), SSD 500 GB y GPU Nvidia RTX 3050.

Además, la naturaleza aleatoria de las máscaras utilizadas en RISE puede dar lugar a mapas de calor con cierto nivel de ruido. Al generarse las máscaras de manera aleatoria, algunas de ellas pueden no capturar adecuadamente las características relevantes de la imagen o incluso incluir áreas que no son verdaderamente informativas para el modelo. Como consecuencia, los mapas de calor resultantes pueden presentar patrones no deseados o ruido, lo que podría afectar la precisión de la segmentación.

Otro aspecto a considerar es que RISE calcula los puntajes de importancia al promediar los resultados de múltiples máscaras, lo que puede suavizar los patrones de importancia y reducir la nitidez del mapa de calor. Esta aleatoriedad en la selección de las máscaras introduce cierta incertidumbre en los resultados obtenidos. Es importante destacar que, a pesar de la presencia de ruido en los mapas de calor generados por RISE, estos resultados no necesariamente invalidan las conclusiones obtenidas. En el contexto de esta tesis, es posible realizar ajustes en los parámetros de RISE, como el tamaño de las máscaras o la cantidad de máscaras generadas. También se podría considerar el uso de múltiples GPU para acelerar el cálculo y mejorar la calidad de los mapas de calor.

5.5. Conclusión del capítulo

Se han explorado y aplicado métodos explicativos, como LIME y SHAP, para comprender el funcionamiento de un clasificador de imágenes sanguíneas. Estos métodos proporcionan una visualización detallada de las regiones más relevantes en las imágenes que influyen en las decisiones del modelo, lo que contribuye a su validez y robustez. Sin embargo, se ha demostrado que el método explicativo RISE ofrece ventajas distintivas en el contexto de la segmentación semántica no supervisada de células sanguíneas. RISE se destaca por su capacidad para localizar áreas importantes en las imágenes, lo cual es esencial para lograr una segmentación precisa de estructuras y elementos en células sanguíneas. Además, al generar múltiples máscaras aleatorias y promediar los resultados, RISE proporciona explicaciones estables y consistentes, lo que aumenta la confianza en las predicciones del modelo. La interpretabilidad a nivel de píxel que ofrece RISE es especialmente valiosa para entender el proceso de segmentación, ya que resalta las regiones de la imagen que más influyen en las decisiones del modelo.

No obstante, es esencial tener en cuenta que la aplicación de RISE también plantea ciertos desafíos. La notable demanda de recursos computacionales, debido al considerable número de iteraciones requeridas para generar las máscaras aleatorias, puede resultar en un tiempo de procesamiento prolongado. Por ejemplo, la generación de los mapas de calor presentados en este trabajo conlleva alrededor de una hora para lograr una visualización significativa de las áreas de interés. Además, este proceso requiere hardware potente para lograr eficiencia en los resultados obtenidos. Otro aspecto a considerar es la inherente naturaleza aleatoria de las máscaras generadas por RISE, lo que puede introducir un nivel de ruido en los mapas de calor. Aunque este ruido no influye en la precisión de la segmentación, es importante destacar que, a pesar de estos desafíos, RISE mantiene su valía como herramienta fundamental para obtener información relevante y explicativa en la segmentación de imágenes de células sanguíneas. Mediante la realización de ajustes adecuados en los parámetros y la consideración de la utilización de múltiples unidades de procesamiento gráfico (GPU), se puede mejorar la calidad de

los mapas de calor y reducir el ruido inherente a la metodología de RISE. Esto, a su vez, ampliaría la confiabilidad y utilidad de este enfoque en el análisis de imágenes de células sanguíneas, brindando una comprensión más sólida y precisa de los procesos subyacentes.

Finalmente, con el propósito de permitir futuras comparaciones y el desarrollo potencial de técnicas de Inteligencia Artificial Explicable (IAE), el código de la CNN con el método RISE adaptado se encuentra disponible en el siguiente repositorio de GitHub: `https://github.com/gomarjo/Unsupervised-RISE`. Esta iniciativa no solo fomenta la transparencia en la investigación, sino también el avance colectivo en la interpretación y el análisis de imágenes médicas en el contexto de la segmentación semántica no supervisada.

Capítulo 6

Conclusiones y Trabajo futuro

6.1. Conclusiones

Durante el proceso de entrenamiento de las redes neuronales en este estudio, se optó por no utilizar redes neuronales preentrenadas. La razón detrás de esta decisión fue permitir la generación de mapas de calor utilizando el método RISE, que nos permite visualizar el aprendizaje realizado por la red neuronal de forma autónoma. Los primeros mapas de calor obtenidos en este proceso, como se muestra en el capítulo 4, corresponden a imágenes que no forman parte del conjunto de datos utilizado. Estos mapas resaltan los píxeles más relevantes, especialmente los contornos de las figuras geométricas como círculos, triángulos y cuadrados, mientras dejan de lado el fondo de la imagen.

Después de iniciar el entrenamiento de la red neuronal convolucional para la tarea de segmentación semántica no supervisada en imágenes de leucemia linfoblática aguda en combinación con el método explicativo RISE, se requirió realizar preprocesamiento en las imágenes. En este caso, se redujo el tamaño original a 30 x 30 debido a limitaciones de recursos computacionales. Al realizar la reducción, se pudo ampliar el número máximo de máscaras aleatorias generadas pasando de un máximo de 500 en el tamaño original a 9000 en el tamaño propuesto, lo cual nos brindó la oportunidad de obtener diferentes mapas de calor con un mayor número de máscaras, el número de máscaras que generan mapas de calor con regiones de interés son 3000, 5000 y 7000, tomando en cuenta que en ocasiones fue necesaria la variación de los hiperpárametros que rigen al método RISE. A medida que aumentaba el número de máscaras, se resaltaban regiones de mayor importancia. Sin embargo, un aspecto a tener en cuenta fue que, al trabajar con imágenes pequeñas, se presentó un nivel constante de ruido en los mapas de calor Para eliminar este ruido y mejorar los resultados, sería necesario aumentar los recursos computacionales disponibles o realizar otro tipo de preprocesamiento de las imágenes.

En última instancia, los resultados y hallazgos obtenidos en este trabajo proporcionan una base sólida para futuras investigaciones en el campo de la segmentación semán-

tica no supervisada en imágenes de células sanguíneas. La combinación de métodos explicativos como RISE con redes neuronales convolucionales representa una potente herramienta para el análisis y diagnóstico de enfermedades hematológicas, con el potencial de mejorar la precisión y eficacia del diagnóstico clínico. Con esfuerzos continuos para mejorar la calidad de los mapas de calor y reducir el consumo de recursos, el método RISE y enfoques similares prometen seguir siendo una valiosa contribución en el campo de la visión por computadora aplicada a la hematología

6.2. Contribuciones

Con base en los resultados obtenidos en este trabajo, se derivan las siguientes contribuciones clave que enriquecen el ámbito de la visión por computadora, la inteligencia artificial explicativa en combinación con el aprendizaje no supervisado:

- Entrenamiento una red neuronal convolucional para la tarea de segmentación semántica no supervisada enfocada a imágenes sanguíneas.
- Investigación sobre métodos de inteligencia artificial explicable agnósticos enfocados a modelos de caja negra, es decir, que no se requiere modificar la arquitectura original de la red.
- Generación de mapas de calor utilizando el método RISE para tareas de segmentación semántica no supervisada.
- Los resultados de esta tesis pueden ser publicados como un nuevo enfoque en el campo de la visión por computadora en aprendizaje no supervisado.

6.3. Trabajo futuro

Los resultados obtenidos en este trabajo son prometedores en el campo de la segmentación no supervisada explicativa. Se ha identificado una brecha en la investigación, ya que existe poca información sobre el uso de la Inteligencia Artificial Explicable en este campo. Se propone explorar el entrenamiento de diferentes arquitecturas enfocadas en la segmentación no supervisada y adaptar métodos explicativos como LIME y SHAP. Además, se busca mejorar la eficiencia del método RISE mediante la variación de parámetros como el número de máscaras, tamaño y probabilidad de activación del píxel. Para lograr esto, se considera aumentar los recursos computacionales disponibles y mejorar el tamaño de las imágenes para evitar el ruido en los mapas de calor, lo cual permitiría obtener resultados más confiables de las redes neuronales.

Asimismo, se ha observado que el uso de redes neuronales con bajos bits puede ser eficiente [39]. Estas técnicas buscan reducir la precisión de los números utilizados en

las operaciones de la red sin comprometer su rendimiento. Esto conlleva ventajas como acelerar los cálculos y mejorar la eficiencia energética, lo cual es especialmente relevante en aplicaciones en tiempo real o con restricciones de energía. Sin embargo, es importante tener en cuenta que reducir la precisión también implica una pérdida de información relevante. Para mitigar este efecto negativo, existen técnicas como la cuantización, la normalización dinámica y el uso de esquemas de entrenamiento especializados. No obstante, estas técnicas no se alinean con los objetivos y metas propuestas en este trabajo de investigación.

Bibliografía

[1] Mathworks (2023) Procesamiento de imágenes y visión artificial. Segmentación semántica. Recuperado de: https://la.mathworks.com/solutions/image-video-processing/semantic-segmentation.html

[2] Tran. T., Kwon, O. H., Kwon, K. R., Lee, S. H., & Kang, K. W. (2018, December). Blood cell images segmentation using deep learning semantic segmentation. In 2018 IEEE international conference on electronics and communication engineering (ICECE) (pp. 13-16). IEEE.

[3] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

[4] Gramegna A, Giudici P. SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. Front Artif Intell. 2021 Sep 17;4:752558. doi: 10.3389/frai.2021.752558. PMID: 34604738; PMCID: PMC8484963.

[5] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Transactions on Medical Imaging, 36(7):1550–1560, 2017.

[6] Peter Naylor, Marick Lae, Fabien Reyal, and Thomas Walter. ´ Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Transactions on Medical Imaging, 2018.

[7] Hao Chen, Xiaojuan Qi, Lequan Yu, Qi Dou, Jing Qin, and Pheng-Ann Heng. Dcan: Deep contour-aware networks for object instance segmentation from histology images. Medical Image Analysis, 36:135–146, 2017.

[8] Emadi. A & Law JY. (2022). MSD Manual, Professional version. Overview of leukemia. Recuperado de: https://www.msdmanuals.com/professional/hematology-and-oncology/leukemias/overview-of-leukemia

[9] Daniel A. Arber, Attilio Orazi, Robert Hasserjian, Jürgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola, James W. Vardiman; The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016; 127 (20): 2391–2405. doi: https://doi.org/10.1182/blood-2016-03-643544

[10] American Cancer Society (2023): Cancer Facts and Statistics. Recuperado de: https://www.cancer.org/research/cancer-facts-statistics.html

[11] Jiang, Z., He, Y., Ye, S., Shao, P., Zhu, X., Xu, Y., ... & Yang, G. (2023). O2M-UDA: Unsupervised dynamic domain adaptation for one-to-multiple medical image segmentation. Knowledge-Based Systems, 265, 110378.

[12] Kim, W., Kanezaki, A., & Tanaka, M. (2020). Unsupervised learning of image segmentation based on differentiable feature clustering. IEEE Transactions on Image Processing, 29, 8055-8068.

[13] Kanezaki, A. (2018, April). Unsupervised image segmentation by backpropagation. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 1543-1547). IEEE.

[14] Xia, X.,& Kulis, B. (2017). W-net: A deep model for fully unsupervised image segmentation. arXiv preprint arXiv:1711.08506.

[15] Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., ... & Gao, M. (2023). Techniques and Challenges of Image Segmentation: A Review. Electronics, 12(5), 1199.

[16] Rother, C., Kolmogorov, V., & Blake, A. (2004). "GrabCutïnteractive foreground extraction using iterated graph cuts. ACM transactions on graphics (TOG), 23(3), 309-314.

[17] Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient nd image segmentation. International journal of computer vision, 70(2).

[18] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePrisot, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J.: A guide to deep learning in healthcare. Nature Medicine 25(1), pp. 24-29 (2019).

[19] Valliani, A. A., ranti, D., Oermann, E. K.: Deep Learning in Neurology: A Systematic Review. Neurology and Therapy 8(2), pp. 351-365 (2019).

[20] Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H-J., Kim, N.: Deep Learning in Medical Imaging. Neurospine16(4), pp. 657-668 (2019).

[21] Meske, C., & Bunde, E. (2020). Transparency and trust in human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support. In Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22 (pp. 54-69). Springer International Publishing.

[22] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,ïn IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[23] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[24] Silburt, Ari & Ali-Dib, Mohamad & Zhu, Chenchong & Jackson, Alan & Valencia, Diana & Kissin, Yevgeni & Tamayo, Daniel & Menou, Kristen. (2018). Lunar Crater Identification via Deep Learning. Icarus. 317. 10.1016/j.icarus.2018.06.022.

[25] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). Why should i trust you?. Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[26] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[27] Petsiuk, V., Das, A., & Saenko, K. (2018). Rise: Randomized input sampling for explanation of black-box models.arXiv preprint arXiv:1806.07421.

[28] Alharbi, A. H., Aravinda, C. V., Lin, M., Venugopala, P. S., Reddicherla, P., & Shah, M. A. (2022). Segmentation and Classification of White Blood Cells Using the UNet. Contrast media & molecular imaging, 2022, 5913905. https://doi.org/10.1155/2022/5913905

[29] Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., & Freeman, W. T. (2022). Unsupervised semantic segmentation by distilling feature correspondences. arXiv preprint arXiv:2203.08414.

[30] Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). Picie: Unsupervised semantic segmentation using invariance and equivariance in clustering. In Proceedings

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16794-16804).

[31] Dardouillet, P., Benoit, A., Amri, E., Bolon, P., Dubucq, D., & Crédoz, A. (2022, August). Explainability of Image Semantic Segmentation Through SHAP Values. In ICPR-XAIE.

[32] Gupta, A., & Gupta, R. (2019). ALL Challenge dataset of ISBI 2019 [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.dc64i46r

[33] A. Kanezaki, Unsupervised image segmentation by backpropagation, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE, 2018.

[34] Velázquez Arreola, J. J., (2019), Identificación de peatones en imágenes aéreas con redes neuronales explicativas y fusión de sensores, Tesis de Maestría, Instituto Nacional de Astrofísica, Óptica y Electrónica.

[35] Hasany, S. N., Petitjean, C., & Mériaudeau, F. (2023). Seg-XRes-CAM: Explaining Spatially Local Regions in Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3732-3737).

[36] Fel, T., Ducoffe, M., Vigouroux, D., Cadène, R., Capelle, M., Nicodème, C., & Serre, T. (2023). Don't Lie to Me! Robust and Efficient Explainability with Verified Perturbation Analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16153-16163).

[37] Oh, C., Shim, D., & Kim, H. J. (2023). AURA: Automatic Mask Generator using Randomized Input Sampling for Object Removal. arXiv preprint arXiv:2305.07857.

[38] Montavon, G., Kauffmann, J., Samek, W., Müller, KR. (2022). Explaining the Predictions of Unsupervised Learning Models. In: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, KR., Samek, W. (eds) xxAI - Beyond Explainable AI. xxAI 2020. Lecture Notes in Computer Science(), vol 13200. Springer, Cham. https://doi.org/10.1007/978-3-031-04083-2_7

[39] Becking, D., Dreyer, M., Samek, W., Müller, K., & Lapuschkin, S. (2020, July). Ecq x: explainability-driven quantization for low-bit and sparse DNNs. In International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers (pp. 271-296). Cham: Springer International Publishing.

[40] Velázquez-Arreola, J., Zarraga-Vargas, O.A., Díaz-Hernández, R., Altamirano-Robles, L. (2023). Evaluation of Heatmaps as an Explicative Method for Classifying Acute Lymphoblastic Leukemia Cells. In: Rodríguez-González, A.Y., Pérez-Espinosa, H., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Olvera-López, J.A.

(eds) Pattern Recognition. MCPR 2023. Lecture Notes in Computer Science, vol 13902. Springer, Cham. https://doi.org/10.1007/978-3-031-33783-3_24

Apéndice A

Algoritmo código de segmentación semántica no supervisada con mapas de calor RISE.

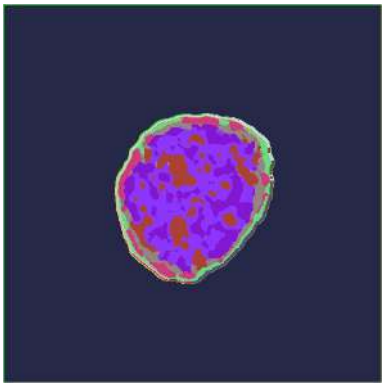
Algoritmo 1: Segmentación semántica no supervisada con mapas de calor RISE

Data: Wm, bm, Wc = Inicializar()
Input: I = {vn $\in \mathbb{R}^3$}//$ImagenRGB$,μ // Peso para Lcon, T // Número de iteraciones
for *t = 1 to T* **do**
    {xn} = GetFeats({vn}, {Wm, bm});
    {rn} = { Wcxn };
    {r0n} = Norm({rn});
    {cn} = arg máx$_i$ $r0n, i$ $Lsim = Lsim(\{r0n, cn\})$ $Lcon =$$\mu$Lcon({r0n});
    L = Lsim + Lcon;
    {Wm, bm}, Wc = Update({r0n, cn}, L);
end
for *imagen en imágenes_enmascaradas* **do**
    características = GetFeats(imagen, {Wm, bm});
    explicación = ModeloExplicativo(feature_map);
    agregar explicación a explicaciones;
end
IMPORTANCIA = Promedio(explicaciones);
Result: L, IMPORTANCIA

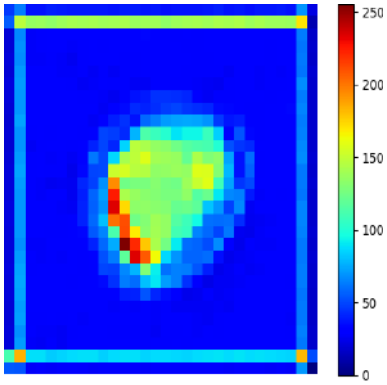
Apéndice B

Mapas de calor adicionales.

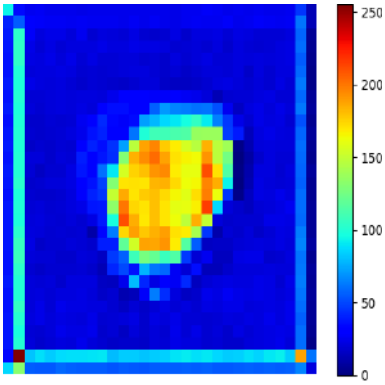
(a)



(b)

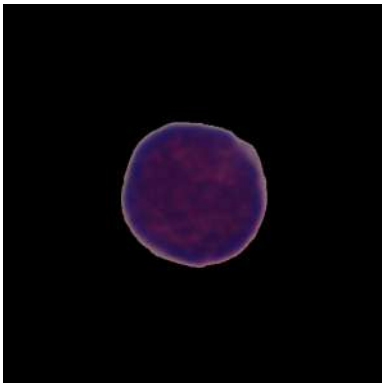


(c) Mapa de calor RISE 7000 máscaras, p1 = 0.5

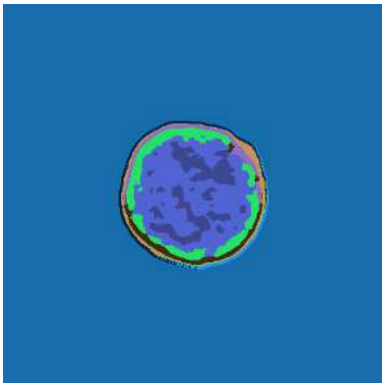


(d) Mapa de calor RISE 7000 máscaras, p1 = 0.9

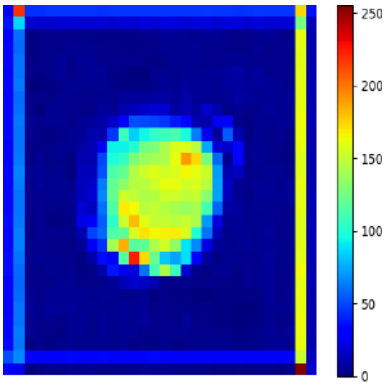
Figura B.1: .



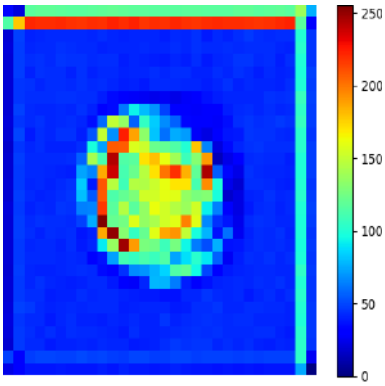
(a)



(b)

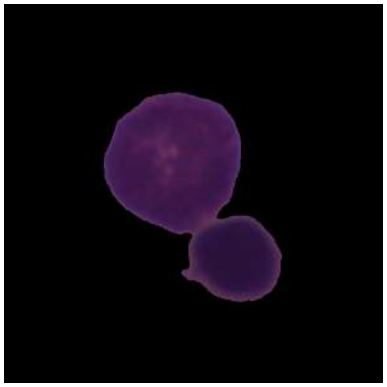


(c) Mapa de calor RISE 7000 máscaras, p1 = 0.5

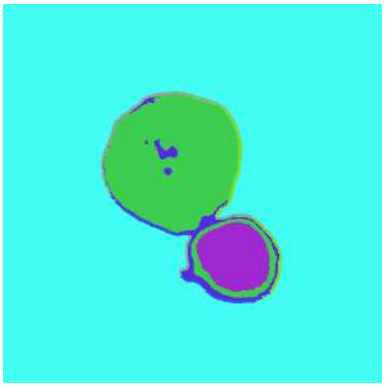


(d) Mapa de calor RISE 7000 máscaras, p1 = 0.9

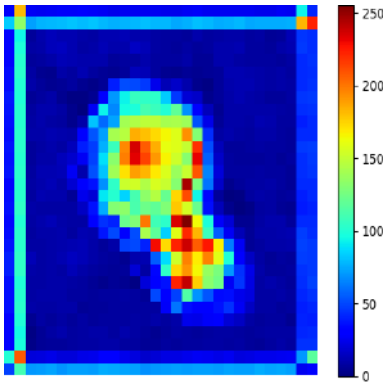
Figura B.2: .



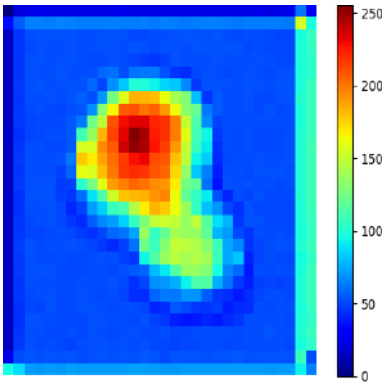
(a)



(b)

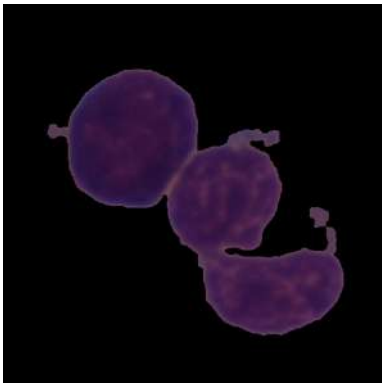


(c) Mapa de calor RISE 7000 máscaras, p1 = 0.5

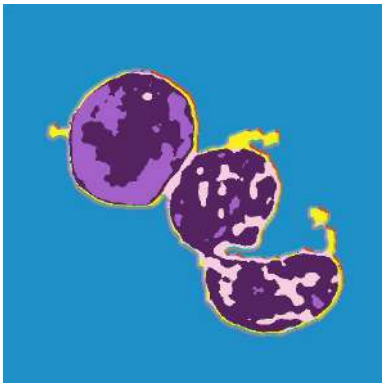


(d) Mapa de calor RISE 7000 máscaras, p1 = 0.9

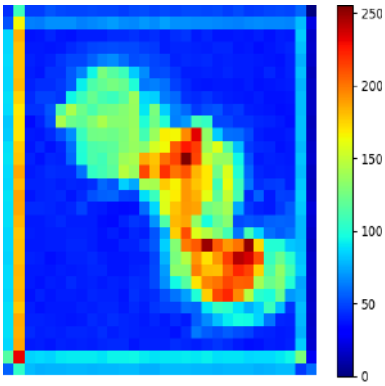
Figura B.3: .



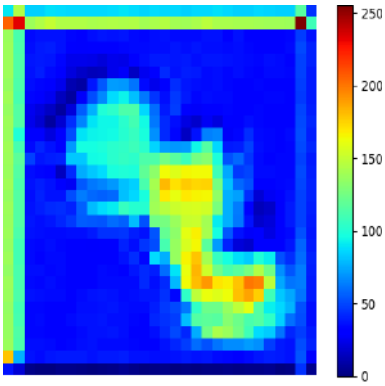
(a)



(b)



(c) Mapa de calor RISE 7000 máscaras, p1 = 0.5



(d) Mapa de calor RISE 7000 máscaras, p1 = 0.9

Figura B.4: .