



**Universidad  
Europea**

**UNIVERSIDAD EUROPEA DE  
MADRID**

**ESCUELA DE ARQUITECTURA, INGENIERÍA  
Y DISEÑO**

**GRADO EN FÍSICA**

**TRABAJO DE FIN DE GRADO**

**Detección de objetos en mapas del fondo  
cósmico de microondas usando  
aprendizaje profundo**

**Pablo Agudo Aguado**

**Dirigido por**

**Dr. Marcos López Caniego Alcarria**

**CURSO 2022-2023**

---

**TÍTULO:** DETECCIÓN DE OBJETOS EN MAPAS DEL FONDO CÓSMICO DE MICROONDAS USANDO APRENDIZAJE PROFUNDO

**AUTOR:** : PABLO AGUDO AGUADO

**TITULACIÓN:** GRADO EN FÍSICA

**DIRECTOR/ES DEL PROYECTO:** : DR. MARCOS LÓPEZ CANIEGO ALCARRIA

**FECHA:** JUNIO DE 2023

## RESUMEN

El proyecto consiste en el diseño, entrenamiento, y evaluación de un modelo de aprendizaje profundo capaz de detectar fuentes compactas en mapas del fondo cósmico de microondas utilizando datos reales de la misión Planck.

La detección de fuentes es un campo de gran importancia en la cosmología actual, al ser un paso necesario para obtener los mapas del fondo cósmico de microondas, y este tipo de técnicas, el aprendizaje automático, pueden servir para automatizar muchos procesos facilitando el análisis de los grandes volúmenes de datos.

Por otro lado, se validará el modelo obtenido aplicándole datos externos y se acotará su rango de utilidad en función de la señal/ruido de las observaciones.

Por último, se estudia la influencia del plano galáctico en este tipo de detecciones separando las fuentes por su latitud.

El modelo obtenido a resultado satisfactorio, con una precisión del 95.83%. Considerándose el modelo de utilidad para observaciones de como mínimo 4 de señal/ruido con una precisión del 93%. Este último resultado se ha obtenido utilizando un catálogo externo lo cual implica la generalización del modelo para otros conjuntos de datos.

Respecto al estudio del plano galáctico se ha observado la existencia de un hilo central donde la detección es muy alta, una zona intermedia del plano galáctico donde el ruido dificulta la detección y la zona fuera del plano galáctico donde la precisión crece rápidamente.

**Palabras clave:** Fondo Cósmico de Microondas. Detección de Fuentes Compactas. Aprendizaje Automático. Redes Neuronales Convolucionales. Reconocimiento de Imagen.

## ABSTRACT

The project consist of the design, training and evaluation of a deep learning model able to detect compact sources in the Cosmic Microwave Background maps by using real data of Plank Mision. The sources detection is a very important field research on cosmology because of it's an inevitable step to obtain the Cosmic Microwave Background maps. In this field, deep learning can be of great help to automate data analysis.

On the other hand, the model obtained will be validated by applying external data and its utility range will be delimited based on the signal/noise of the observation.

Finally, we will study how the influence of the galactic plane affected of this tip of detections. To achieve this we will separate the observations by latitude.

The model obtained has been satisfactory, with an accuracy of 95.83%. Considering the utility model for observations of at least 4 signal/noise with a precision of 93%. This last result has been obtained using an external catalog which implies the generalization of the model to other datasets. Regarding the study of the galactic plane, the existence of a central thread has been observed where precision is very high, an intermediate area of the galactic plane where noise makes detection difficult, and the area outside the galactic plane where precision grows rapidly.

**Keywords:** Cosmic Microwave Background. Compact Sources Detection. Machine Learning. Convolutional Neuronal Network. Image Recognition.



## **AGRADECIMIENTOS**

Quería agradecer a mi tutor Marcos López por su paciencia y las constantes revisiones que han permitido ampliar y pulir enormemente este trabajo.

Me gustaría también aprovechar para dar las gracias a mi madre y a mi pareja por su apoyo, y a mis compañeros y amigos en especial a Irene, Claudia, Laura y Pablo que me han acompañado y ayudado enormemente durante estos cuatro años de carrera.

## Cita

Los mitos de creación de los diferentes pueblos y religiones del mundo palidecen en comparación con la gloria del Big Bang. La historia de tres mil millones de años de la evolución de la vida, desde las moléculas autorreproductoras hasta la civilización, contiene giros y romances que no pueden ser igualados por ningún mito o epopeya.

---

*Liu Cixin, El Problema de los Tres Cuerpos*

**TABLA RESUMEN**

	<b>DATOS</b>
<b>Nombre y apellidos:</b>	Pablo Agudo Aguado
<b>Título del proyecto:</b>	Detección de objetos en mapas del fondo cósmico de microondas usando aprendizaje profundo
<b>Directores del proyecto:</b>	Dr. Marcos López Caniego Alcarria
<b>El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:</b>	NO
<b>El proyecto ha implementado un producto:</b>	NO
<b>El proyecto ha consistido en el desarrollo de una investigación o innovación:</b>	SI
<b>Objetivo general del proyecto:</b>	Diseño, entrenamiento y evaluación de una red neuronal para la detección de fuentes compactas

## Índice de figuras

1.	Historia del Universo . . . . .	13
2.	Mapa de temperatura del FCM . . . . .	15
3.	Ampliación del mapa de polarización Planck 2018 . . . . .	
4.	Mapa de intensidad de emisión sincrotrón . . . . .	19
5.	Mapa de intensidad de emisión de polvo térmico . . . . .	20
6.	Dependencia con la frecuencia de las principales fuentes de contaminación difusa del FCM . . . . .	22
7.	Mexican Hat Wavelet . . . . .	24
8.	Familia de ondículas MHWF . . . . .	24
9.	Ejemplo capa convolucional . . . . .	27
10.	Ejemplo de un bloque convolucional . . . . .	28
11.	Mapas a 30GHz . . . . .	
12.	Estructura del modelo . . . . .	37
13.	Implementación del modelo en python. . . . .	37
14.	Ejemplo de matriz de confusión . . . . .	39
15.	Parches a distintos tamaños . . . . .	
16.	Distribución de centros a distintos tamaños de parche . . . . .	
17.	La precisión y la función de pérdida ( <i>loss</i> ) durante el proceso de entrenamiento . . . . .	46
18.	Matriz de confusión para el modelo entrenado sobre PCCS2. . . . .	47
19.	Parches para las fuentes falladas en la evaluación . . . . .	48
20.	Casos de error en la evaluación . . . . .	48
21.	Resultados de comparar distintos tamaños de parche . . . . .	49
22.	Resultados de comparar la latitud . . . . .	51
23.	Precisión sobre nuevos datos según la S/R de las fuentes. . . . .	52
24.	Distribución de centros del nuevo catálogo para los intervalos de S/R de (5,6] y (6,7] . . . . .	53
25.	Precisión en función de la latitud de los nuevos datos . . . . .	54
26.	Matrices de confusión de la comparación a distintos tamaños de parche . . . . .	64
27.	Matrices de confusión correspondientes al entrenamiento li- mitando la latitud . . . . .	65

## Índice de tablas

1.	Parámetros para los distintos tamaños de parche . . . . .	40
2.	Número de fuentes y negativos según el umbral de latitud. . .	44
3.	Número de fuentes para los distintos intervalos de S/R . . . .	45
4.	Precisión en la detección de fuentes utilizando distintos ta- maños de parche . . . . .	50

## Índice

Índice de figuras	7
Índice de tablas	8
<b>1. RESUMEN DEL TFG</b>	<b>11</b>
1.1. Contexto y justificación . . . . .	11
1.2. Planteamiento del problema . . . . .	11
1.3. Objetivos del proyecto . . . . .	11
1.4. Resultados obtenidos . . . . .	11
1.5. Estructura de la memoria . . . . .	11
<b>2. ANTECEDENTES / ESTADO DEL ARTE</b>	<b>12</b>
2.1. Estado del arte: Cosmología . . . . .	12
2.1.1. La Historia del Universo . . . . .	13
2.1.2. Fondo Cósmico de Microondas . . . . .	15
2.1.3. Contaminación . . . . .	19
2.1.4. Detección de fuentes compactas . . . . .	23
2.2. Estado del arte: Aprendizaje automático . . . . .	26
2.2.1. Aprendizaje profundo y redes neuronales convolucio- nales . . . . .	26
2.2.2. Uso de las redes neuronales en cosmología . . . . .	28
2.3. Contexto y justificación . . . . .	30
2.4. Planteamiento del problema . . . . .	31
<b>3. OBJETIVOS</b>	<b>32</b>
3.1. Objetivos generales . . . . .	32
3.2. Objetivos específicos . . . . .	32
3.3. Beneficios del proyecto . . . . .	32
<b>4. DESARROLLO DEL TFG</b>	<b>33</b>
4.1. Planificación del trabajo fin de grado . . . . .	33
4.2. Descripción de la solución, metodologías . . . . .	34
4.2.1. Datos utilizados . . . . .	34
4.2.2. Generación de los parches negativos . . . . .	36
4.2.3. Creación de las imágenes para el <i>dataset</i> . . . . .	36
4.2.4. Modelo . . . . .	37

4.2.5. Entrenamiento . . . . .	38
4.2.6. Evaluación . . . . .	39
4.2.7. Comparaciones . . . . .	39
4.2.8. Estudio de los límites del modelo evaluándolo sobre un catálogo más amplio. . . . .	45
4.2.9. Estudio de la dependencia con la latitud utilizando nuevos datos. . . . .	46
4.3. Resultados del TFG . . . . .	46
4.3.1. Entrenamiento y Evaluación sobre el catálogo PCCS2. . . . .	46
4.3.2. Distintos tamaños de parche. . . . .	49
4.3.3. Umbrales mínimos de latitud . . . . .	51
4.3.4. Resultados de la dependencia con la S/R sobre nuevos datos . . . . .	52
4.3.5. Resultados de la dependencia de la latitud sobre nue- vos datos . . . . .	54
4.4. Herramientas empleadas . . . . .	55
<b>5. DISCUSIÓN</b>	<b>56</b>
5.1. Utilización de Imágenes . . . . .	56
5.2. Limitación Computacional . . . . .	56
5.3. Limitación en el método . . . . .	56
<b>6. CONCLUSIONES</b>	<b>58</b>
6.1. Conclusiones del trabajo . . . . .	58
6.2. Conclusiones personales . . . . .	59
<b>7. FUTURAS LÍNEAS DE TRABAJO</b>	<b>60</b>
<b>Referencias</b>	<b>61</b>
<b>8. ANEXOS</b>	<b>64</b>
8.1. Matrices de confusión . . . . .	64

## **1. RESUMEN DEL TFG**

### **1.1. Contexto y justificación**

El fondo cósmico de microondas supone una importante fuente de información para la cosmología actual, pero para utilizarlo, se debe detectar y eliminar la contaminación producida por otras fuentes.

La utilización de aprendizaje profundo se espera que permita automatizar parte del proceso de detección de fuentes compactas.

### **1.2. Planteamiento del problema**

Se pretende utilizar la información obtenida en la misión Planck, para el desarrollo de un modelo de aprendizaje profundo (DL) que permita la detección de fuentes compactas, observando la viabilidad de esta técnica con los datos actuales.

### **1.3. Objetivos del proyecto**

El objetivo del proyecto consiste en diseñar, entrenar, evaluar y estudiar las limitaciones, de un modelo de DL para la detección de fuentes compactas. Además de observar cómo afecta la presencia del plano galáctico a estas detecciones.

### **1.4. Resultados obtenidos**

Se ha conseguido una precisión del 95.83 % distinguiendo si una imagen dada corresponde o no a una fuente compacta. El modelo podría ser de utilidad para la detección de fuentes hasta un mínimo de señal/ruido de 4, obteniendo para estas fuentes una precisión del 93 %.

### **1.5. Estructura de la memoria**

En antecedentes, se da contexto y explica la detección de fuentes, además de explicar el funcionamiento de una red neuronal convolucional. En objetivos se detalla lo que se espera realizar en el proyecto. En el desarrollo se explican los datos y el modelo utilizado, así como las tareas que se han tenido que realizar y los resultados. En discusión se expresan las limitaciones surgidas en el proyecto. En conclusiones se contextualizan y valoran los resultados.



## 2. ANTECEDENTES / ESTADO DEL ARTE

### 2.1. Estado del arte: Cosmología

El fondo cósmico de microondas (FCM) desde su detección en 1964, ha constituido una parte fundamental de la cosmología. Aspectos como su equivalencia con un cuerpo negro y el estudio de las anisotropías por parte de misiones espaciales como COBE, WMAP y Planck, nos han permitido entender mejor la evolución del Universo, su geometría y su composición.

En la actualidad, el modelo cosmológico  $\Lambda$ CDM (Lambda-Cold Dark Matter) está basado en la teoría del Big Bang, en la existencia de energía oscura ( $\Lambda$ ) y en la materia oscura fría (materia oscura cuya velocidad es no-relativista). Utilizando este modelo misiones como Planck calculan los valores que deben tener los parámetros del Universo para estar en concordancia con principalmente 4 observaciones, la estructura a gran escala del universo, la expansión acelerada del universo, las abundancias de los elementos ligeros (hidrogeno, helio y litio) y por supuesto, el fondo cósmico de microondas. Por esta razón, y debido al modelo de concordancia es necesario detectar y limpiar de contaminantes la radiación del FCM, puesto que modificaciones en el espectro de potencias de las anisotropías, afectaría a parámetros tan importantes como la constante de Hubble o la curvatura del Universo.

### 2.1.1. La Historia del Universo

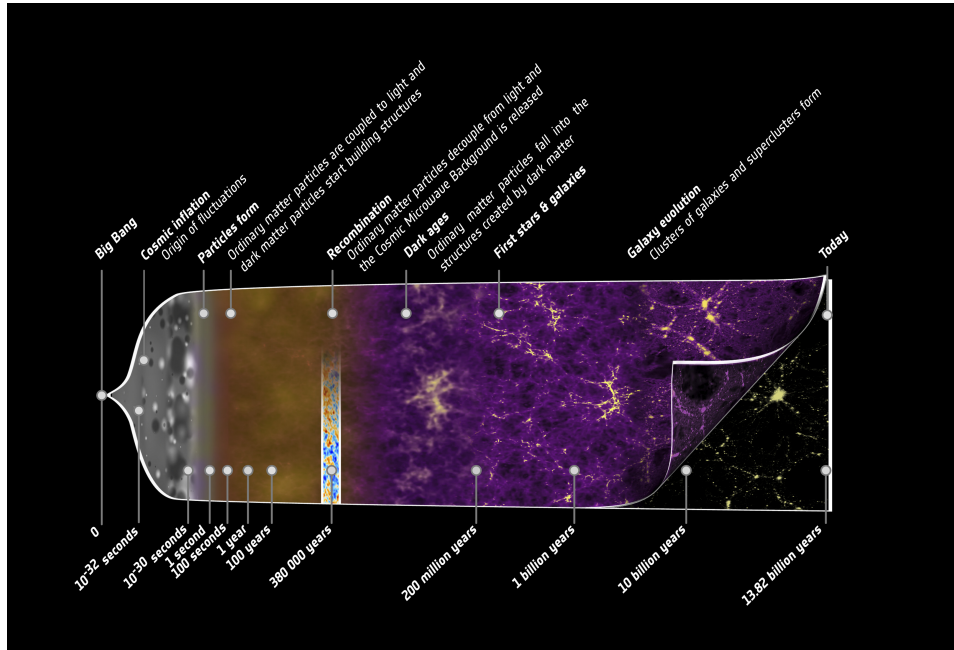


Figura 1: Historia del Universo. Derechos: ESA ([http://www.esa.int/spaceinimages/Images/2013/03/Planck\\_history\\_of\\_Universe\\_zoom](http://www.esa.int/spaceinimages/Images/2013/03/Planck_history_of_Universe_zoom))

El Big Bang constituye el inicio del espacio y el tiempo desde un punto infinitamente denso. Aproximadamente, entre los  $10^{-36}$  y  $10^{-32}$  segundos tras el Big Bang, se produce una expansión exponencial del Universo, aumentando su tamaño en  $10^{26}$ . Este proceso se llama inflación, y, aunque existen discrepancias en su funcionamiento, es aceptado como solución al problema del horizonte y la planicidad del Universo [1]. Consistiendo el problema del horizonte en que, utilizando la edad estimada del Universo, podemos observar puntos alejados entre sí una distancia tal que, la luz, y, por tanto, la información, no tuvo tiempo de viajar entre ellos, pero que, aun así, son equivalentes. Esto invita a pensar que en el Big Bang, el Universo era perfectamente homogéneo, algo imposible basándose puramente en la teoría del Big Bang debido, de nuevo, a la inconexión casual entre distintos puntos. El proceso de expansión permite más tiempo para alcanzar el equilibrio térmico en el Universo, a la vez que, cualquier diferencia queda reducida drásticamente por el enfriamiento causado por la expansión.

A los 3 minutos del Big Bang, la densidad de energía ya había bajado lo suficiente para crear nucleones y se empezaron a formar los primeros núcleos de elementos ligeros, hidrógeno y helio [2]. Aun así, la energía seguía siendo demasiado alta y los electrones interactuaban fuertemente con la luz. Como consecuencia, la luz no podía viajar libremente y los electrones no podían formar átomos estables.

Después de 380000 años, se produce la recombinación. La temperatura había descendido hasta aproximadamente los 3000 °C. Los protones y electrones se combinaron formando átomos. La luz que había estado ‘encerrada’ en el plasma caliente dejó de sufrir colisiones, viajando libremente en línea recta. Ese proceso, ocurrido en todo el Universo, generó una radiación que se propaga desde todas direcciones y que nos muestra información de como era el Universo en el momento de la recombinación.

Esta radiación es actualmente llamada radiación de fondo cósmico de microondas (RFCM), dado que en la actualidad, los haces de luz que correspondían a 3000 °C han perdido energía por la expansión del universo sufriendo corrimiento al rojo hasta las ondas microondas.

Tras la recombinación, empieza la conocida como época oscura. Se conoce así dado que no existían estrellas, y la única luz existente era la RFCM.

Gracias a la atracción gravitatoria, las primeras estrellas se formarían aproximadamente tras 150 millones de años del Big Bang. En este punto, empieza la época llamada como reionización debido a que la energía aportada por la fusión nuclear en el interior de las estrellas permite ionizar los átomos que habían estado millones de años estables.

### 2.1.2. Fondo Cósmico de Microondas

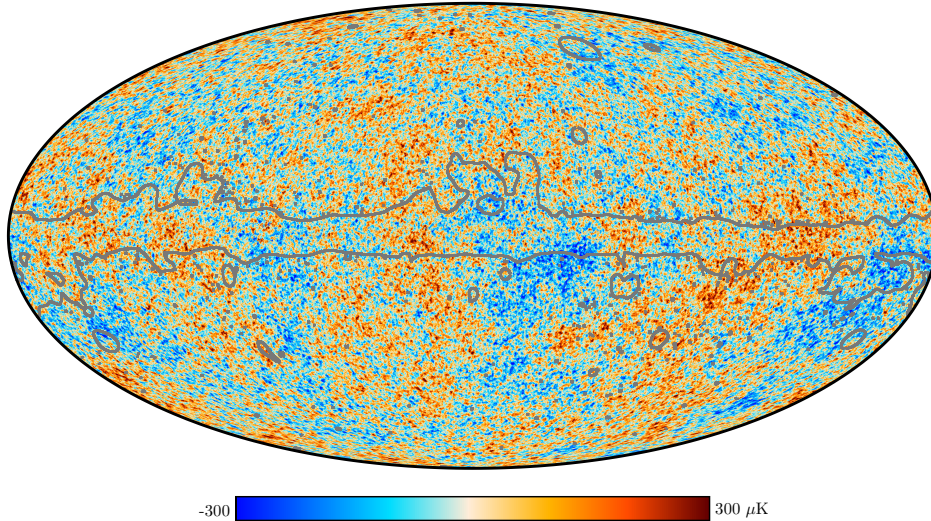


Figura 2: Mapa de temperatura del FCM. Planck 2018 SMICA CMB temperature map. Derechos: ESA ([https://www.aanda.org/articles/aa/full\\_html/2020/09/aa33880-18/F6.html](https://www.aanda.org/articles/aa/full_html/2020/09/aa33880-18/F6.html))

La detección del fondo cósmico de microondas se produjo en 1964 por Arno Penzias y Rober Wilson [3]. Fue un descubrimiento clave para asentar la teoría del Big Bang, que ya había sido propuesta tras el descubrimiento de la expansión del Universo por Hubble en 1929.

El Explorador del Fondo Cósmico (COBE) fue la primera misión espacial centrada en el estudio del FCM. Lanzada en 1989, duró 4 años. Hasta ese momento, el FCM suponía una radiación casi perfectamente homogénea e isotrópica, gracias al instrumento DMR del satélite fueron medidas las inhomogeneidades del FCM en 1990 [4].

Estas pequeñas diferencias en la temperatura se conocen como anisotropías.

La misión espacial COBE demostró, además que la RFCM suponía la radiación de cuerpo negro más perfecta descubierta, siendo su temperatura de 2.725 K. Esto fue posible en 1996 por D. Fixsen y sus compañeros gracias al instrumento FIRAS [5].

Se distinguen dos tipos de anisotropías dependiendo del origen de estas.

#### Anisotropías primarias

Son aquellas que ocurrieron en el momento de la formación del FCM, es decir, las inhomogeneidades presentes en el momento de la recombinación. Estas fueron causadas por fluctuaciones en el plasma primordial, y serían el origen de la estructura a gran escala del Universo.

### **Anisotropías secundarias**

Son aquellas que se producen en el camino de la radiación entre su origen y nosotros. Se deben a distintas interacciones.

### **Efectos gravitacionales**

El método por el cual la temperatura del FCM se ve afectada por la gravedad se llama *Sachs-Wolfe effect* (ISW).

La situación es la siguiente: cuando un fotón que cae en un pozo de potencial y vuelve a salir, este no sufre ningún cambio neto en su energía. Pero esto solo se cumple si la profundidad del pozo es constante. Si no se cumple, la energía del fotón cambiará, en nuestro caso, alterando la temperatura medida del FCM.

Efectos gravitatorios causantes de anisotropías secundarias:

- *Early ISW*: Antes de la recombinación, la contribución de la radiación a la densidad del Universo no era despreciable, esto era debido a que la luz interactuaba fuertemente con electrones y bariones. En el momento de la recombinación, estas interacciones desaparecen o disminuyen drásticamente, permitiendo viajar a los fotones libremente, pero, también, disminuyendo la densidad de energía, siendo esto equivalente, en relatividad general, a la disminución de densidad del Universo provocando el efecto anterior.
- *Late ISW*: En un Universo dominado por la energía oscura  $\Lambda$ , la geometría de este sería abierta. Llegaría un momento de rápida expansión donde la densidad del Universo decaería provocando el efecto ISW.
- *Rees-Sciama*: Consiste en el efecto ISW causado por los cambios en la estructura o disposición de la materia.
- *Lentes Gravitacionales*: Las lentes gravitacionales, causadas por la acumulación de grandes cantidades de materia, deforman la métrica del espacio-tiempo, si bien no cambian la energía de los fotones como en los casos anteriores, sí modifican la trayectoria de estos, cambiando el FCM que observamos desde la Tierra.

- *Ondas gravitacionales:* Aunque las ondas gravitacionales puedan provocar el efecto ISW, sería apreciable a distancias mayores a la del propio FCM.

### **Efectos de dispersión por reionización**

Tras la recombinación y la época oscura, comienza la formación de estrellas. Estas consiguen generar energía por fusión nuclear, esa energía generará iones y electrones libres que interactuarán con los fotones del FCM.

Una parte de la RFCM sufrirá dispersión, esto significa que los fotones recibidos podrían haber sido emitidos en una dirección distinta a la observada. La contribución de este efecto dependerá de la distancia a la que se considere que empezó la reionización.

También existe el efecto *Sunyaev-Zel'dovich (SZ)*. Los fotones del FCM podrían atravesar las nubes de gas caliente ionizado que existen en el interior de cúmulos galácticos. Este gas caliente, no solo generaría cambios en la trayectoria, sino que cambiaría la energía de los fotones por el efecto Compton inverso.

### **Polarización del FCM**

Además de la intensidad de la luz, otro campo de estudio es la polarización de la luz recibida.

Para crear polarización, se necesita una interacción de dispersión o scattering de un fotón con un cuádruplo. Este tipo de interacción solo se podría dar en la fase final antes de la recombinación, por esta razón, observar la polarización nos da información sobre cómo era el Universo antes del FCM.

Existirían dos modos de polarización en el FCM. El primero, E-mode, fue detectado por primera vez en el 2002 [6]. Los haces polarizados en este modo son perpendiculares entre sí.

En segundo modo de polarización B-mode, sus haces forman 45 grados entre sí. La misión LiteBIRD de la agencia espacial japonesa tiene por objetivo principal estudiar esta polarización en el FCM. Es interesante debido a que la inflación cósmica, dada al principio del Universo, debería haber creado este tipo de polarización a través de ondas gravitacionales. Con ella se obtendría una confirmación de la inflación, además de información anterior a la recombinación.



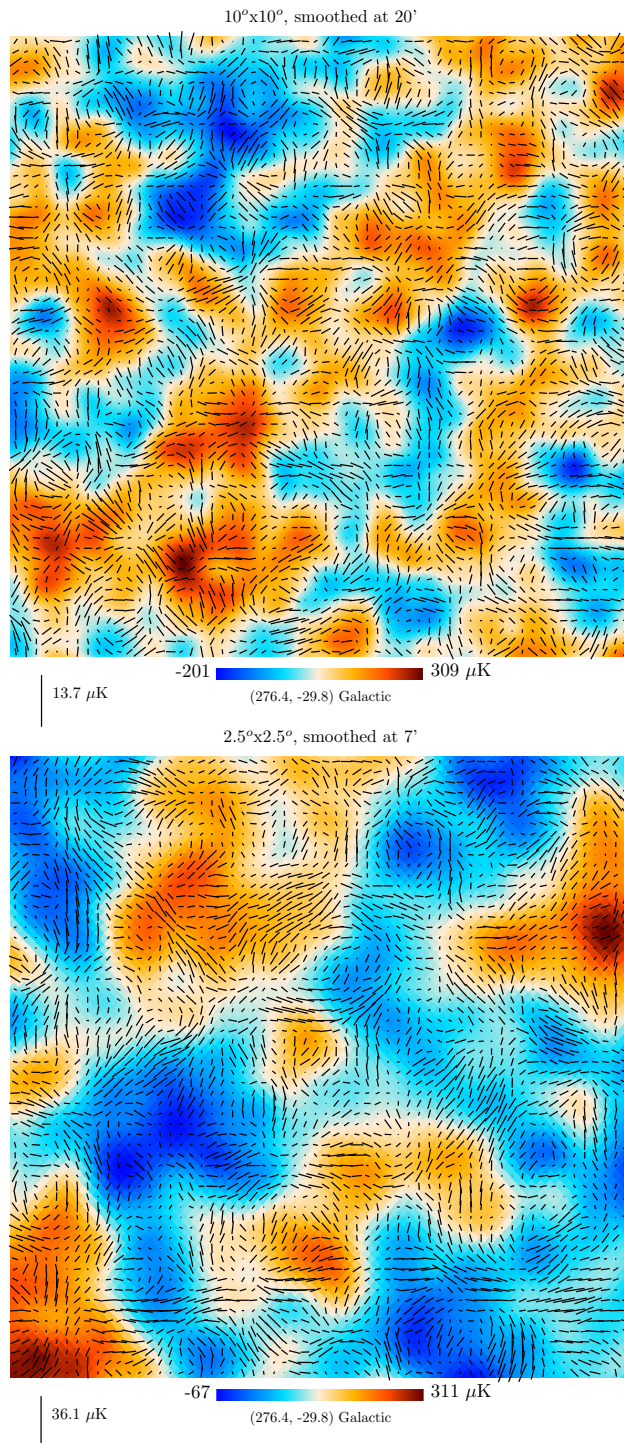


Figura 3: Ampliación del mapa de polarización Planck 2018. El fondo coloreado representa las anisotropías de temperatura suavizadas y superpuesta está la polarización, para observar la correlación. La figura superior es un parche de 10°x10° del polo sur de la eclíptica, la figura inferior es un parche de 2.5°x2.5° de la misma zona. Derechos: ESA ([https://www.aanda.org/articles/aa/full\\_html/2020/09/aa33880-18/F7.html](https://www.aanda.org/articles/aa/full_html/2020/09/aa33880-18/F7.html)) [7]

### 2.1.3. Contaminación

A la hora de observar el FCM, se está observando un rango de frecuencias donde existen otras fuentes de radiación, que se juntarán con los fotones procedentes del FCM. Estudiar cuál es la contribución de estas fuentes y eliminarlas de los datos obtenidos es imprescindible para el estudio del FCM y sus anisotropías.

La mayoría de estas fuentes provienen de nuestra galaxia, se pueden ver algunos ejemplos y su dependencia con la frecuencia en la imagen 6.

Debemos distinguir entre la emisión difusa y la emisión compacta de la galaxia.

**La emisión difusa** afectará a grandes zonas celestes. Por esta razón, su contribución se representa en mapas celestes, en proyecciones mollweide de todo el cielo, como las que se pueden ver en las imágenes 4 o 5.

Algunos ejemplos son:

- *Emisión sincrotrón:* Esta se emite cuando una partícula cargada es acelerada. En nuestro contexto, esto ocurre cuando los rayos cósmicos interactúan con el campo magnético galáctico. Este tipo de radiación es dominante sobre los 20 GHz.

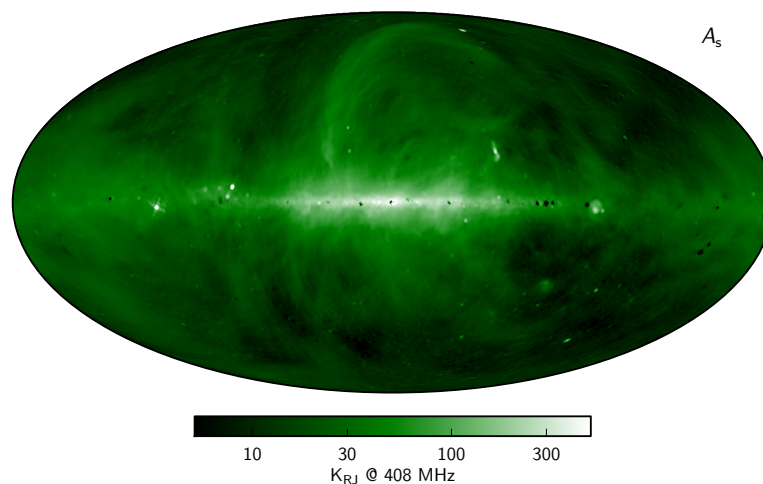


Figura 4: Mapa de intensidad de emisión sincrotrón en 408 MHz. Propiedad: ESA ([https://www.aanda.org/articles/aa/full\\_html/2016/10/aa25967-15/F8.html](https://www.aanda.org/articles/aa/full_html/2016/10/aa25967-15/F8.html))

- *Libre-Libre (Free-Free):* Hace referencia a la interacción de desacele-



ración entre dos partículas libres cargadas. Ocurre cuando electrones de alta energía interactúan con iones en el espacio interestelar. Es dominante en un rango pequeño de frecuencias alrededor de los 50 GHz y carece de polarización.

- *Polvo térmico:* Consiste en granos de polvo que absorben rayos ultravioleta y re-emiten fotones en el infrarrojo lejano. Este polvo, aproximadamente, constituye un cuerpo negro de 20K y su emisión está fuertemente polarizada.

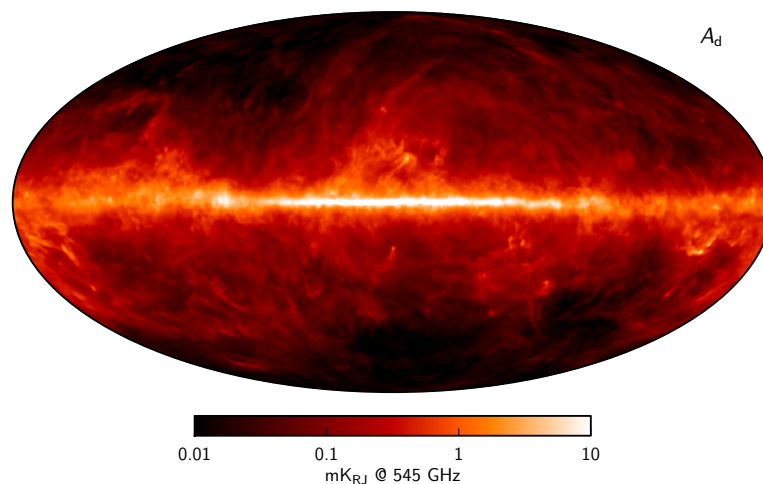


Figura 5: Mapa de intensidad de emisión de polvo térmico en 545 GHz. Propiedad: ESA ([https://www.aanda.org/articles/aa/full\\_html/2016/10/aa25967-15/F11.html](https://www.aanda.org/articles/aa/full_html/2016/10/aa25967-15/F11.html))

- *Polvo en rotación:* Polvo nanométrico que tiene un momento dipolar eléctrico y una rotación en el rango de frecuencias de las microondas.
- *Emisión de CO:* Algunas transiciones entre niveles energéticos del monóxido de carbono pueden contaminar los datos del FCM, principalmente entre los 110 y 345 GHz.

**La emisión compacta** se compone de observaciones puntuales que son detectados en los mapas y guardados en catálogos.

Separándolas según si son galácticas o extragalácticas:

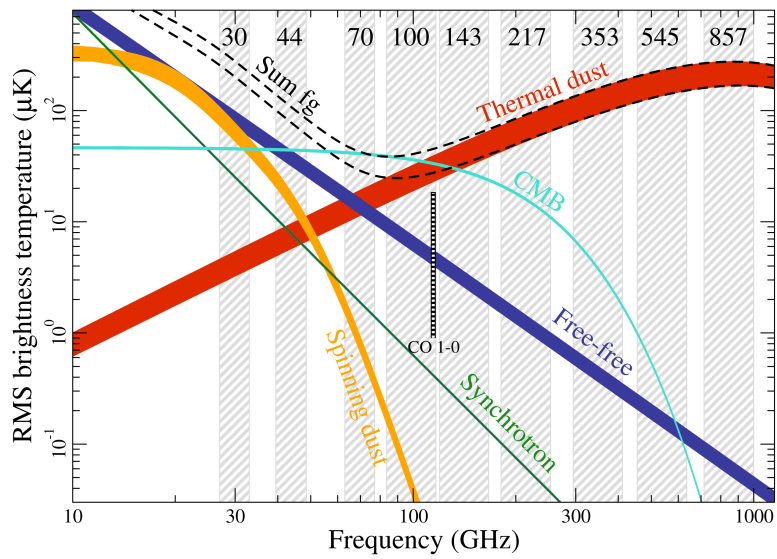
- *Emisiones galácticas:* Estas consisten en emisiones de fuentes puntuales dentro de la galaxia. Algunas de las posibles fuentes son, remanentes

de supernova o concentraciones de polvo interestelar frío.

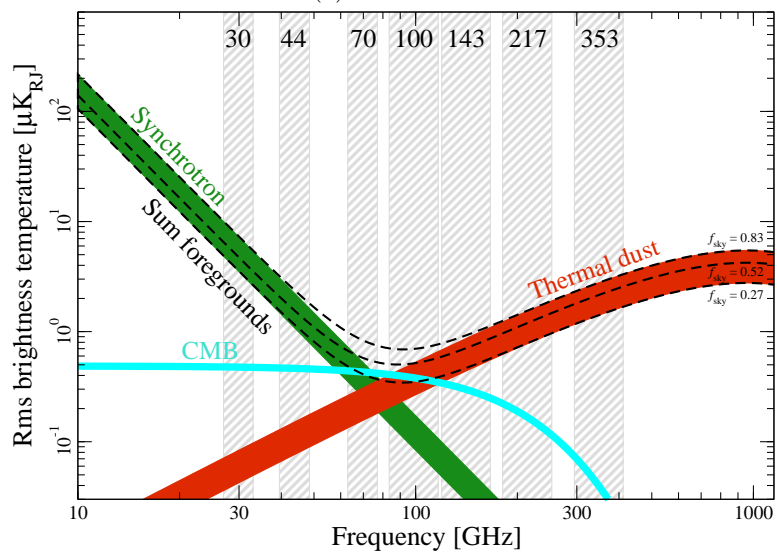
- *Emisiones extragalácticas:* Este tipo de emisiones, debido a la distancia, se ven como fuentes compactas en los MFCM.

Varias de estas emisiones se dan alrededor de núcleos galácticos activos, donde la materia que rodea al agujero negro se ioniza y es acelerada a velocidades cercanas a la luz, emitiendo, por tanto, radiación de sincrotrón. Además, en algunos núcleos llamados blazar, la materia se proyecta en corros de plasma a velocidades relativistas generando desde rayos gamma a emisión en radio. Otras fuentes de contaminación son las galaxias que emiten en radio o galaxias infrarrojas que emiten principalmente en esta frecuencia debido a contener muchas estrellas en formación [8].

Por último, nombrar el efecto Sunyaev-Zel'dovich en cúmulos de galaxias que, como se ha visto antes también es causante de anisotropías secundarias. Consistiendo en la emisión de infrarrojo lejano al impactar fotones, sobre nubes de plasma en el interior de cúmulos galácticos.



(a) Intensidad



(b) Polarización

Figura 6: Dependencia con la frecuencia de las principales fuentes de contaminación difusa del FCM. Las zonas sombreadas corresponden a las zonas observadas por la misión Planck. Derechos: ESA (<http://www.cosmos.esa.int/web/planck/picture-gallery>)

#### 2.1.4. Detección de fuentes compactas

En este trabajo, utilizaremos el segundo catálogo de fuentes compactas de la misión Planck perteneciente a la revisión de 2015 [9]. Por ello, nos centraremos en las técnicas usadas para su creación.

Para el desarrollo del catálogo, los investigadores utilizaron cuatro medidas de la densidad de flujo para cada observación. Utilizando estas densidades se puede hacer un cálculo preliminar de la señal/ruido (S/R) asociada a cada fuente, siendo este, un parámetro que representa la densidad de flujo recibido en el telescopio entre el error asociado.

- 'Detection pipeline photometry' (DETFLUX). Correspondiente a la luz incidente directamente en el telescopio y el tamaño de la fuente.
- 'Aperture photometry' (APERFLUX). Consiste en una integración alrededor de la fuente, utilizando también el fondo para estimar el error.
- 'PSF fit photometry' (PSFFLUX). Consiste en el ajuste del modelo PSF, que tiene en cuenta la posición de la fuente, su tamaño y el ruido de fondo.
- 'Gaussian fit photometry' (GAUFLUX). Corresponde al ajuste de un modelo gaussiano.

En la misión Planck se han utilizado, principalmente, dos métodos para la detección de fuentes compactas [10]: "The Mexican Hat Wavelet 2 algorithm"(MHW2) y "Matched Filter".

Explicaremos el funcionamiento de algoritmo MHW2, ya que es el utilizado en el catálogo PCCS2, utilizado en este trabajo.

##### **The Mexican Hat Wavelet 2 algorithm [11]**

La *Ricker wavelet* (ondícula), usualmente conocida como Mexican Hat Wavelet (MHW), consiste en la segunda derivada de la función gaussiana.

Para la detección de fuentes, se trabaja en el plano, por lo que, partiendo de la gaussiana en dos dimensiones, se aplica un laplaciano para obtener el MHW. Tras otra aplicación de un laplaciano sobre HHW obtenemos MHW2. Esto se puede aplicar sucesivamente, generando una familia de ondículas, como se muestra en la figura 8

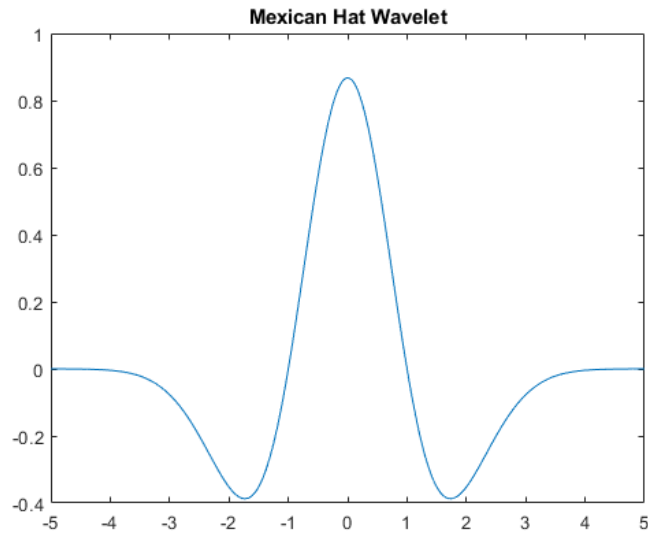


Figura 7: Mexican Hat Wavelet

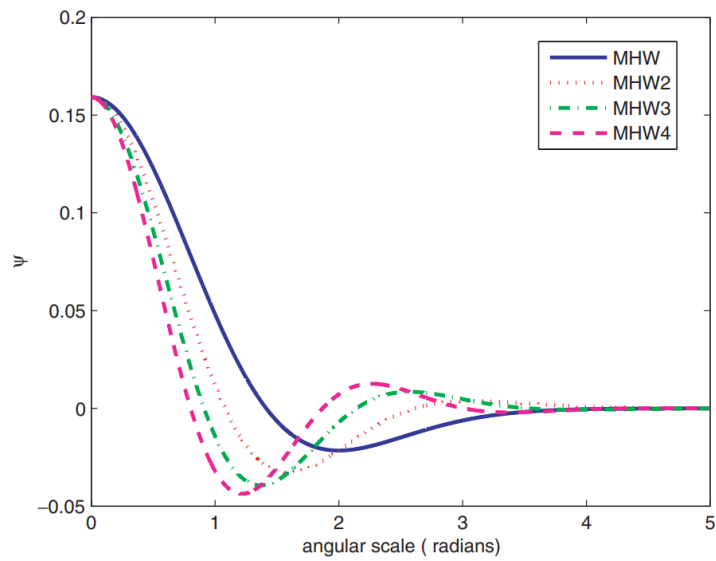


Figura 8: Perfil radial de las 4 primeras ondículas de la Mexican Hat Wavelet Family (MHWF). Propiedad: Oxford Academic (J. González-Nuevo y otros [11])

*The Mexican Hat Wavelet 2 algorithm* utiliza la segunda ondícula de la familia y tiene un único parámetro libre (la escala  $R$ ), que puede ser optimizado

individualmente en cada caso. Se considera un método poco sensible a artefactos o puntos demasiado brillantes, siendo en general, una solución robusta para todas las latitudes [12].

Existen dos implementaciones distintas de este algoritmo, una para los mapas de baja frecuencia y otra para alta frecuencia. El funcionamiento general es común para ambas:

El mapa completo es dividido en parches cuadrados los cuales además de la fuente, deben de contener algo del ruido de fondo, pero deben ser lo suficientemente pequeños para que ese ruido sea casi homogéneo [12]. En concreto, se separa el mapa en 371 parches cuadrados de  $14.65^\circ \times 14.65^\circ$  con  $3^\circ$  de solapamiento entre los parches [10]. Se optimiza de forma independiente en cada parche el parámetro de escala  $R$ , de forma que se maximice la  $S/R$  de las posibles fuentes presentes en el parche. Por último, se debe enmascarar cualquier fuente de señal ruido por encima de 30. Tras la optimización, el parámetro de escala suele rondar la unidad, siendo bajo cerca del plano galáctico y mayor cuando la fuente es pequeña respecto al tamaño del píxel.

La detección se contabiliza dependiendo de la  $S/R$  obtenida en cada fuente. Debido al solapamiento de los parches, se deben quitar las fuentes duplicadas, dejando en el catálogo el valor más alto de  $S/R$ .

Adicionalmente, a la detección por este o el otro método, las fuentes se validan mediante simulaciones y se contrastan con otros catálogos.

A continuación, se expresan las ecuaciones correspondientes al método de detección para la familia de ondículas MHWF.

La amplificación  $\lambda_n$  corresponderá al aumento en  $S/R$ , por tanto, será el parámetro a maximizar. Además, podemos observar la importancia de que el parche contenga ruido de fondo en los parámetros  $\sigma$  y  $\sigma_{w_n}$ .

Ecuación general para la familia MHW, donde  $n$  es el número de la ondícula:

$$\Psi_n(x) = \frac{(-1)^n}{2^n n!} \Delta^n \psi(x) \quad \text{donde} \quad \psi = \frac{e^{-x^2}/2}{2\pi} \quad \text{es la gaussiana 2D}$$

Transformada de Fourier de  $\Psi_n(x)$  sobre un punto arbitrario  $x$  del plano:

$$\hat{\Psi}_n(k) = \int_0^\infty dx x J_0(kx) \Psi_n(x) \quad \text{donde } k \text{ es el numero de onda y } J_0(kx) \text{ la función de Bessel}$$

Coefficientes de la ondícula (*wavelet coefficients*):

$$w_n(b, R) = \int dk e^{-ik \cdot b} f(k) \hat{\Psi}_n(kR)$$

La RMSD (*Root-mean-square deviation*) del fondo después de ser filtrado por el coeficiente de onda  $w_n$  a una escala  $R$  ( $\sigma_{w_n}$ ). En función de la RMSD del fondo sin filtrar ( $\sigma$ ):

$$\sigma_{w_n} = \frac{\sigma l_p \sqrt{(2n)!}}{R \pi 2^n n!} \quad \text{donde } l_p \text{ es el tamaño del píxel}$$

La amplificación para un valor inicial de temperatura  $T_0$ :

$$\lambda_n = \frac{w_n / \sigma_{w_n}}{T_0 / \sigma}$$

## 2.2. Estado del arte: Aprendizaje automático

### 2.2.1. Aprendizaje profundo y redes neuronales convolucionales

El aprendizaje profundo (DL) consiste en un tipo de aprendizaje automático o computacional, que concatena varias capas de neuronas. La neurona, la unidad más pequeña de la red, no es más que un valor dependiente de una o varias neuronas anteriores y cuya salida estará sometida a una función de activación. Cuanto afecta a una neurona los valores anteriores, son los llamados pesos, que serán ajustados en el entrenamiento.

Para la clasificación de imágenes, que es lo que se realizará en este trabajo, comúnmente se utilizan redes neuronales convolucionales (CNN) con un proceso de aprendizaje supervisado. Las CNN son un tipo de aprendizaje profundo que recibe su nombre por el tipo de capas utilizadas y el aprendizaje supervisado consiste en suministrar a la red ejemplos ya clasificados de la tarea que se quiere realizar como método de entrenamiento.

A continuación, explicaremos el funcionamiento de las capas más importantes de una CNN y que han sido usadas en este trabajo.

La capa más importante es la convolucional 2D. Consiste en un kernel, una matriz normalmente cuadrada, que será multiplicada por secciones de la matriz de entrada.

Empezando en el primer elemento, el kernel se multiplicará por este, y se desplazará una cantidad fijada, llamada *strides*. Dependiendo del tamaño

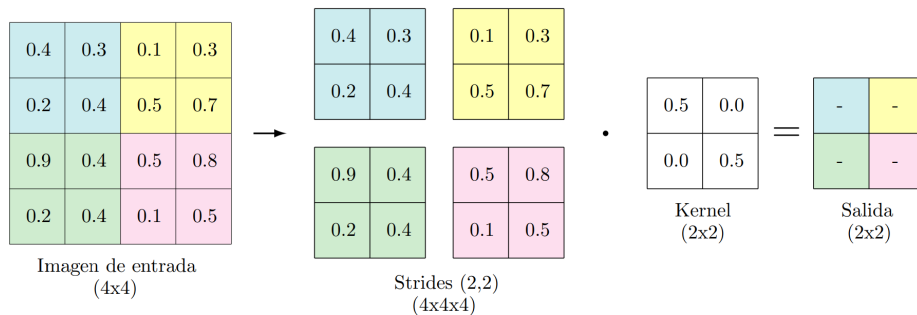


Figura 9: Proceso de multiplicación de un kernel en una capa convolucional

del kernel y cuántos elementos se desplace tras cada multiplicación (*strides*), tendremos el tamaño de los mapas de salida. Este proceso se ilustra en la imagen 9.

Normalmente, no se utiliza un único kernel, sino que se aplican varios de estos, llamándolos filtros.

Cada filtro o kernel genera un mapa independiente, y los elementos del filtro son parámetros que optimizaremos en el proceso de entrenamiento.

Todos los valores de las matrices de salida corresponderán a una neurona, y antes de la siguiente capa, se someten a una función de activación.

Tras una capa convolucional, se crean gran cantidad de neuronas, para el ejemplo de la imagen 9, si aplicamos, por ejemplo, 4 filtros, obtendríamos 4 matrices de 2x2, es decir, 16 neuronas para una imagen 4x4. Si aplicamos esta misma capa sobre una imagen de 128x128, la salida sería de 16 mil neuronas.

Dependerá de cómo se configure la capa convolucional, pero, en general, se generan demasiadas neuronas y no es posible concatenar varias de estas capas. Por ello, se utilizan las llamadas *pooling layers*. Este tipo de capa está pensada para reducir el número de neuronas de la capa siguiente. Una matriz de entrada se separará en grupos y se devolverá, únicamente el valor máximo o la media de los valores.



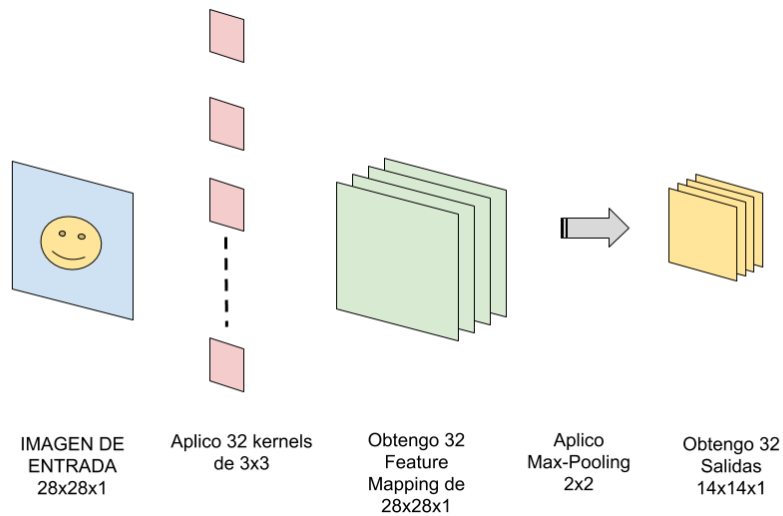


Figura 10: Ejemplo de funcionamiento de un bloque de capas convolucionales. Juan Ignacio Bagnato, 2018, (<https://www.aprendemachinelearning.com/wp-content/uploads/2018/11/cnn-07.png>)

Una vez se han añadido tantos pares (convolucional, *pooling*) como se quiera, se deben colocar un par de capas que nos permitan obtener resultados. En primer lugar, como se puede ver en la figura 10, nuestra imagen de entrada se ha convertido en un grupo de varias matrices. Es necesario utilizar una capa tipo Flatten para concatenar todas estas neuronas en una única fila.

A continuación, se coloca una capa tipo ‘Dense’, que hace referencia a que está densamente conectada, en concreto, esta capa conectará todas las neuronas anteriores con una, un par o unas decenas de neuronas. Esta última fila de neuronas será la salida, y cada una de las neuronas representa una categoría, por lo que se colocarán tantas como número de categorías queramos clasificar las imágenes.

### 2.2.2. Uso de las redes neuronales en cosmología

Se van a revisar una serie de artículos donde se utiliza aprendizaje profundo para la detección de distintas fuentes en el ámbito de la cosmología.

- *‘Constraining the polarisation flux density and angle of point sources*

*by training a convolutional neural network'* por J. M. Casas y otros (13 Ene 2023). Referencia: [13].

En este trabajo realizan una estimación de la densidad de flujo de polarización y el ángulo de polarización para las fuentes compactas superpuestas al FCM.

Utilizan parches simulados de forma realista del FCM añadiendo en el centro una fuente compacta y contaminación por polvo térmico y ruido de los instrumentos.

Evaluando el modelo resultante sobre las fuentes polarizadas del catálogo PCCS2, obtienen resultados similares a los de este. Concluyendo que es un modelo prometedor para la estimación de la densidad de flujo y ángulo de polarización comparándolo con el rendimiento de otras alternativas de estimación ciega.

- *'Astronomical source detection in radio continuum maps with deep neural networks'* por S. Riggi y otros (5 Dic 2022). Referencia: [14].

Este trabajo no está centrado en el FCM, sino en la detección de señales de radio, pero el método utilizado consiste en una red neuronal convolucional con lo cual sigue siéndonos de sumo interés.

Los datos que utilizan fueron tomados por el australiano SKA Pathfinder telescope (ASKAP) [15] los cuales forman parte de un desafío de detección, para la creación de un mapa de la evolución del universo (EMU) por sus siglas en inglés [16]. Para la creación de este mapa de la evolución del universo, se espera hacer uso de un nuevo radiotelescopio llamado SKA (*Square Kilometre Array*) [17], este supondría el mayor radiotelescopio del mundo que podría detectar millones de fuentes. En este contexto se entiende la necesidad de automatizar, sin perder precisión, todos los procesos para la detección de las fuentes.

Algunas de las mejoras que mencionan que han conseguido son:

- A diferencia de los algoritmos mayormente usados se puede aplicar sobre grandes trozos del mapa y no necesariamente en parches individuales.
- Puede devolver una máscara de segmentación de las fuentes detectadas, no únicamente su posición con una sección cuadrada que la contenga.

- Puede detectar artefactos creados por fuentes muy brillantes.
- Puede detectar fuentes extensas que estén partidas en varias islas.

Estas mejoras aunque pueden no reflejarse directamente en la precisión sí pueden ser de gran utilidad para la automatización de estos procesos. Sus resultados finales fueron de un 90 % de precisión sobre fuentes compactas, y un 80 % sobre fuentes extensas aunque comentan que tienen una confianza moderada del 60 %.

- '*Machine learning approach to the detection of point sources in maps of the CMB temperature anisotropies*' por P. Diego-Palazuelos y otros (28 Feb 2023). Referencia: [18].

Este artículo es la referencia más parecida que se ha encontrado al problema planteado en este trabajo. Consiste en la detección ciega por medio de una red neuronal convolucional de fuentes compactas extragalácticas.

Los datos utilizados son simulaciones del FCM a 143 GHz donde se han incluido ruido instrumental simulado, contaminación difusa de sincrotrón, polvo térmico, polvo en rotación, free-free y contaminación compacta por efecto SZ extragaláctico. El mapa obtenido ha sido troceado en parches de  $3.62^\circ \times 3.62^\circ$  y estos parches han sido divididos en primer lugar, según el brillo de fondo y luego cada categoría en anillos de distinta latitud.

Sus resultados preliminares muestran que para las zonas oscuras, donde el FCM sobresale de la contaminación, los resultados son tan buenos como en la utilización de *Matched Filter* (MH) y lo supera en rendimiento en los casos más brillantes cerca y sobre el plano galáctico.

### 2.3. Contexto y justificación

Las numerosas misiones espaciales dedicadas al estudio del FCM han generado una gran cantidad de datos, tardando años en convertirlos en resultados. Este es, sin lugar a duda, un proceso necesario, donde se van mejorando los métodos utilizados y donde siempre se utilizan los datos anteriores como validación de los nuevos. Por esta razón, y viendo los importantísimos resultados que están dando los modelos de DL en campos muy diversos, parece razonable preguntarnos si con los resultados ya obtenidos en estas misiones espaciales, podemos crear un modelo capaz de replicar estos resultados y en

un futuro ayudarnos a obtener nuevos, integrándose en la cadena de trabajo para ayudar a la automatización del análisis de los datos.

Además, ya se ha visto en otras disciplinas, cómo modelos de inteligencia artificial consiguen mejorar el resultado de algoritmos clásicos, debido a que, estos internamente, pueden estar manejando una cantidad de parámetros muchísimo mayor a la de un algoritmo clásico, convirtiéndose, en muy poco tiempo, en la mejor opción para el reconocimiento de patrones o imágenes.

Otra razón para la realización de este proyecto, es el aprendizaje propio sobre el uso de las redes neuronales como método de análisis de datos. Aplicando esta técnica, que tiene importantes aplicaciones en investigación y en la empresa privada, a un problema real de la cosmología actual.

#### **2.4. Planteamiento del problema**

Se pretende desarrollar un modelo de DL que permita la detección de fuentes compactas a 30GHz.

A diferencia de los artículos anteriores, se utilizará datos reales de la misión Planck, tanto el mapa a 30GHz como el catálogo de fuentes ya detectadas. Será necesario observar la viabilidad de conseguir un modelo de detección fiable con los ejemplos actualmente disponibles, para ello, se tendrá que estudiar los límites de utilidad del modelo. Esto se realizará validando los resultados sobre un segundo catálogo de Planck.

### **3. OBJETIVOS**

#### **3.1. Objetivos generales**

El objetivo general es el desarrollo de un modelo de aprendizaje profundo capaz de identificar fuentes compactas en mapas del fondo cósmico de microondas.

#### **3.2. Objetivos específicos**

- Análisis y optimización de los parámetros de entrada y entrenamiento, tales como la resolución.
- Análisis de las limitaciones del modelo, acotando su umbral de detección.
- Estudiar las características de la bóveda celeste y cómo afectan a la detección de fuentes. Principalmente, la presencia del plano galáctico.
- Estudio del parámetro S/R de una fuente, así como su importancia en la detección o determinación de una fuente.

#### **3.3. Beneficios del proyecto**

En primer lugar, el desarrollo de este tipo de herramientas permitirá automatizar procesos en la detección de fuentes. Para el caso más concreto de este proyecto sirve como estudio de viabilidad de la utilización de DL con los datos reales ya obtenidos.

## 4. DESARROLLO DEL TFG

### 4.1. Planificación del trabajo fin de grado

- **Estado del arte:** Repaso de los artículos relacionados con el proyecto, principalmente, de la misión Planck, así como los artículos relacionados con la utilización de aprendizaje profundo en este campo.  
Familiarización con las bibliotecas de python: keras, healpy y astropy, y con el sistema de archivos *Flexible Image Transport System (fits)*, el cual, es utilizado comúnmente en astronomía para la transmisión de datos correlacionados.
- **Generación de los parches negativos o vacíos:** Buscar puntos del MFCM que carezcan de una fuente en nuestro catálogo para crear imágenes que no contengan ninguna fuente.
- **Creación de las imágenes.**
- **Estudio de las redes convolucionales y creación de la estructura del modelo.**
- **Entrenamiento y evaluación.**
- **Comparaciones:** Estudio del comportamiento del modelo en diferentes situaciones.  
Este punto implica repetir la generación de parches, las imágenes, el entrenamiento y la evaluación varias veces, efectuando los cambios necesarios en cada caso.
- **Estudio de los límites del modelo:** Para ello, se ha evaluado el modelo con nuevos datos. Proceso que ha supuesto filtrar, agrupar y crear las imágenes de todos ellos.

## 4.2. Descripción de la solución, metodologías

### 4.2.1. Datos utilizados

Los archivos utilizados han sido descargados a través del portal Planck Legacy Archive [19].

#### **Mapa del fondo cósmico de microondas a 30GHz**

El MFCM que se ha usado como base para obtener las imágenes es el Planck Release 2018 - Light Maps 30GHz NS1024. Este consiste en un mapa basado en Healpix. La definición del mapa es de  $N_{side} = 1024$ , lo que corresponde a 12,582,912 píxeles con una resolución de aproximadamente  $3.43$  ( $'/\text{pix}$ ), es decir, cada píxel ocupa  $3.43$  minutos de arco.

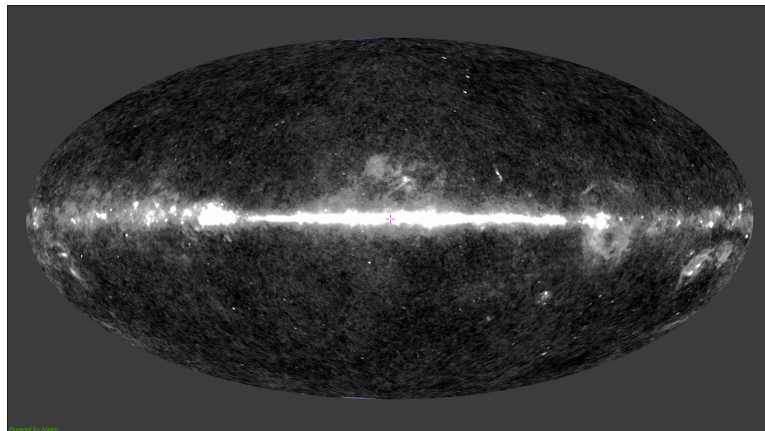
El mapa se muestra en la figura 11 acompañado de un mapa de emisión difusa y otro de emisión compacta, ambos a 30GHz. El mapa de emisión contiene radiación sincrotrón, free-free y emisión anómala. El mapa de radiación del FCM contendrá, en mayor o menor medida, la contaminación mostrada en ambos mapas de emisión.

#### **Catálogo de fuentes para 30GHz**

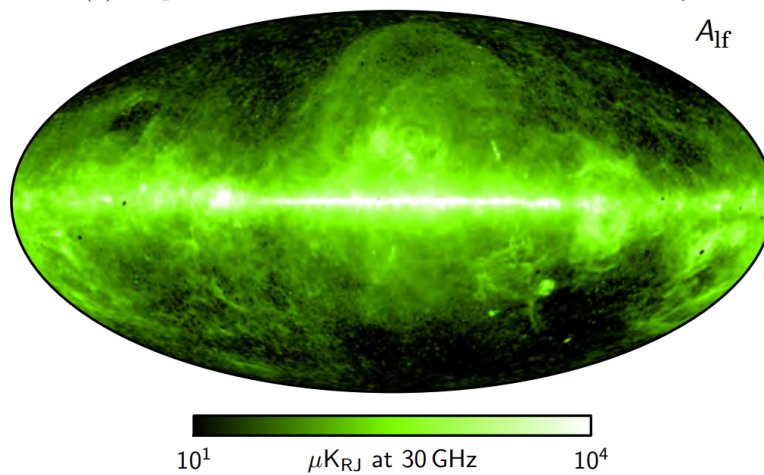
El catálogo de fuentes es el anteriormente nombrado, The Second Planck Catalogue of Compact Sources (PCCS2), que contiene las posiciones en coordenadas galácticas de las detecciones obtenidas a través de las pipelines de detección de fuentes del proyecto Planck, así como, otros datos útiles, como la densidad de flujo o si la fuente ya fue detectada en otro catálogo.

En su artículo de referencia [9], podemos ver que consta de 1560 fuentes compactas, de las cuales, 745 están situadas fuera del plano galáctico, fijado para el rango de 30GHz en  $\pm 30^\circ$  de latitud.

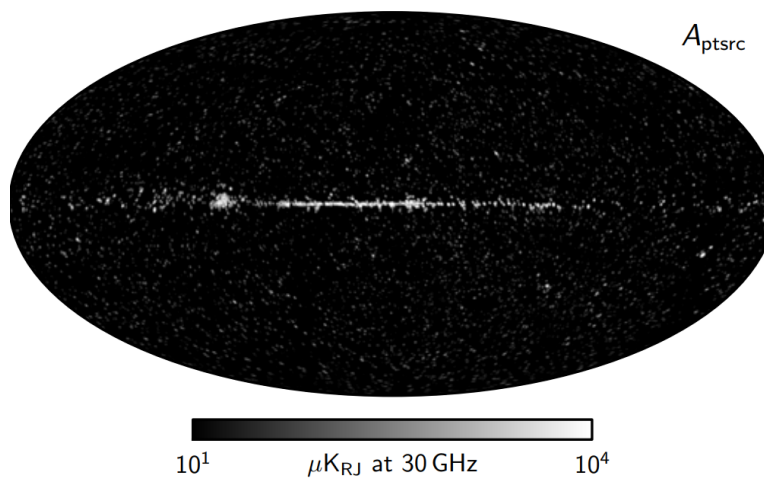
Este catálogo se ha creado considerando un umbral de S/R de 4 para las observaciones en 30 GHz. Todas las observaciones por debajo de este valor, aunque muchas podrían ser fuentes, para asegurar la fiabilidad del catálogo, quedaron descartadas.



(a) Mapa del FCM a 30GHz utilizado en este trabajo



(b) Mapa de emisión difusa a 30GHz. Contiene radiación de sincrotrón, free-free y radiación anómala



(c) Mapa de emisión aproximada de fuentes compactas a 30GHz

Figura 11: Mapas a 30GHz. Propiedad: ESA ([https://www.aanda.org/articles/aa/full\\_html/2020/09/aa33881-18/F53.html](https://www.aanda.org/articles/aa/full_html/2020/09/aa33881-18/F53.html))



#### 4.2.2. Generación de los parches negativos

Con el fin de entrenar una red neuronal en la tarea especificada, es necesario ofrecer dos opciones de elección, fuente o carencia de fuente, así como ejemplos de ambas situaciones. Para esto, se han generado parches negativos, que no son más que imágenes de igual resolución y tamaño a las creadas para las fuentes, pero que muestren un trozo del MFCM que carece de fuente. Nos referiremos a este tipo de parche como parches negativos, o, simplemente negativos, dado que representan el contrario al parche que contiene una fuente.

La técnica usada ha consistido en generar pares de números aleatorios representativos de la longitud y latitud del centro negativo, y, usando el teorema del coseno, comprobar su distancia con todo el resto de los centros, desechando todo par de números cuya distancia con otro centro sea menor a los grados del tamaño del parche.

Con este método, es seguro que no aprovechamos al máximo el espacio disponible, pero es poco probable que alguna implementación más compleja consiga un número significativamente superior de centros válidos para los parches negativos. En concreto, se ha probado un mínimo de un millón de pares aleatorios para asegurarnos de rellenar el espacio disponible.

Debido a errores a la hora de codificar la generación de estos centros, se tuvieron que repetir numerosas pruebas posteriores. Por esta razón, todos los grupos de centros se han exportado y dibujado junto al catálogo original en aladín, comprobando que estuviesen correctamente colocados. Un ejemplo de esto se puede ver en las imágenes de la figura 16.

#### 4.2.3. Creación de las imágenes para el *dataset*

Obtenidas dos listas con los centros de las fuentes del catálogo y los negativos. En primer lugar, se mezclan aleatoriamente para evitar cualquier patrón del posible catálogo. Luego, se utiliza *healpy* para obtener sobre el mapa una proyección cuadrada (*Gnomonic projection*) centrada en cada fuente y del tamaño correcto. Estos datos se exportarán como imágenes en escala de grises.

#### 4.2.4. Modelo

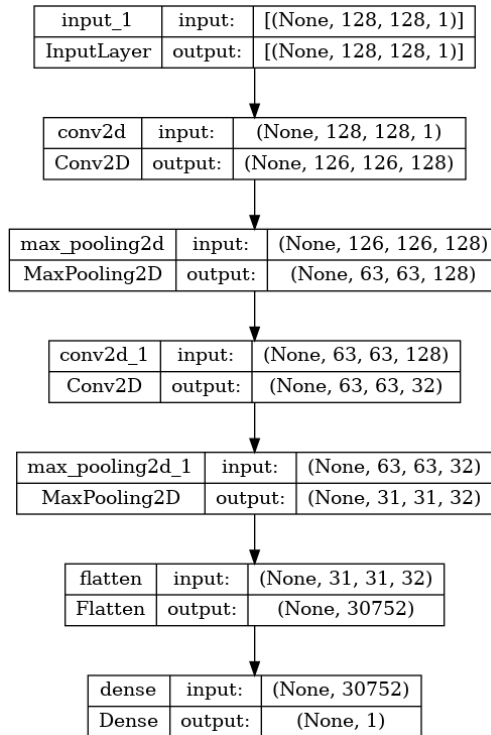


Figura 12: Estructura del modelo

El modelo está basado en las capas anteriormente explicadas.

La primera capa es la entrada, una imagen de 128x128 píxeles en un canal de color.

A continuación dos bloques de capa convolucional + *MaxPooling*. Para terminar, una capa “Flatten” que colapsa las matrices de neuronas en una única línea, y la capa “Dense”, que conecta todas las neuronas anteriores con una única neurona de salida.

El modelo consta de 36161 parámetros entrenables, 1280 de la primera capa convolucional, 4128 de la segunda capa convolucional y 30753 de la capa altamente conectada ‘Dense’.

```

## CAPAS DEL MODELO
input_shape = (128,128)+(1,) #Imagen 128x128 píxeles en un canal de color
inputs = keras.Input(shape=input_shape)

#Primer bloque convolucional
x = layers.Conv2D(filters = 128 ,kernel_size = 3)(inputs)
x = layers.MaxPooling2D(pool_size=(2,2))(x)

#Segundo bloque convolucional
x = layers.Conv2D(filters = 32 , kernel_size = 1)(x)
x = layers.MaxPooling2D(pool_size=(2,2))(x)

x = layers.Flatten()(x)
outputs = layers.Dense( 1, activation = "sigmoid")(x) #Una neurona de salida

model = keras.Model(inputs, outputs)
    
```

Figura 13: Implementación del modelo en python.

En la imagen 13 se puede ver la implementación en python del modelo usando keras, donde aparecen los parámetros usados en cada capa. La primera

capa convolucional utiliza 128 filtros de 3x3, la segunda utiliza 32 filtros de 1x1. Por otro lado, las dos capas 'pooling' seleccionan el valor máximo en grupos de 2x2, en ambos casos, dividiendo las neuronas anteriores entre 4. Por último, la capa 'dense', conecta todas las neuronas de la capa anterior con una única neurona de salida. Anteriormente, se ha explicado que en redes neuronales para clasificación, la última capa consta de tantas neuronas como clases se quiere dividir el *dataset*. Sin embargo, en el caso de elegir entre dos categorías, como es el nuestro, es preferible utilizar una única neurona donde el valor se interpreta como la probabilidad de pertenecer a la categoría positiva.

#### 4.2.5. Entrenamiento

El modelo ha sido entrenado con la función fit de keras, separando el conjunto de datos según se muestra a continuación y aunque en posteriores apartados se utilicen *datasets* con distinta cantidad de datos, en todos los casos la separación sigue los mismos porcentajes.

- **Entrenamiento 75 %:** Utilizado para el entrenamiento optimizando los parámetros de la red.
- **Validación 10 %:** Se utiliza en el proceso de entrenamiento, en cada ciclo tras ajustar las neuronas según el conjunto de entrenamiento, se evalúa sobre la red este conjunto para valorar los cambios efectuados. Sirve principalmente para evitar el "overfitting" (que la red se sobreentrene en los datos recibidos en el entrenamiento y este perdiendo precisión en cualquier otro conjunto).
- **Evaluación 15 %:** Tras el entrenamiento completo de la red. Este conjunto se usa para evaluar la precisión del modelo. Comparando las predicciones sobre este conjunto con los valores reales.

Los entrenamientos se han realizado con 25 iteraciones. En cada iteración, la función de entrenamiento, pasa por la red todos los datos de entrenamiento ajustando los pesos de las neuronas, así como, el conjunto de validación. En todos los entrenamientos se ha monitorizado la precisión del modelo y la función de pérdida a lo largo de las iteraciones, un ejemplo se puede ver en la imagen 17. La función de pérdida o coste, que debe ser optimizada ha sido '*BinaryCrossentropy*' la cual es usada para la clasificación binaria.

#### 4.2.6. Evaluación

En la evaluación, se obtienen las matrices de confusión. Estas son matrices que muestran el número de datos en las distintas categorías, si son fuentes o no, y cuál ha sido su predicción. Como en nuestro problema tenemos dos categorías, la matriz de confusión es 2x2. El significado de cada celda se puede ver en la figura 14.

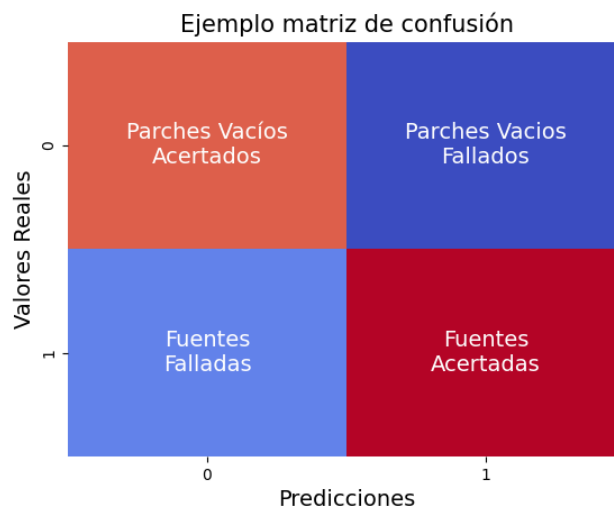


Figura 14: Ejemplo de las matrices de confusión utilizadas

#### 4.2.7. Comparaciones

Se han realizado una serie de comparaciones con el fin de comprender mejor distintos aspectos tanto de la red neuronal como de los MFCM.

Estas consisten en modificar los valores de distintos aspectos de los datos. Generando varios *datasets* distintos, que se usarán para entrenar modelos completamente independientes. Por tanto, estos modelos comparten la estructura de capas, pero su entrenamiento ha sido distinto, obteniendo cada uno distintas predicciones. Los resultados de todos ellos se graficarán juntos para observar posibles tendencias.

**Tamaño del parche de entrada** Se ha probado comparar el comportamiento de la red neuronal, si esta recibe imágenes de distinto tamaño. En cada caso se ha realizado un entrenamiento independiente del mismo modelo. El número de píxeles de las imágenes es fijo en todos los casos (128x128)

siendo lo que cambiamos, el tamaño de cielo mostrado en cada imagen, llamaremos tamaño del parche a los grados cuadrados mostrados en la imagen. Se respeta la relación:

$$\text{Tamaño del parche } [^\circ] = \text{Píxeles de la imagen} \cdot \text{Resolución del parche } (^\circ/\text{pix})$$

donde los píxeles de la imagen son 128 y la resolución del parche son los grados celestes por cada píxel.

Los tamaños usados están mostrados en la tabla 1. Estos tamaños han sido elegidos de tal forma que la resolución, en minutos por píxel ( $'/\text{pix}$ ) tengan como máximo dos cifras decimales. Para el caso de 12.80 grados cuadrados, no se ha conseguido colocar ningún negativo, por tanto, este como cualquier tamaño superior quedan descartados.

A tamaños más grandes, la ventaja reside en que la red recibe más contexto de lo rodeado a la fuente. A tamaños más pequeños, la ventaja está en dedicar más píxeles a la propia fuente, es decir, se está ganando en definición. Además, al ser los parches más pequeños es posible generar más cantidad de negativos válidos, aumentando el tamaño del *dataset*.

En la figura 15 se puede ver la diferencia entre los distintos tamaños de parches, y en la figura 16 como quedan distribuidos los negativos en los distintos casos.

Tamaño [grados cuadrados]	Resolución ( $'/\text{pix}$ )	Núm. Negativos	Relación F/N
3.200	1.5	3872	0.4
5.888	2.76	576	2.7
6.400	3.00	388	4.02
7.400	3.47	197	7.92
9.820	4.65	24	65
12.80	6.00	0	-

Tabla 1: La resolución, el número de negativos generados y la relación entre el número de fuentes y el número de negativos (F/N) para distintos tamaños de parche.

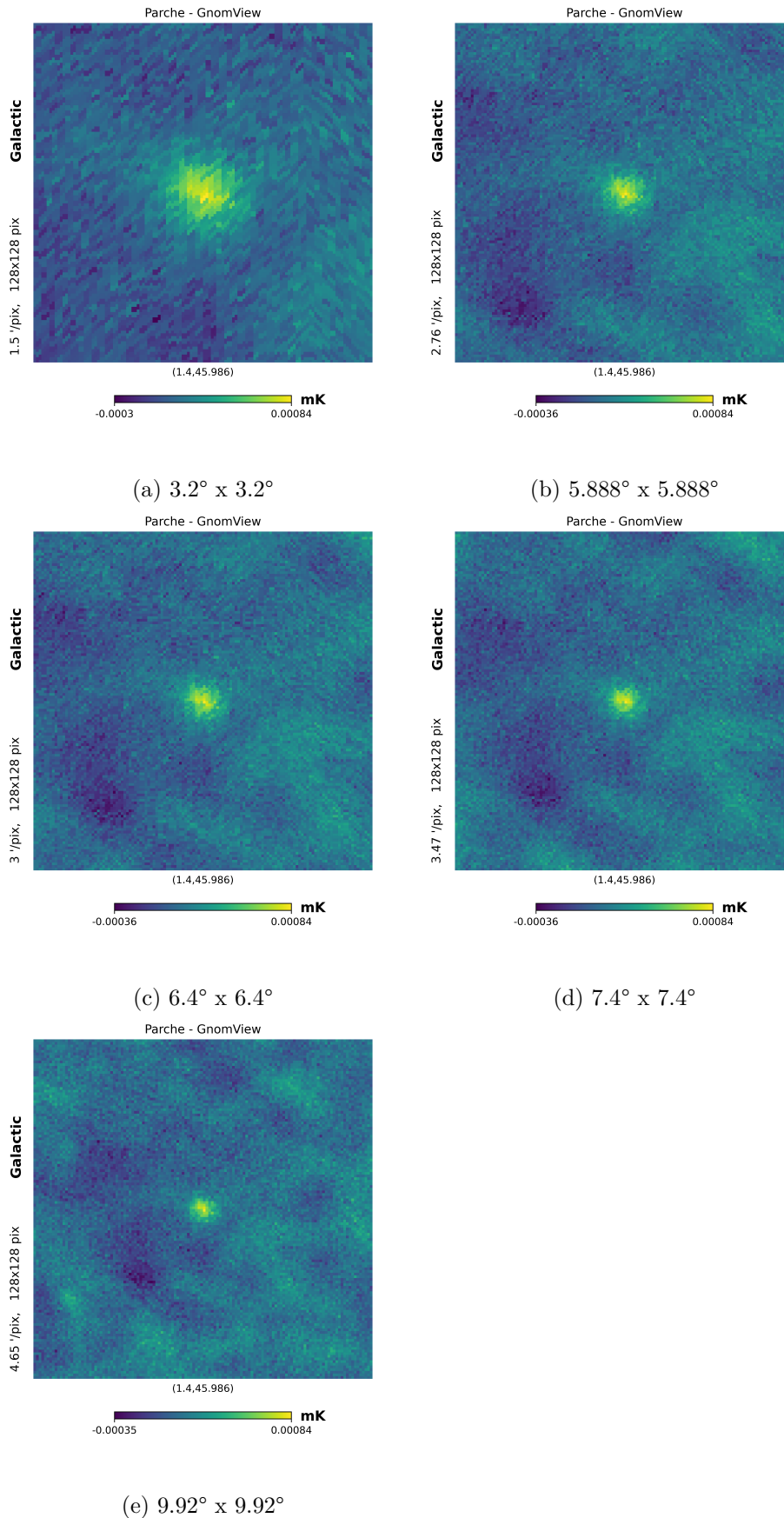
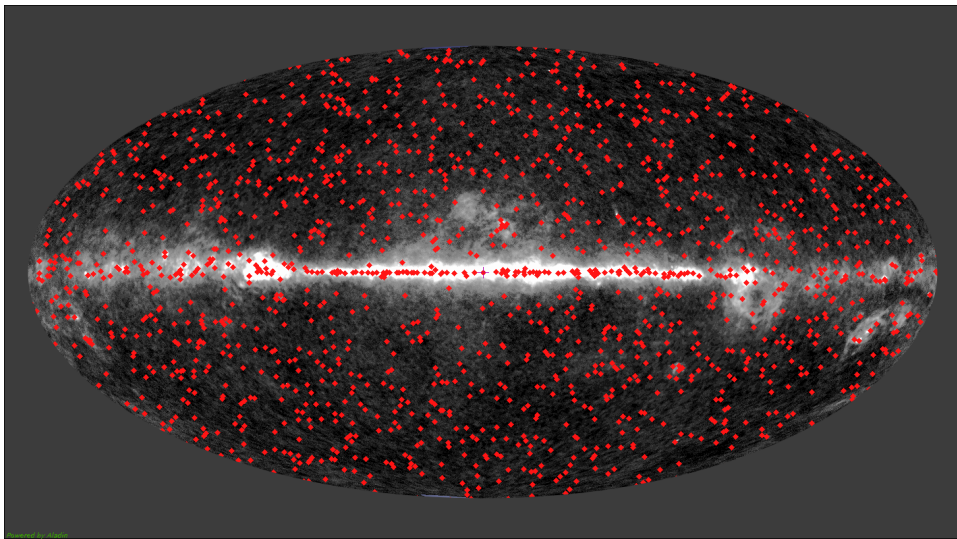
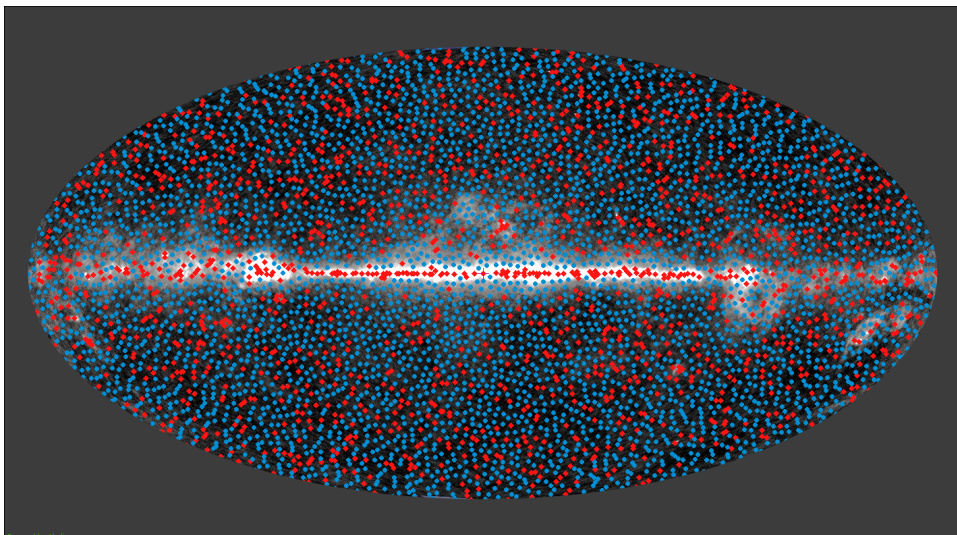


Figura 15: Ejemplo de la imagen de una fuente utilizando los distintos tamaños de parche.

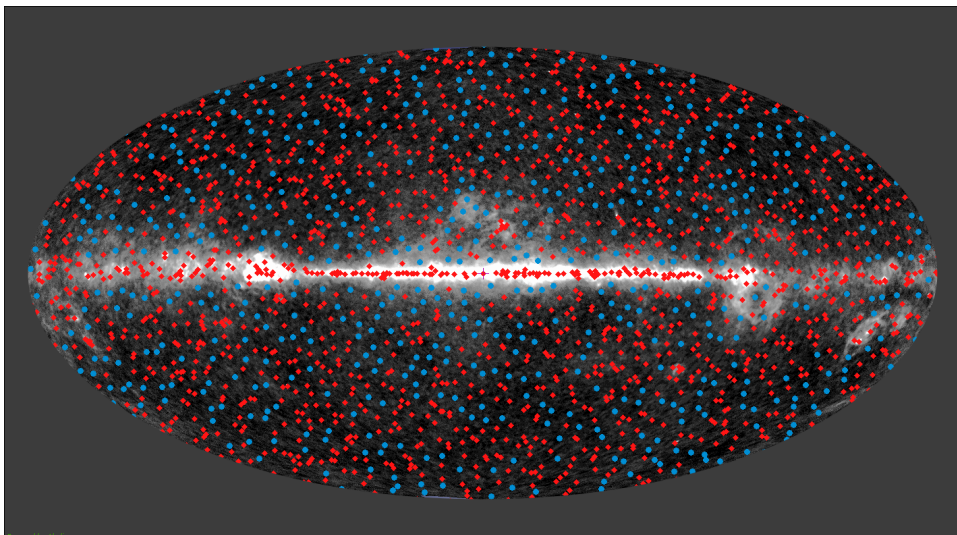




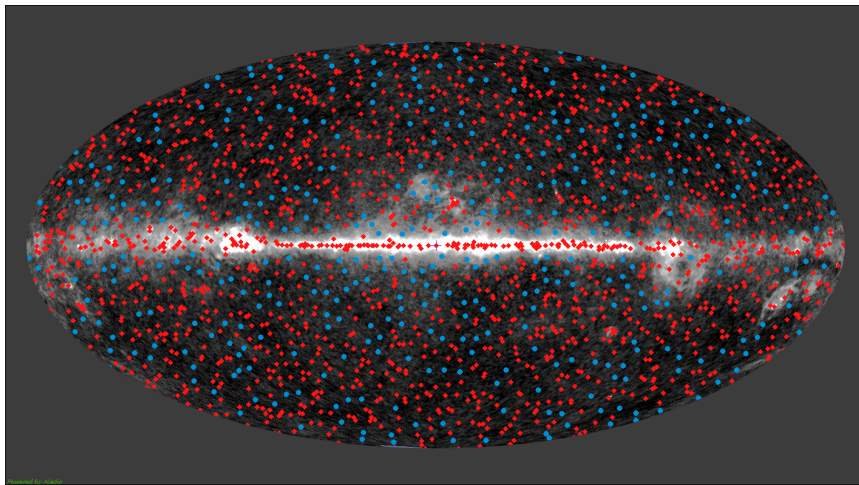
(a) Centros del catálogo PCCS2



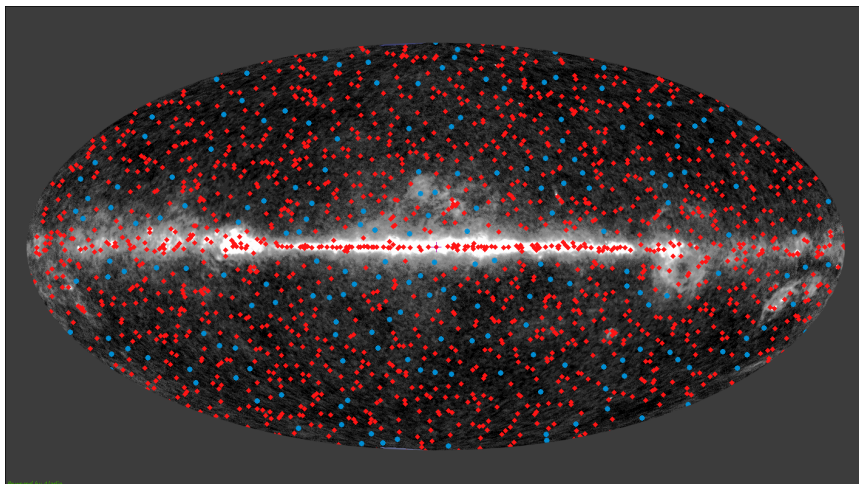
(b)  $3.2^\circ \times 3.2^\circ$



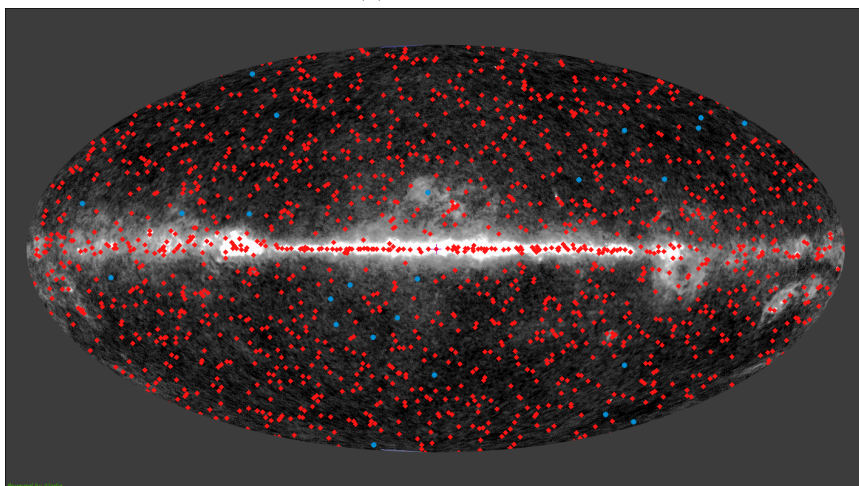
(c)  $5.888^\circ \times 5.888^\circ$



(d)  $6.4^\circ \times 6.4^\circ$



(e)  $7.4^\circ \times 7.4^\circ$



(f)  $9.92^\circ \times 9.92^\circ$

Figura 16: Distribución de los centros sobre el FCM para los distintos tamaños. (a) Muestra las fuentes del catálogo PCCS2, (b, c, d, e, f) Muestra las fuentes del catálogo PCCS2 en rojo y en azul los centros negativos respectivos de cada tamaño.



### Umbrales mínimos de latitud.

La segunda comparación que se ha llevado a cabo consiste en delimitar la latitud que pueden tener las fuentes respecto a la galaxia. El disco galáctico donde estamos internos genera la mayor cantidad de interferencias con el fondo cósmico de microondas, allí se encontrarán la mayor cantidad de fuentes y la mayor cantidad de ruido. Por ello es interesante ver cómo se comporta nuestro modelo cuando salimos del plano galáctico.

Para esto se ha separado el *dataset* utilizando distintos umbrales mínimos en la latitud. En el primer caso, se utilizan todas las fuentes. En el segundo, con un umbral de  $\pm 5$ , se eliminan cualquier fuente que este a menos de  $5^\circ$  del plano galáctico, es decir, cualquier fuente o negativo en el intervalo  $(-5,5)^\circ$ . Se hace lo mismo con el resto de umbrales.

Una vez creado los distintos *dataset*, se han realizado entrenamientos independientes con cada uno. Aclarar que el tamaño de parche utilizado es de  $7.4^\circ \times 7.4^\circ$ .

El número de fuentes y negativos en los distintos casos se muestra en la tabla 2 y como es de esperar según nos acercamos a los polos, la cantidad de datos va disminuyendo.

Corte de Latitud [ $^\circ$ ]	Núm. Fuentes	Porcentaje de Fuentes	Núm. Negativos	Total
0	1560	100 %	197	1757
$\pm 5$	1320	85 %	197	1517
$\pm 10$	1207	77 %	167	1374
$\pm 15$	1090	70 %	151	1241
$\pm 20$	960	62 %	133	1093
$\pm 30$	745	48 %	105	850
$\pm 45$	422	27 %	61	483
$\pm 60$	205	13 %	27	232

Tabla 2: Número de fuentes y negativos según el umbral de latitud.

#### 4.2.8. Estudio de los límites del modelo evaluándolo sobre un catálogo más amplio.

Con el objetivo de validar el modelo obtenido y, además, limitar cuál es su rango de utilidad se ha obtenido la precisión sobre nuevas fuentes, en concreto sobre otro catálogo de Planck de fuentes compactas. El catálogo usado ha sido ‘2015 Planck multi-frequency catalogue of non-thermal sources’ [20], el cual, contiene fuentes de frecuencias entre 30 y 857 GHz. Hemos seleccionado aquellas fuentes que tienen contribución a 30GHz, basándonos en la S/R a 30GHz asociada a cada fuente.

Para poder obtener el rango de utilidad del modelo, no se ha calculado la precisión sobre todo el catálogo simultáneamente, en su lugar, se han formado grupos según la S/R de cada fuente. Ahora sí, sobre cada grupo, se ha pedido al modelo que determine cuáles observaciones considera fuentes, a diferencia de antes todas son fuentes catalogadas, por lo que esperamos que responda en todos los casos que existe fuente. Los distintos grupos formados se pueden ver en la tabla 3, los intervalos de S/R y la cantidad de fuentes en cada caso.

Intervalo Señal/Ruido	Núm. Fuentes
<1	10296
[1,2)	9467
[2,3)	3434
[3,4)	1084
[4,5)	536
[5,6)	337
[6,7)	262
[7,8)	210
[8,9)	122
[9,10)	101
[10,15)	291
[15,20)	112
[20,25)	62
[25,35)	56
[35,50)	37
>50	20

Tabla 3: Número de fuentes para los distintos intervalos de S/R

#### 4.2.9. Estudio de la dependencia con la latitud utilizando nuevos datos.

Para reforzar el estudio sobre el efecto del plano galáctico, empezado en el apartado 4.2.7, se ha efectuado otra prueba sobre el nuevo catálogo.

En primer lugar, al igual que en el apartado anterior, se han creado grupos según la S/R, aunque, en este caso, no se han utilizado todas las fuentes, sino que se han usado los 3 intervalos siguientes: (3,4], (4,5] y (>5), dejando fuera las fuentes por debajo de 3 de S/R.

Después de esto, cada grupo se ha vuelto a subdividir dependiendo de la latitud haciendo uso del mismo criterio del apartado 4.2.7, es decir, se han creado 8 conjuntos donde, cada vez, se ha aplicado un umbral mínimo en la latitud más alto.

Cada uno de estos subgrupos, ha sido evaluado por el modelo ya entrenado en el apartado 4.2.7 correspondiente a ese umbral de latitud.

### 4.3. Resultados del TFG

#### 4.3.1. Entrenamiento y Evaluación sobre el catálogo PCCS2.

El catálogo cuenta con 1560 fuentes, utilizando un tamaño de parche de  $7.4^\circ \times 7.4^\circ$ , contamos con 197 negativos. Utilizando los porcentajes anteriormente explicados:

- El conjunto de entrenamiento cuenta con 1318 imágenes.
- El conjunto de validación cuenta con 175 imágenes.
- Para la evaluación, el conjunto de test contiene 264 imágenes.

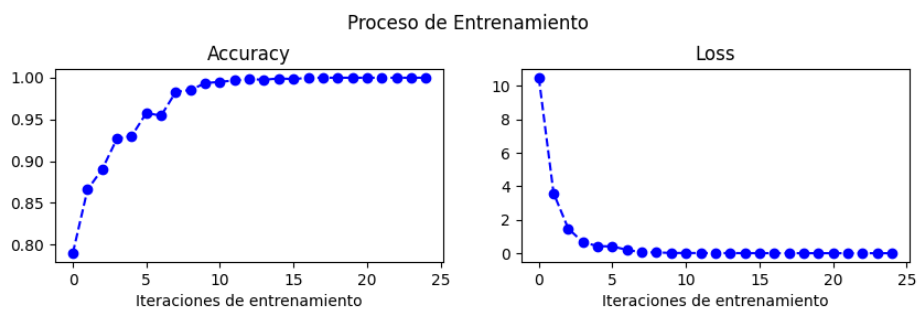


Figura 17: Variación de la precisión y la función de pérdida durante el proceso de entrenamiento sobre el catálogo PCCS2.

La evolución de la precisión del modelo y la función de pérdida a lo largo del entrenamiento se puede ver en la imagen 17.

La precisión conseguida sobre el conjunto de test es de 95.83% como se puede ver en la imagen 18.

La matriz de confusión nos permite ver el comportamiento del modelo y como podemos observar ha detectado correctamente el 98% de las fuentes. Por el contrario, ha acertado el 73% de los negativos.

Si bien esto nos está indicando que el modelo tiene predilección por catalogar una entrada como fuente, es algo esperado teniendo en cuenta que ha sido entrenada con 1170 fuentes y únicamente con 148 negativos.

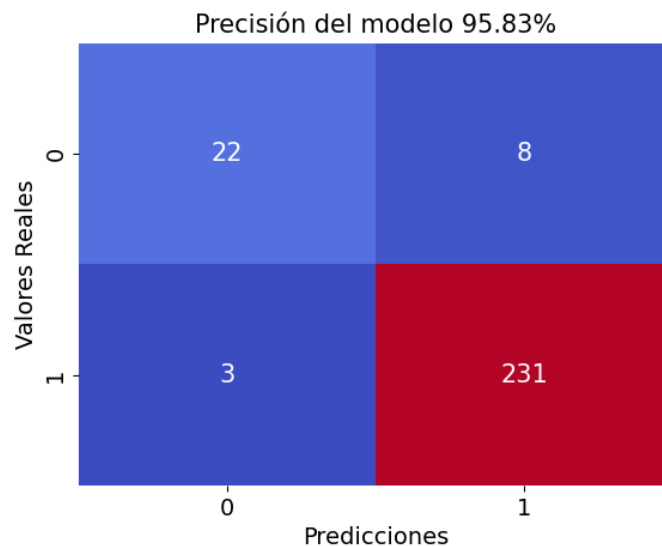


Figura 18: Matriz de confusión para el modelo entrenado sobre PCCS2.

Se ha decidido dibujar sobre el mapa los casos de error. De esta forma podremos observar si existe algún patrón en los errores. En nuestro caso, como se puede ver en la imagen 20, no parece existir ningún patrón claro. Las fuentes falladas se pueden ver en la figura 19, donde parece evidente la presencia de una fuente en su centro, para el caso superior derecho y el inferior. Observando el valor de S/R de las fuentes falladas tenemos: 11.63, 10.15 y 15.8. Por lo tanto, tampoco son fuentes con S/R especialmente bajas.

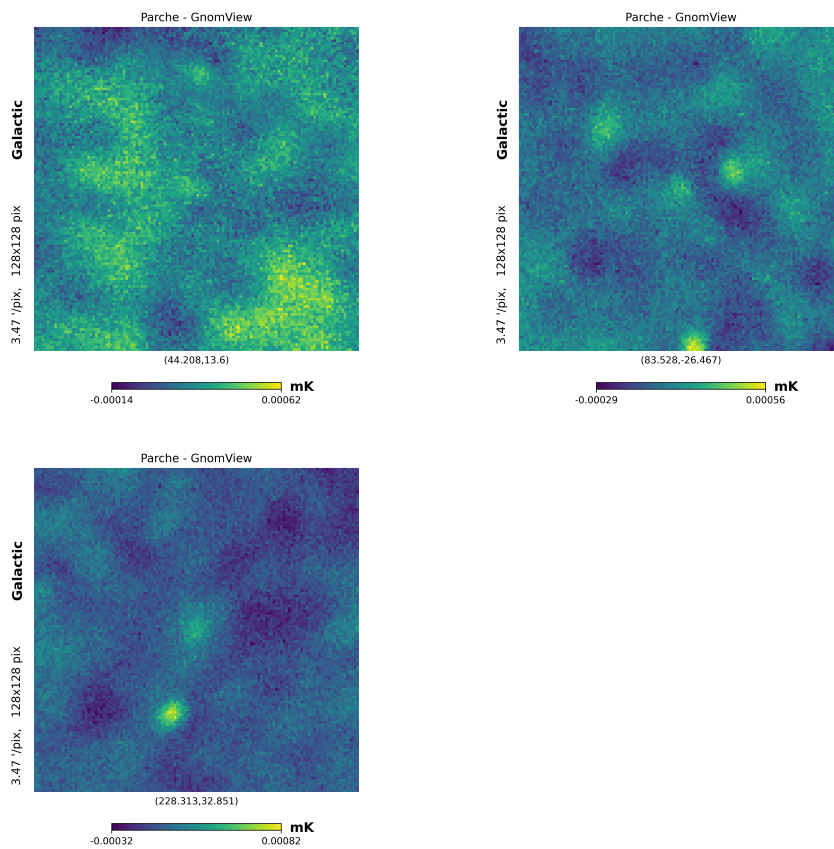


Figura 19: Parches para las fuentes falladas en la evaluación. Las S/R son: 11.63 (superior izquierda), 10.15 (superior derecha), 15.8 (inferior).

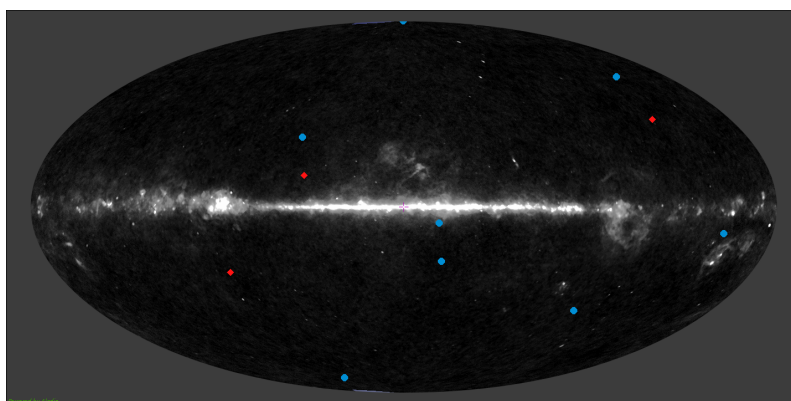


Figura 20: Casos de error en la evaluación. En rojo las fuentes falladas, en azul los negativos fallados

### 4.3.2. Distintos tamaños de parche.

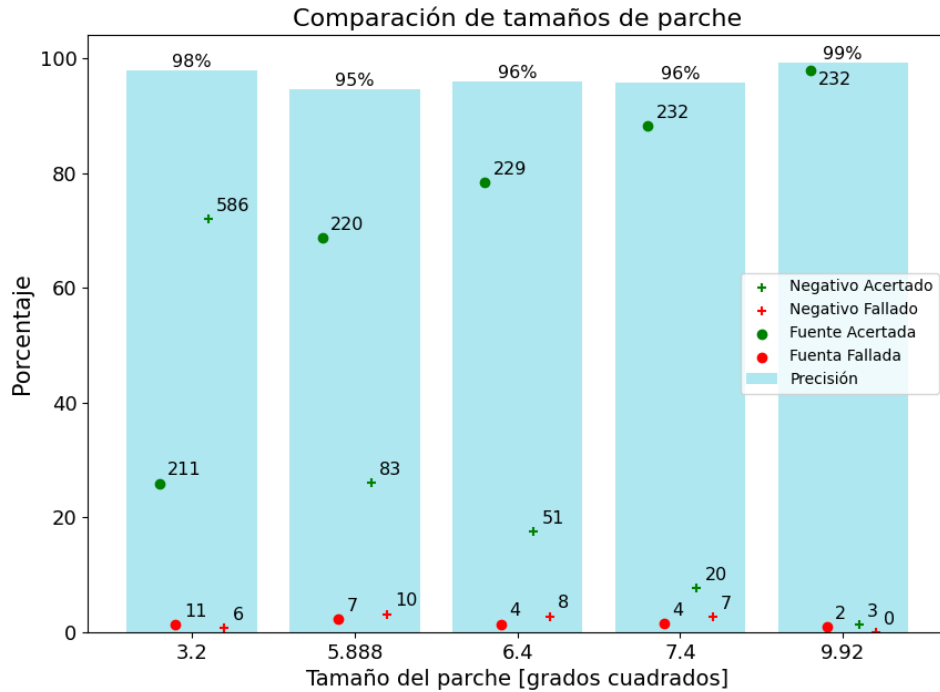


Figura 21: Resultados del entrenamiento de modelos con distintos tamaños de parche. Las barras verticales representan la precisión obtenida, mientras que, los puntos y cruces representan la matriz de confusión correspondiente a cada caso. La altura de los símbolos representa su porcentaje dentro del total de la columna

Los resultados obtenidos se muestran en la figura 21.

Las matrices de confusión individuales de cada tamaño se encuentran en el anexo 8.1.

Dado que la precisión no sigue ningún patrón concreto, aumentando en ambos extremos. Utilizando las matrices de confusión, se ha calculado el porcentaje de éxito únicamente al catalogar las fuentes, esto se muestra en la tabla 4.

Tamaño [grados cuadrados]	Precisión Total	Precisión en la detección de fuentes
3.200	97.91	95.04
5.888	94.64	96.91
6.400	95.89	98.28
7.400	95.82	98.30
9.920	99.16	99.14

Tabla 4: Precisión en la detección de fuentes utilizando distintos tamaños de parche

Como se puede ver, la detección de fuentes mejora al aumentar el tamaño del parche. Pero esto podría no estar ligado a un mejor rendimiento del modelo. Recordando la tabla 1, para el caso de  $9.92^\circ \times 9.92^\circ$  se ha entrenado con 65 veces más fuentes que negativos. Un modelo así tendrá tendencia a dar falsos positivos.

Teniendo en cuenta los resultados y a pesar de la tendencia en la detección, parece conveniente mantener un tamaño de parche de  $7.4^\circ \times 7.4^\circ$  correspondiente a la resolución de  $3.47$  ( $'/\text{pix}$ ), ya que es la utilizada por el propio mapa de Planck.

### 4.3.3. Umbrales mínimos de latitud

Los resultados obtenidos se muestran en la figura 22.

Las matrices de confusión individuales de cada tamaño se encuentran en el anexo 8.1.

Los resultados obtenidos son, por un lado, satisfactorios, puesto que la estructura del modelo seleccionada consigue buenos resultados en las distintas situaciones, pero, por otro lado, estos resultados no nos permiten distinguir si existe más o menos complejidad según nos alejamos del plano galáctico.

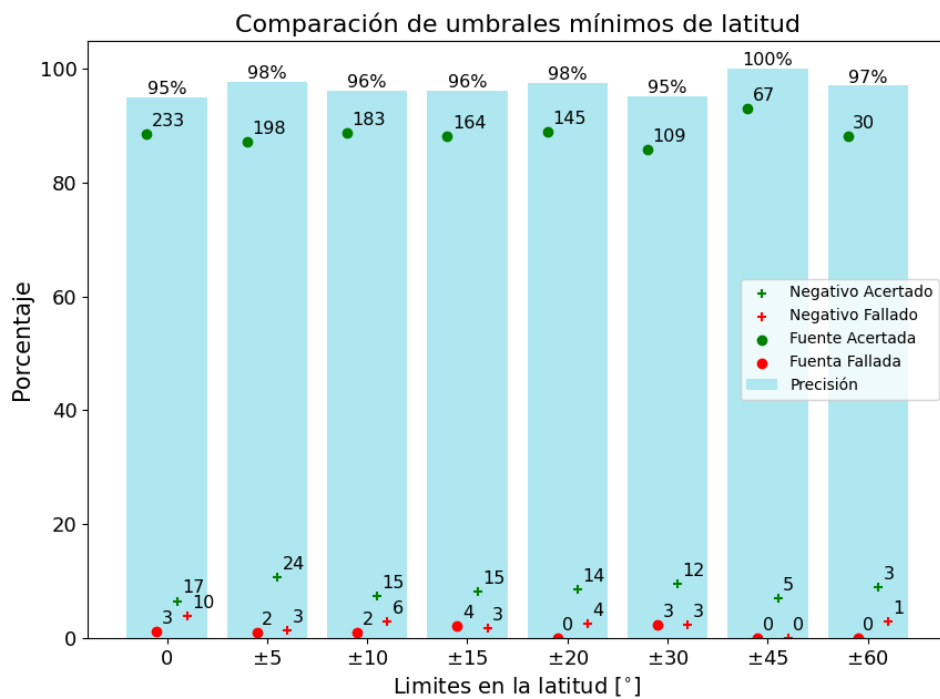


Figura 22: Resultados del modelo limitando la latitud de los datos. Las barras verticales representan la precisión obtenida, mientras que, los puntos y cruces representan la matriz de confusión correspondiente a cada caso. La altura de los símbolos representa su porcentaje dentro del total de la columna



#### 4.3.4. Resultados de la dependencia con la S/R sobre nuevos datos

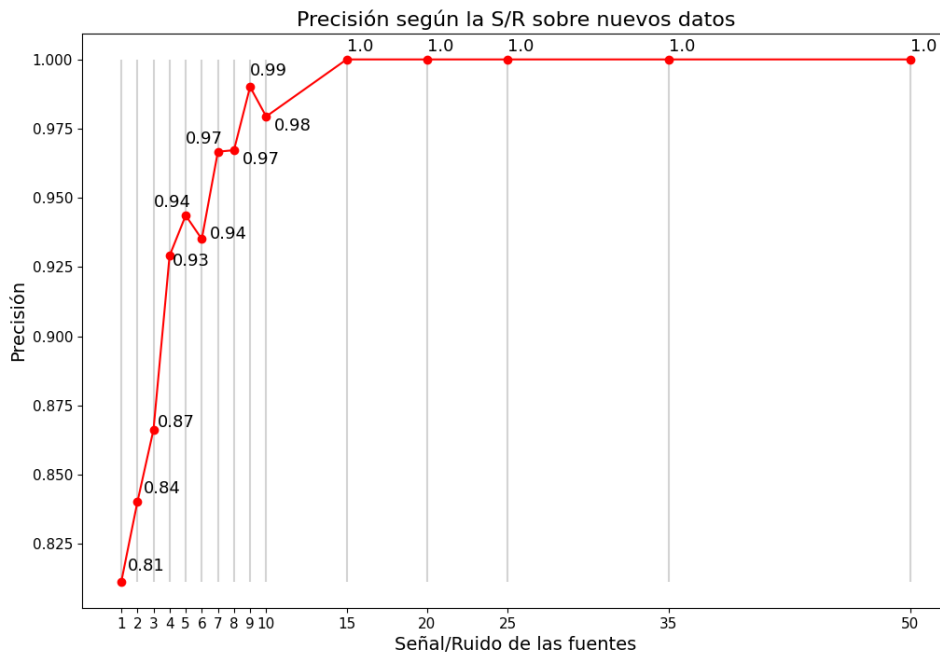


Figura 23: Precisión sobre nuevos datos según la S/R de las fuentes.

Los resultados están mostrados en la imagen 23. La tendencia de la gráfica es la esperable, la precisión crece rápidamente según aumenta la relación de S/R.

Existen pequeñas caídas entre algunos valores. Para el caso más pronunciado, al pasar de (9,10] a (10,15], se podría deber al aumento del número de fuentes, 101 en el primer caso y 291 en el segundo, aunque claro estas observaciones siguen siendo de mejor calidad. En el otro caso notable al pasar del intervalo (5,6] a (6,7], el segundo conjunto contiene menos imágenes como es lo esperado. Se ha probado a dibujar ambos conjuntos de fuentes, pero como se ve en la imagen 24 no se observa ningún patrón. Por tanto, no se ha encontrado una explicación clara para esto más allá de que se trate de un defecto del entrenamiento.

Independientemente de esto, se puede ver como el modelo acierta todos los casos de S/R por encima de 15, estas suponen 287 fuentes (tabla 3). Luego, con un 3% de error tenemos las fuentes entre 7 y 8 de S/R. Por tanto,

podemos considerar que el modelo es muy fiable para observaciones por encima de 7 S/R, consiguiendo más del 97 % de precisión en 1011 fuentes. Fijándonos en la gráfica podemos decir que el modelo deja de ser fiable aproximadamente a partir de 4 o 5 de S/R, con un 93 % y 94 % de precisión respectivamente. La caída considerable para fuentes con menos de 4 de S/R tiene sentido debido a que este, era el umbral aplicado para el catálogo PCSS2 desde el cual ha entrenado la red.

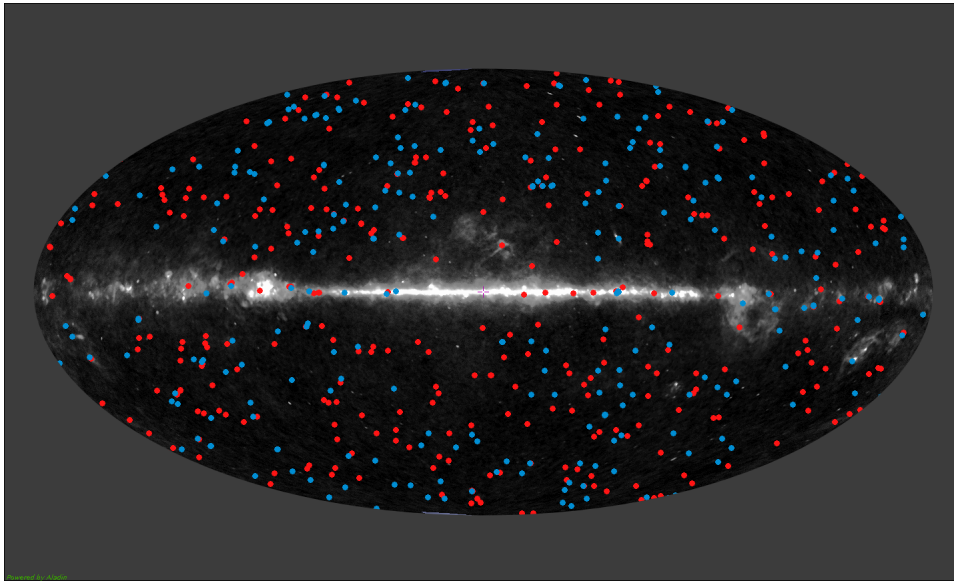


Figura 24: Distribución de centros del nuevo catálogo para los intervalos de S/R de  $(5,6]$  en rojo y  $(6,7]$  en azul

#### 4.3.5. Resultados de la dependencia de la latitud sobre nuevos datos

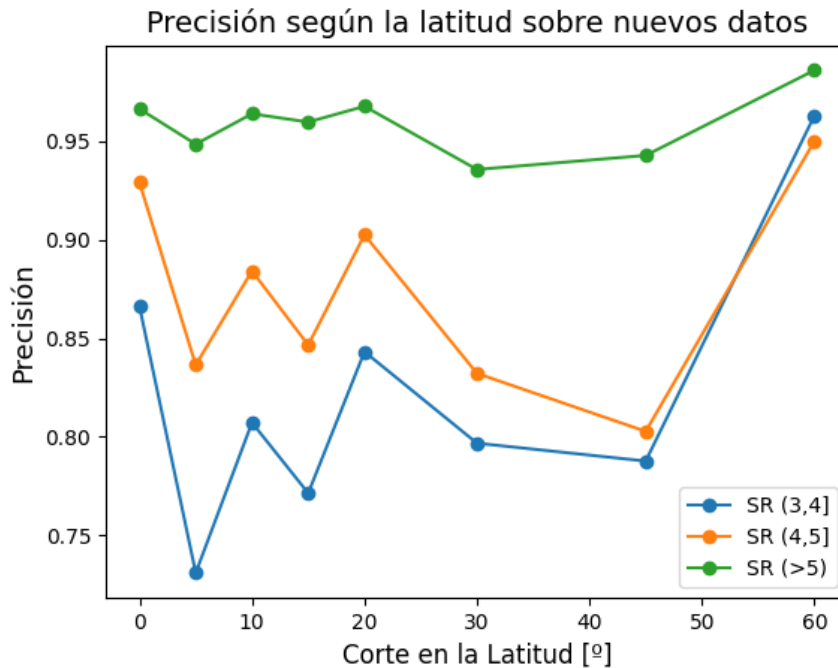


Figura 25: Precisión en función de la latitud de los nuevos datos, separados en 3 intervalos de S/R.

Estos nuevos resultados nos otorgan información útil sobre el comportamiento en función de la latitud. En primer lugar, observar que, los tres conjuntos describen el mismo patrón, pero cuanto menor es la S/R más sufren las variaciones por la latitud.

Se observa una caída importante entre los 0 y  $\pm 5$  grados. La diferencia entre estos dos conjuntos son todas las fuentes pertenecientes al hilo central del plano galáctico. En esta zona las fuentes están muy juntas entre sí superponiéndose en los parches. Además, las provenientes de la propia galaxia serán muy brillantes al encontrarse cerca de nosotros. También añadir que no tenemos en esta franja parches negativos que pudieran servir a la RN de contraste. Por todo esto, es razonable pensar, que aunque haya mucho ruido por esta zona, este tipo de fuentes son fáciles de detectar por la RN. Al eliminar estas fuentes poniendo el corte en  $\pm 5^\circ$  perdemos precisión por

eliminar este tipo de fuente.

Después de este punto se va ganando precisión debido a que el cielo es cada vez más limpio de ruido.

Se sufre otra caída en la precisión sobre los  $\pm 30^\circ$ , este es el corte utilizado en el catálogo PCCS2 para marcar las fuentes consideradas fuera del plano galáctico para el caso de bajas frecuencias, como es nuestro caso. Teniendo en cuenta esto, observamos que al final del plano galáctico perdemos precisión, pero esta aumenta enormemente cuando estamos completamente fuera del plano galáctico y el cielo es mucho más limpio. Aclarar que situar una fuente fuera del plano galáctico, no nos asegura que esta provenga de fuera de nuestra galaxia.

#### 4.4. Herramientas empleadas

- Python: El trabajo se ha creado utilizando este lenguaje de programación. Se han utilizado bibliotecas comunes como numpy o matplotlib y otras más especializadas como astropy, healpy o keras.
- Aladin [21]: Programa para la visualización de mapas celestes. Contiene las proyecciones mayormente utilizadas, así como distintos tipos de coordenadas. Se ha utilizado para visualizar el mapa de Planck, superponiendo fuentes o negativos en distintas situaciones para realizar comprobaciones. Los mapas celestes de este trabajo han sido exportados utilizando este programa, a excepción de aquellos que tengan citada una fuente externa.
- Healpix [22]: Sistema de granulado/píxelado de superficies esféricas. Dado que trabajamos con una esfera, la forma de seccionar la superficie en píxeles no es trivial, y healpix es el estándar más utilizado. En concreto, se ha utilizado el módulo de python healpy, basado en healpix y matplotlib, para obtener los parches de cada fuente del tamaño y resolución deseados.
- Keras: Para todos los procesos concernientes al modelo de inteligencia artificial se ha utilizado la biblioteca de keras, así como algunas herramientas de TensorFlow. Por tanto, se ha utilizado para implementar el modelo, hacer las predicciones, evaluarlo, entrenarlo, etc.

## 5. DISCUSIÓN

### 5.1. Utilización de Imágenes

En todo el trabajo los datos de los parches se convierten a imágenes en formato PNG. Esto se ha realizado así porque al principio del trabajo encontramos varios problemas para pasar directamente los arrays de datos a la red. Al principio hubo problemas con el input que debía recibir la RN y los arrays, pero tras solucionar esto la red seguía sin entrenar correctamente, en algunos casos errores en otros devolvía siempre el mismo valor. Al pasar los datos a imagen se pudieron cargar utilizando una función nativa de keras lo que solucionó los problemas.

El problema de utilizar imágenes reside en que los datos se normalizan para adaptarlos al formato. Esto puede causar que resalten más las diferencias relativas dentro de un mismo parche lo cual puede ser útil para la red, pero probablemente sea un inconveniente al perder la información de la densidad de flujo absoluta para las observaciones.

### 5.2. Limitación Computacional

El entrenamiento de una red neuronal debido al número de cálculos es, sin duda, un proceso muy costoso desde el punto de vista computacional.

Este trabajo se ha realizado desde un portátil de uso personal y esta limitación ha sido clave a la hora de utilizar un modelo con pocas capas. Cuando se intentó probar estructuras de capas mucho más complejas, los tiempos de entrenamientos subían a varias horas. Además, los resultados preliminares solían ser entrenamientos erróneos, que sin duda estaban dados por una mala configuración en las capas o en el input, pero que, debido a los tiempos de entrenamiento, se habría malgastado demasiado tiempo en hacerlos funcionar correctamente, para obtener un modelo que seguirá funcionando demasiado lentamente para el resto de pruebas.

### 5.3. Limitación en el método

El enfoque de este proyecto, de utilizar datos reales para entrenar la red neuronal y evaluarla, ya ha sido justificado en apartados anteriores y sin menoscabar lo ya expuesto, cabe remarcar algunas limitaciones.

A la hora de intentar mejorar una RN, nuestras opciones principalmente

son: cambiar su estructura y reentrenarla, o cambiar los datos que recibe. En nuestro contexto, se pueden utilizar resultados más o menos fiables, pero inevitablemente a la hora de aumentar la cantidad de datos estaremos ocupando un espacio que antes posiblemente contenía negativos.

Esto es un proceso inevitable según mejoren la resolución de los mapas obtenidos, se podrán observar nuevas fuentes donde antes no se detectaba nada. Cualquier red neuronal necesitará ejemplos negativos porque en caso contrario siempre devolverá el mismo valor careciendo de utilidad.

Una posible solución a este problema ya ha sido tratada en este trabajo, que consistiría en reducir el tamaño del parche para colocar mayor número de negativos. Pero esto, por lo menos en las condiciones que nosotros teníamos (aumentan los negativos, pero no las fuentes), suponía perder precisión en la detección.

La solución más lógica parece ser utilizar simulaciones de igual forma que se usan para la validación de las fuentes o de los métodos de detección. Superponer a simulaciones del FCM la contaminación causada por fuentes compactas permitiendo obtener tantos datos como se desee. Si se quisiera seguir un poco la filosofía original de este trabajo, se podrían utilizar las fuentes reales y utilizar para los negativos simulaciones del FCM a los que se les superponga emisiones difusas si fuera conveniente.

## 6. CONCLUSIONES

### 6.1. Conclusiones del trabajo

Basándonos en los resultados del apartado 4.3.1, con una precisión del 95.83 % podemos concluir que se ha diseñado y entrenado una red neuronal exitosamente para la detección de fuentes compactas.

Además, se ha acotado la utilidad del modelo en función de la S/R, pudiendo concluir que tendrá un gran rendimiento para observaciones con S/R superior a 7 con una precisión superior al 97 %. Y dependiendo las necesidades podría ser utilizado hasta una S/R de 4 con una precisión superior al 93 %. Estos últimos resultados al ser obtenidos sobre un segundo catálogo nos permiten concluir que el modelo tiene un comportamiento consistente, no sufriendo de overfitting y cuya utilidad no está limitada al catálogo de entrenamiento. En general, otorgan al modelo mucha más validez.

Respecto a las características de la bóveda celeste. Podemos afirmar que existe un anillo en el intervalo  $(-5^\circ, 5^\circ)$  donde las fuentes son fácilmente detectables principalmente por su cantidad y la presencia de algunas fuentes muy brillantes. Un tramo intermedio que se ve afectado por el ruido galáctico, donde la detección se complica, pero que mejora según nos alejamos del plano galáctico. Y por último, la zona fuera del plano galáctico que mejora drásticamente la precisión según nos acercamos a los polos, donde el ruido es mínimo. En el corte de salida del plano galáctico ( $30^\circ$ ) aumenta también ligeramente la dificultad, posiblemente por la presencia de parches que en su mayoría estén libres de ruido galáctico, pero en alguna parte si contienen algo, siendo poco homogéneo y dificultando ligeramente que la fuente sobresalga.

Para un mejor estudio probablemente sea necesario introducir máscaras, pudiendo con estas separar de forma más precisa las zonas brillantes del plano galáctico. Por contra, complican el análisis y por eso no han sido utilizadas en este trabajo.

Por último, sobre el comportamiento de la R.N. ante distintos tamaños de parche, los resultados indican que si bien la precisión total no siguen ningún patrón, la precisión a la hora de evaluar las fuentes mejora al aumentar el tamaño. Pero esto parece más un efecto de la relación entre el número de fuentes y el de negativos, que una mejora real en la detección del modelo.

Por tanto, a falta de otras pruebas, como someter el nuevo catálogo a distintos tamaños de parche de forma similar a lo realizado con la latitud, no podemos afirmar un comportamiento definido.

## 6.2. Conclusiones personales

El desarrollo de este proyecto ha supuesto, por un lado, un aprendizaje en detalle sobre las redes neuronales, sobre su funcionamiento, como utilizarlas, etc. Y, por otro lado, también ha supuesto un aprendizaje sobre el fondo cósmico de microondas, sobre cómo se estudia este en las pipelines de detección de Planck, el camino que siguen los datos desde el telescopio, hasta que una observación se considera fuente compacta.

Sin duda, los conocimientos obtenidos en este proyecto me serán de utilidad en el futuro, debido a la alta implementación que se está dando de este tipo de métodos de análisis.



## 7. FUTURAS LÍNEAS DE TRABAJO

Algunas de las continuaciones lógicas de este trabajo consistirían en solucionar las limitaciones expuestas en el apartado de discusión (5).

En primer lugar, mayor capacidad computacional permitiría utilizar modelos más complejos, que si bien esto no asegura mejores resultados, sería ilógico pensar que el modelo utilizado es el mejor posible. En segundo lugar, la utilización de los datos en bruto, directamente desde los arrays, sin pasar la información a imágenes, como se ha comentado, permitiría estudiar la densidad de flujo de las observaciones, siendo útil, por ejemplo, para obtener la precisión según la densidad de flujo de las fuentes y determinar así la densidad de flujo mínimo necesario para garantizar una precisión concreta, el cual, es un valor comúnmente usado para valorar el rendimiento de los métodos de detección.

Fuera de las limitaciones encontradas, otras opciones de continuación podrían pasar por la utilización de dos o más frecuencias simultáneamente. El análisis multifrecuencia entraña otros problemas, la más obvia es que una observación puede contribuir fuertemente a una frecuencia y ser invisible en otra, pero, a pesar de la aparición de nuevos problemas, parece interesante evaluar si una red neuronal podría ser capaz de encontrar estos patrones con el suficiente número de ejemplos e iteraciones, o si, por el contrario, sería necesario fijarle reglas para obtener resultados fiables.

Por último, si imaginamos una red neuronal para la detección de fuentes, entrenada utilizando una gran cantidad de datos, probablemente simulados, se esperará que esta tenga utilidad a largo plazo, y esto pasará necesariamente por qué el modelo no pierda precisión si, por ejemplo, se le entregan datos en distintos tamaños, distinta resolución, etc. De forma que siga siendo útil para el procesamiento de datos de un nuevo satélite, o un satélite que por la razón que sea ha mejorado o empeorado la calidad de sus datos.

Como se puede ver, este es un amplio campo de estudio donde esperamos obtener potentes herramientas en un futuro próximo, las cuales ayuden con los futuros proyectos.

## Referencias

- [1] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981.
- [2] R. A. Alpher, H. Bethe, and G. Gamow. The origin of chemical elements. *Phys. Rev.*, 73:803–804, Apr 1948.
- [3] Arno A. Penzias and Robert G. Wilson. A measurement of excess antenna temperature at 4080-mc/s. *The Astrophysical Journal*, 142:419–421, 1965.
- [4] G.F. Smoot et al. Structure in the COBE Differential Microwave Radiometer First-Year Maps. , 396:L1, September 1992.
- [5] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright. The cosmic microwave background spectrum from the full coBE\* firas data set. *The Astrophysical Journal*, 473(2):576, dec 1996.
- [6] J. M. Kovac, E. M. Leitch, C. Pryke, J. E. Carlstrom, N. W. Halverson, and W. L. Holzapfel. Detection of polarization in the cosmic microwave background using DASI. , 420(6917):772–787, December 2002.
- [7] Planck Collaboration. Planck 2018 results - i. overview and the cosmological legacy of planck. *A&A*, 641:A1, 2020.
- [8] D. B. Sanders and I. F. Mirabel. Luminous Infrared Galaxies. , 34:749, January 1996.
- [9] Planck Collaboration. Planck 2015 results - xxvi. the second planck catalogue of compact sources. *A&A*, 594:A26, 2016.
- [10] S. M. Leach et al. Component separation methods for the PLANCK mission. , 491(2):597–615, November 2008.
- [11] González-Nuevo et al. The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps. *Monthly Notices of the Royal Astronomical Society*, 369(4):1603–1610, 07 2006.
- [12] Planck Collaboration. Planck 2013 results. XXVIII. The Planck Catalogue of Compact Sources. , 571:A28, November 2014.

- [13] J. M. Casas et al. Constraining the polarisation flux density and angle of point sources by training a convolutional neural network. *A&A*, 670:A76, 2023.
- [14] S. Riggi et al. Astronomical source detection in radio continuum maps with deep neural networks. *Astronomy and Computing*, 42:100682, 2023.
- [15] S. Johnston et al. Science with askap. *Exp Astron*, 22:151–273, 2008.
- [16] A. M. Hopkins et al. The askap/emu source finding data challenge. *Publications of the Astronomical Society of Australia*, 32:e037, 2015.
- [17] M. A. Garrett, J. M. Cordes, D. R. Deboer, J. L. Jonas, S. Rawlings, and R. T. Schilizzi. Square kilometre array: a concept design for phase 1, 2010.
- [18] P. Diego-Palazuelos et al. Machine learning approach to the detection of point sources in maps of the cmb temperature anisotropies. 2023.
- [19] European Space Agency. Planck legacy archive. [pla.esac.esa.int/pla](http://pla.esac.esa.int/pla).
- [20] Planck Collaboration. Planck intermediate results-liv. the planck multi-frequency catalogue of non-thermal sources. *Astronomy & Astrophysics*, 619:A94, 2018.
- [21] F. Bonnarel et al. The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. , 143:33–40, April 2000.
- [22] K. M. Górski et al. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, apr 2005.
- [23] L. F.Lanz Oca. Detección de fuentes extragalácticas en mapas de la radiación del fondo cósmico de microondas mediante técnicas multifrecuenciales [Tesis Doctoral]. *Universidad de Cantabria*.
- [24] Marcos López-Caniego Alcarria. PhD thesis, 9 2007.
- [25] Clive Dickinson. Cmb foregrounds - a brief review, 2016.
- [26] M. López-Caniego, J. González-Nuevo, D. Herranz, M. Massardi, J. Luis Sanz, G. De Zotti, L. Toffolatti, and F. Argüeso. Nonblind

---

catalog of extragalactic point sources from the wilkinson microwave anisotropy probe (wmap) first 3 year survey data. *The Astrophysical Journal Supplement Series*, 170(1):108, may 2007.

## 8. ANEXOS

### 8.1. Matrices de confusión

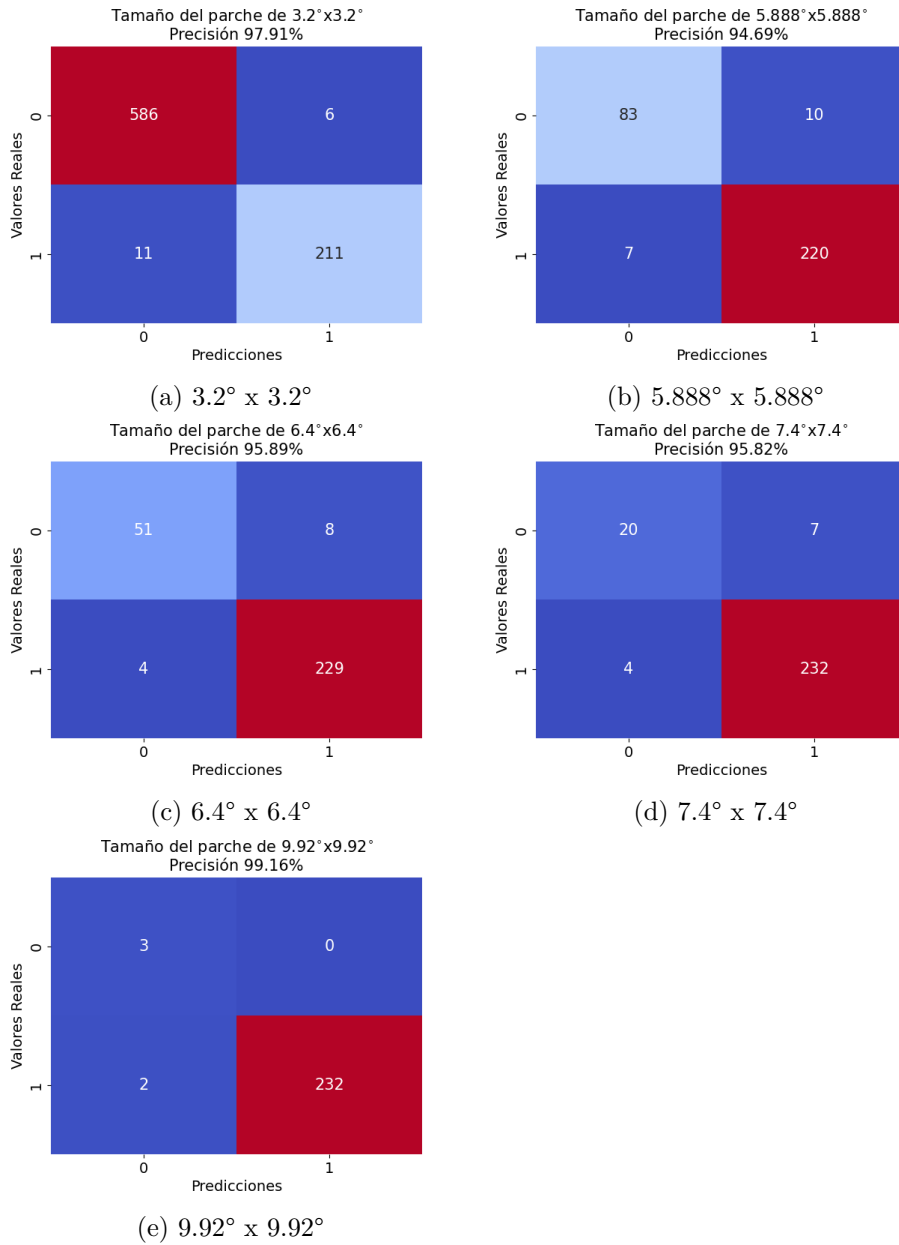


Figura 26: Matrices de confusión correspondientes a la comparación de distintos tamaños de parche.

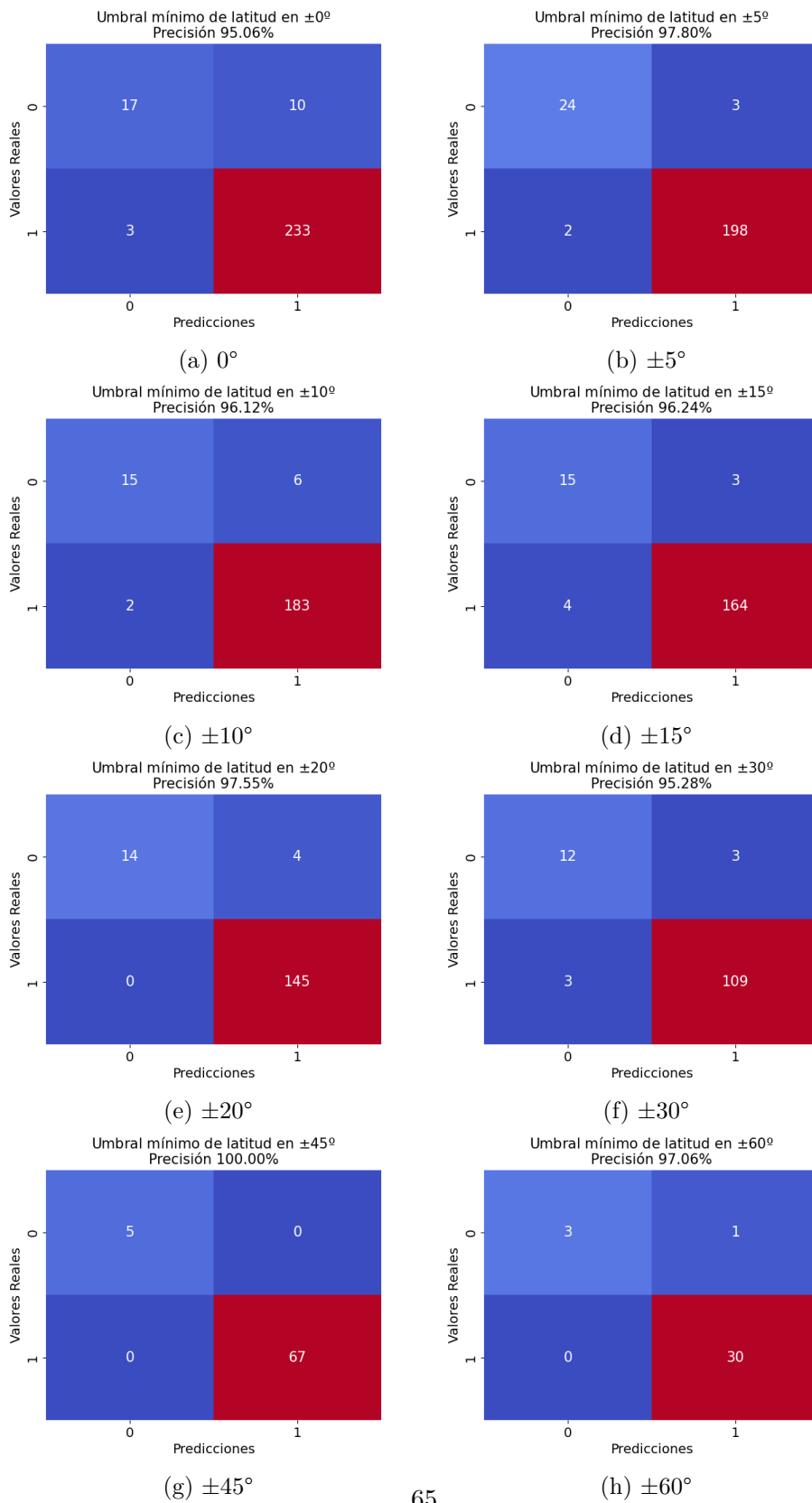


Figura 27: Matrices de confusión correspondientes al entrenamiento limitando la latitud

