

# Modelo para la Detección y Eliminación de Ecos de Tierra en Radares Meteorológicos en Banda C

Rafael Orlando Cubillos Sánchez<sup>1</sup>, Ernesto Gómez Vargas<sup>2</sup>, Oscar Danilo Montoya<sup>2</sup>.

## Resumen

La labor de identificación de fenómenos climáticos mediante la utilización de mediciones de radar se ve notablemente dificultada por la presencia de interferencias provenientes del entorno terrestre. La presencia de objetos u obstáculos que causan desviaciones en el haz de radar resulta en la generación de ecos de alta intensidad que comprometen la integridad de las mediciones y vician al proceso de detección. En el contexto de este estudio, se propone la consecución de un modelo concebido con la finalidad de discernir y catalogar las células contaminadas por ecos de tierra, procediendo posteriormente a la sustitución de los datos asociados a dichas células por información confiable y precisa. El enfoque de detección de ecos no deseados se materializa mediante un algoritmo basado en el análisis de diversas variables polarimétricas, que posibilitan la discriminación efectiva de las señales pertinentes a fenómenos climáticos genuinos de aquellas que se encuentran contaminadas por perturbaciones terrestres, permitiendo así una identificación y evaluación precisa de los eventos meteorológicos en cuestión. La tarea de identificar ecos terrestres puede servir como una herramienta eficaz para mejorar la detección de fenómenos meteorológicos, aportando beneficios significativos a un país como el nuestro, donde los eventos climáticos periódicos pueden ocasionar daños tanto económicos como sociales.

**Palabras Clave:** desorden de tierra, lógica difusa, radar meteorológico, polarimetría de radar.

## Abstract

The task of identifying climatic phenomena through radar measurements is significantly hindered by the presence of interferences originating from the terrestrial environment. The existence of objects or obstacles causing deviations in the radar beam results in the generation of high-intensity echoes that compromise the integrity of the measurements and distort the detection process. In the context of this study, the attainment of a model is proposed, conceived with the purpose of discerning and categorizing cells contaminated by ground echoes, subsequently proceeding to replace the data associated with these cells with reliable and accurate information. The approach to unwanted echo detection is realized through an algorithm based on the analysis of various polarimetric variables, which enable the effective discrimination of signals relevant to genuine climatic phenomena from those tainted by terrestrial disturbances, thus allowing for a precise identification and assessment of the meteorological events in question. The task of identifying ground echoes can serve as an effective tool to enhance the detection of meteorological phenomena, providing significant benefits to a country like ours, where periodic weather events can cause both economic and social damages.

**Keywords:** Ground clutter, fuzzy logic, meteorological radar, radar polarimetry.

**Recepción:** 04-October-2023

**Aceptación:** 25-Noviembre-2023

<sup>1</sup> Universidad Santo Tomás, Bogotá, Colombia. Dirección electrónica: [rafaelcubillos@usta.edu.co](mailto:rafaelcubillos@usta.edu.co)

<sup>2</sup> Universidad Distrital Francisco José de Caldas Bogotá-Colombia

## 1. Introducción

Colombia es uno de los países de América latina en donde más se producen desastres y de ellos, las inundaciones son superiores en número a las demás calamidades naturales que sufre nuestra nación [8]. Las constantes tragedias causadas por fenómenos climatológicos evidencian que el país no cuenta con un adecuado sistema de detección y prevención, que integrado a unas políticas de reacción y mitigación adecuadas, permitan disminuir las pérdidas económicas y de vidas humanas. Un sistema de detección y prevención apropiado debe contar con un conjunto de radares meteorológicos que funcionen eficientemente, lo cual no se puede lograr si no se tiene una metodología que permita eliminar, entre otros factores, los ecos terrestres que inducen al error. La existencia de una herramienta que elimine las señales no meteorológicas y que permita identificar los diferentes fenómenos climáticos con un mayor grado de acierto y en un menor tiempo, es indispensable para elevar las prestaciones del radar meteorológico y mejorar la detección y predicción del clima.

La detección meteorológica es el proceso de recopilación, análisis y seguimiento de datos relacionados con las condiciones atmosféricas y climáticas en la Tierra [1]. Esta labor tiene como objetivo principal el monitoreo y la comprensión de los fenómenos meteorológicos, lo que permite predecir y entender el clima y el tiempo en una determinada área geográfica. La detección meteorológica es esencial para proporcionar información precisa y oportuna sobre el tiempo, lo que tiene un impacto significativo en numerosos aspectos de la vida cotidiana y en diversas industrias, incluyendo la agricultura, la aviación, la navegación marítima, la gestión de desastres y la toma de decisiones en general [2]. En la detección meteorológica se puede utilizar el radar que es un dispositivo capaz de medir y monitorizar las condiciones meteorológicas emitiendo pulsos de energía electromagnética y registrando la cantidad de energía reflejada por objetos en la atmósfera, como gotas de lluvia, copos de nieve, partículas de polvo, etc [2].

La detección meteorológica por radar cumple varios propósitos fundamentales, entre ellos, la predicción del tiempo al ayudar a los meteorólogos a rastrear las precipitaciones, identificar patrones climáticos y anticipar eventos meteorológicos como lluvias intensas, tormentas eléctricas, nevadas y tornados. Además, desempeña un papel esencial en la emisión de alertas tempranas sobre condiciones climáticas severas, permitiendo que autoridades y población se preparen para eventos adversos y tomen medidas de seguridad [1]. También se emplea en la navegación aérea y marítima para garantizar la seguridad de aeronaves y embarcaciones al evitar condiciones peligrosas. Por último, los datos recopilados por radares meteorológicos son fundamentales en investigaciones científicas que contribuyen a una mejor comprensión de los procesos atmosféricos y el comportamiento de las precipitaciones, entre otros fenómenos meteorológicos [3].

Los principales factores que afectan la labor de detección meteorológica por radar incluyen la intensidad y tipo de precipitación, ya que la cantidad y fuerza de las precipitaciones pueden dificultar la detección de objetos detrás de ellas. Además, la ubicación geográfica del radar y la distancia a las áreas de interés influyen en su capacidad para observar fenómenos meteorológicos en tiempo real, considerando también la limitación impuesta por la curvatura de la Tierra en visibilidad a largas distancias [4]. La interferencia electromagnética de fuentes como señales de radio o transmisiones de televisión puede perturbar las señales del radar y afectar su precisión. Asimismo, obstáculos físicos como edificios, montañas y árboles tienen el potencial de bloquear o distorsionar las señales de radar, reduciendo así su capacidad para detectar con exactitud las condiciones meteorológicas en áreas específicas [5].

A los elementos que dificultan la tarea del radar en la detección de fenómenos meteorológicos hay que sumar los ecos de tierra. Los ecos de tierra son señales reflejadas que provienen de la superficie terrestre y que son captadas nuevamente por el radar. Estos ecos pueden ser problemáticos porque pueden ser confundidos con señales de fenómenos meteorológicos genuinos, lo que lleva a errores en la interpretación de los datos y a la generación de información meteorológica incorrecta [4].

Los ecos de tierra ejercen un impacto significativo sobre los datos generados por el radar, al punto de constituir una de las fuentes de error más relevantes, llegando incluso en algunas circunstancias a invalidar la totalidad de la información meteorológica recabada [6]. Estos ecos terrestres se dividen en dos categorías: en primer lugar, los ecos terrestres de propagación normal, que se originan cuando la señal del radar colisiona con obstáculos tales como montañas, edificios o árboles; y en segundo lugar, los ecos terrestres de propagación anómala, que emergen cuando el haz de radar se curva y colisiona con la superficie terrestre. En ambas instancias, se generan reflectividades de considerable magnitud, las cuales son incorrectamente interpretadas como fenómenos meteorológicos de gran intensidad [7].

Una vez identificados, los ecos de tierra pueden ser sometidos a procedimientos de eliminación o corrección en los datos, lo cual se traduce en una mejora sustancial en la precisión de las observaciones meteorológicas y en la capacidad de interpretación de los datos generados por el radar. Estos procedimientos algorítmicos se sustentan en el análisis exhaustivo de variables polarimétricas y otras características inherentes a la señal, lo que posibilita la diferenciación entre ecos de tierra y señales genuinamente meteorológicas [7].

Es innegable que la implantación de un modelo destinado a la detección y eliminación de ecos de tierra en las imágenes captadas por el radar no solo amplía la confiabilidad de los datos emitidos por el radar, sino que también implica una reducción apreciable en los lapsos de tiempo requeridos para la interpretación de esta información. La validación

del modelo se concreta mediante la evaluación del grado de concordancia entre la clasificación obtenida a través de dicho modelo y la clasificación efectuada por un perito en la materia.

La estructura de este documento se organiza de la siguiente manera: En la Sección 2 se efectúa una breve exposición teórica de la teoría del radar. El capítulo 3 versa sobre la lógica difusa. Los apartados cuarto y quinto se centran tanto en el diseño como en la implementación del modelo de detección y eliminación de ecos. La Sección 6 exhibe los resultados de la validación del clasificador implementado. El séptimo apartado describe el proceso de corrección de los datos de clasificación de manera concisa. La penúltima sección proporciona un resumen general de la estructura del clasificador desarrollado. El documento culmina con las conclusiones.

## 2. Radar y observables polarimétricos

### Radar

El término “Radar”, acrónimo de “Radio Detection and Ranging”, encuentra su origen en una aplicación de índole militar desarrollada en la década de 1930 con el propósito inicial de detectar aeronaves. No obstante, desde sus primeros avances, el radar reveló un notable potencial en la detección y observación de fenómenos meteorológicos [8]. Como se ha registrado, “los estudios del fenómeno atmosférico con radar comenzaron casi tan pronto como se usaron los primeros radares”, debido a que el análisis del clima y los ecos atmosféricos representaban un “desorden indeseable que obstaculizaba la tarea de detección de objetivos militares [5].

El radar, en su esencia, cuantifica tres parámetros fundamentales: la reflectividad, relacionada con la potencia de la señal recibida ( $Z$ ); la velocidad radial media, conocida como velocidad Doppler ( $V_r$ ); y el ancho del espectro ( $W$ ), una medida de la variabilidad de las velocidades Doppler dentro del volumen de muestreo [9]. Inicialmente, los radares operaban exclusivamente con polarización horizontal, lo que conducía a la medición de los tres parámetros bajo esta polarización, identificándolos como ZH, VH y WH. Sin embargo, con el advenimiento de los radares de polarización dual, que emplean tanto polarización horizontal como vertical, se expandió el conjunto de observaciones, incluyendo mediciones para ambas polarizaciones y dando lugar a ZH, ZV, VH, VV, WH y WV [8].

Dentro de estos parámetros, la diferencia entre los factores de reflectividad ZH y ZV se denomina reflectividad diferencial, o ZDR. Por otro lado, la discrepancia en el desplazamiento de fase entre las dos polarizaciones proporciona el desplazamiento de fase diferencial, o  $\phi_{DP}$ . Asimismo, la correlación entre las señales retornadas en polarización horizontal y vertical ofrece el coeficiente de correlación copolar

( $\rho_{hv}$ ), aspecto de importancia en la caracterización de los objetos detectados por el radar [9].

En el contexto de la polarización dual, la matriz de dispersión es una herramienta esencial que relaciona los componentes del campo eléctrico incidente y reflejado (*backscatter*) [9]. La representación de esta matriz se formula como:

$$\begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix}$$

En esta expresión, los subíndices “h” “v” denotan las polarizaciones horizontal y vertical tanto para las señales de transmisión (segundo subíndice) como para las de recepción (primer subíndice). Es relevante señalar que, en el caso de gotas esféricas, las componentes  $S_{hh}$  y  $S_{vv}$  son idénticas, mientras que  $S_{vh}$  y  $S_{hv}$  se anulan. Sin embargo, en la presencia de gotas no esféricas,  $S_{vh}$  y  $S_{hv}$  no son iguales a cero y  $S_{hh}$  y  $S_{vv}$  no son iguales [9].

### 2.1. Características del desorden de tierra

El desorden de tierra ha sido objeto de una extensa investigación. A continuación, se presentan los rasgos predominantes de este tipo de desorden, tal como han sido identificados en estudios llevados a cabo por diversos investigadores:

- “Los ecos de no-precipitación, generalmente, poseen una mayor variabilidad espacial en la horizontal o en la vertical, o en ambas” [10].
- “Las ondas del desorden del suelo son principalmente despolarizadas. El grado de polarización es muy diferente entre el desorden del suelo y los ecos climáticos, lo cual permite la posibilidad de utilizar el grado de polarización para la detección de las células de resolución de rango que son afectadas por el desorden de tierra” [11].
- “La extensión vertical del patrón de reflectividad es un ejemplo. Tanto en la propagación media (MP) como en las condiciones AP, los ecos de saturación solo afectan las inclinaciones más bajas del radar. Por lo tanto, la parte superior del eco (es decir, la extensión vertical de los ecos del radar) en las regiones desordenadas no afectadas por la precipitación se limita a las elevaciones más bajas. Por este motivo, varios autores lo han utilizado como una característica del desorden” [12].
- “En radares con capacidad Doppler, la velocidad radial puede ser útil para discriminar el desorden, porque tiende a tener velocidades bajas. Se implementa en muchas de las técnicas revisadas [ver Giuli et al. 1991; Lee et al. 1995; Bellon y Kilambi 1999; Kessinger et al. 2003 (en este caso, junto con el ancho espectral); y Lee et al. 2005]” [12].

- “Por lo general, el desorden del suelo (ya sea de fuentes NP o AP) se caracteriza por una velocidad radial cercana a cero, un ancho de espectro bajo y una textura alta del campo de reflectividad, entre otros” [13].

Resumiendo, el eco de tierra se distingue por su marcada variabilidad espacial y su limitada extensión vertical, generalmente confinándose a exploraciones a bajas altitudes del radar. Este fenómeno tiende a manifestarse a proximidad del radar, presentando velocidades radiales cercanas a cero o inferiores a 2 metros por segundo, además de exhibir un espectro estrecho.

### 3. Lógica difusa

Los seres humanos destacan por su destreza en la gestión de la imprecisión. Son capaces de interpretar y aplicar instrucciones vagas, cuyo significado es subjetivo, como “bastante caliente”, “más o menos rápido”, “muy alto”. Estas directrices, si bien flexibles y adaptables, introducen cierto nivel de incertidumbre al carecer de precisión y cuantificación. A pesar de desafiar la lógica convencional, la habilidad humana para razonar y tomar decisiones basadas en información vaga e imprecisa es excepcional, aunque compleja de emular para sistemas informáticos que requieren definiciones precisas. La incertidumbre en el conocimiento experto es un fenómeno común en diversas disciplinas, como medicina, ciencia, ingeniería, derecho y educación. En el campo de la Inteligencia Artificial, se afronta esta incertidumbre en áreas como la visión por computadora, el procesamiento de lenguaje natural, el aprendizaje automático y los juegos [14].

La lógica difusa permite modelar la imprecisión y vaguedad inherente en algunos sistemas y procesos. Se basa en la teoría de conjuntos difusos, donde un elemento puede pertenecer parcialmente a un conjunto, en lugar de hacerlo de forma binaria como en la lógica booleana clásica. Esta lógica intenta formalizar matemáticamente el razonamiento aproximado que realizan los humanos para tomar decisiones en condiciones de incertidumbre e información imperfecta. Posibilita así dotar a las máquinas de capacidades de inferencia, aprendizaje y optimización similares a la inteligencia humana [15].

La lógica difusa, desarrollada por Lotfi Zadeh en 1965, lleva más de medio siglo siendo una tecnología clave en el campo de la inteligencia artificial. Zadeh en su principio de incompatibilidad plantea que la descripción precisa de sistemas complejos es inalcanzable, por lo que se deben desarrollar herramientas para manejar información imprecisa confiablemente [16]. Esto involucra la representación de dicha información mediante conjuntos difusos y proposiciones condicionales Si-Entonces. Este principio subraya que el razonamiento exacto es un caso particular del razonamiento aproximado. Permite integrar conocimiento humano en sistemas de ingeniería al fuzzificar cualquier sistema lógico.

Los sistemas difusos resultan útiles en problemas complejos sin modelos matemáticos simples, especialmente con comportamientos no lineales, ya que posibilitan decisiones ante información incierta o incompleta. Además, destacan por expresar conocimiento experto de forma lingüística y generalmente más accesible [16].

#### 3.1. Conjuntos Difusos

En contraste con la lógica booleana tradicional, que es binaria, la lógica difusa es multi-valuada, estableciendo grados de pertenencia que varían desde 0 (completamente falso) hasta 1 (completamente verdadero). Esta lógica considera a la lógica binaria como un subconjunto de ella. Zadeh introdujo esta idea en 1965 con el concepto de Conjunto Difuso, lo que marcó el comienzo de la Teoría de Conjuntos Difusos [17]. En esta teoría, los conceptos se vinculan a conjuntos difusos mediante un proceso conocido como fuzzificación. Esto permite trabajar con reglas lingüísticas y generar una salida que puede ser difusa o defuzzificada para obtener un valor discreto [17].

En contraposición a la teoría de conjuntos tradicional, que se fundamenta en el principio lógico básico de que un individuo pertenece o no a un conjunto, el concepto central de un conjunto difuso es que un elemento es parte de un conjunto con un grado específico de pertenencia. Así, una afirmación no es completamente (sino parcialmente) verdadera o falsa. Este grado se representa mediante un número en el rango  $[0, 1]$  [14]. En la parte superior de la figura 1, se muestra la representación clara (crisp), donde se dibuja una línea que distingue de manera evidente a las personas altas de las que no lo son, basándose en una altura de 1.8 m. No obstante, el conjunto difuso, en la parte por ejemplo, una persona con una altura de 1.9 m es parte del conjunto de personas altas con un grado de 0,9 y es parte del conjunto de personas bajas en un grado de 0.05. De este modo, un conjunto difuso ofrece una transición gradual entre los límites de lo que sería un conjunto claro.

En lógica difusa, el grado de membresía es un valor numérico entre 0 y 1 que indica en qué medida un elemento pertenece a un conjunto difuso. Representa la intensidad con la que un elemento satisface una propiedad, permitiendo modelar conceptos imprecisos. A diferencia de la lógica clásica binaria, los valores intermedios capturan la vaguedad e incertidumbre del razonamiento humano. El grado de membresía posibilita aplicaciones como control difuso, toma de decisiones e inteligencia artificial donde la precisión exacta es innecesaria o difícil de obtener [14].

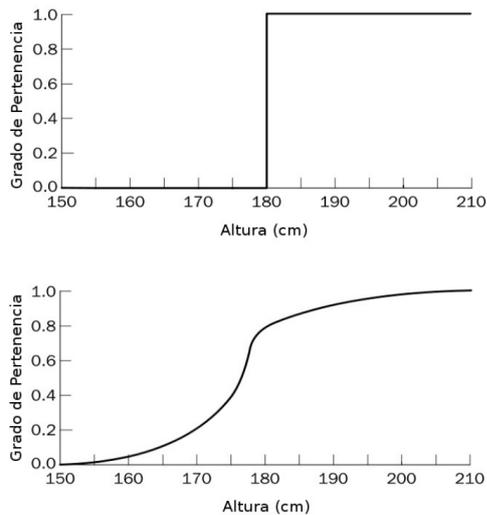


Figura 1: Descripción de conjuntos crisp (arriba) y fuzzy (abajo) de una “persona alta” [18].

La teoría de conjuntos difusos proporciona un marco más general que el de los conjuntos ordinarios, con amplias aplicaciones. Se define un conjunto difuso  $A$  en un espacio  $X$  como una función de pertenencia  $f_A(X)$  que asigna a cada elemento  $x$  en  $X$  un valor entre 0 y 1 [19]. Las propiedades habituales de los conjuntos se extienden a los conjuntos difusos de la siguiente manera:

- El conjunto vacío tiene función de pertenencia 0 para todo  $x$ .
- Dos conjuntos  $A$  y  $B$  son iguales si  $f_A(X) = f_B(X)$  para todo  $x$ .
- El complemento  $A'$  de  $A$  tiene función:  
 $f_{A'}(X) = 1 - f_A(X)$
- $A$  está contenido en  $B$  si  $f_A(X) \leq f_B(X)$  para todo  $x$ .
- La unión  $C = A \cup B$  tiene función:  
 $f_C(X) = \max(f_A(X), f_B(X)), x \in X$
- La intersección  $C = A \cap B$  tiene función:  
 $f_C(X) = \min(f_A(X), f_B(X)), x \in X$ . [19]

De esta forma, la teoría de conjuntos difusos generaliza los conjuntos ordinarios y permite representar conceptos imprecisos de gran utilidad en diversos campos.

La inferencia difusa consiste en aplicar reglas lógicas difusas del tipo “Si-Entonces” que modelan relaciones entre variables. Estas reglas combinan información difusa de múltiples fuentes para producir resultados difusos aproximados. Las reglas capturan conocimiento experto o vaguedad mediante proposiciones condicionales. El proceso implica fuzzificar entradas precisas a grados de membresía difusos, aplicar las reglas para inferir salidas difusas, y finalmente defuzzificar a valores concretos. Este tipo de razonamiento aproximado

permite decisiones ante incertidumbre. Es útil cuando la información es incompleta y las relaciones no son binarias precisas [20].

Los conjuntos difusos representan información vaga mediante membresías graduales, a diferencia de la lógica booleana tradicional binaria. Las reglas difusas relacionan estos conjuntos con proposiciones “SI...ENTONCES”, modelando conceptos imprecisos del mundo real. La inferencia difusa combina estas reglas y valores de entrada difusos para razonar e inferir resultados, que luego se defuzzifican en decisiones concretas [20].

### 3.2. Clasificador Difuso

Un clasificador difuso es un sistema de clasificación que utiliza lógica difusa para asignar una etiqueta a una entrada en función de su grado de pertenencia a diversas categorías. El funcionamiento de un clasificador difuso implica tres pasos clave: **fuzzificación**, **aplicación de reglas difusas** y **Defuzzificación**: [21]:

**Fuzzificación:** La fuzzificación convierte las entradas nítidas en valores difusos. Esto mapea las entradas desde un dominio nítido a un conjunto de funciones de pertenencia difusa. Dichas funciones asignan grados de pertenencia a distintas categorías. Por ejemplo, para temperatura se definirían funciones como “frío”, “templado” y “caliente”, que dan grados de pertenencia de la temperatura actual a esas categorías [21]

**Aplicación de reglas difusas:** En esta etapa se aplican reglas que relacionan las entradas difusas con categorías de salida. Cada regla tiene una premisa que combina los grados de pertenencia de las entradas, y una conclusión que define el grado de salida. Por ejemplo: “SI temperatura es fría Y humedad es alta, ENTONCES salida es invierno”. Esta regla toma los grados de “frío” y “alta” para determinar el grado de “invierno” [21].

**Defuzzificación:** La defuzzificación convierte la salida difusa en una decisión nítida. Esto resume los grados de pertenencia en una categoría específica, mediante técnicas como centroide, máximo grado, etc. Así se obtiene una etiqueta a partir de las salidas difusas [21].

### 3.3. Ejemplo de clasificador difuso

Los métodos probabilísticos y basados en lógica difusa presentan numerosas similitudes, como el uso de valores numéricos entre 0 y 1 para expresar incertidumbre. No obstante, también existen diferencias clave. En probabilidad, la intersección de un conjunto y su complementario es siempre un conjunto vacío, mientras que en lógica difusa esto no tiene por qué ser así. Un ejemplo son las regiones de lluvia y granizo, que podrían contener partículas de fase mixta, como gotas de lluvia mezcladas con granizo. Por tanto, entre un

concepto y su opuesto puede haber una zona difusa no vacía [22]. Los eventos que muestran esta vaguedad se modelan mediante conjuntos difusos. Precisamente los límites imprecisos entre observables polarimétricas hacen que la lógica difusa sea ideal para clasificar partículas, ya que permite transiciones graduales entre los diversos tipos de precipitación. Así, un enfoque difuso posibilita una variación fluida de los límites observables polarimétricos entre categorías [23].

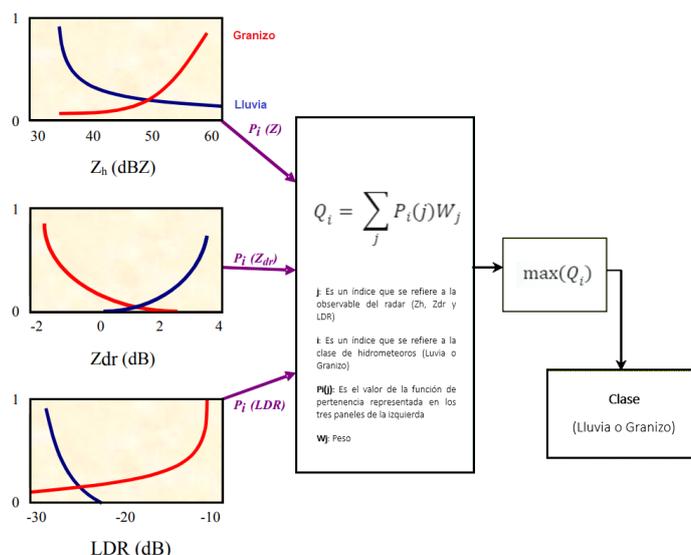


Figura 2: Identificación de partículas basado en lógica difusa. Para cada variable, todas las partículas reciben un valor entre 0 y 1 de la respectiva función de pertenencia (fuzzificación). Estos valores (P) se multiplican luego por el peso apropiado (W) y se suman para cada partícula. Se encuentra el máximo de las sumas ponderadas para determinar el tipo de partícula (defuzzificación) [22].

La Figura 2 muestra un modelo conceptual de una técnica de lógica difusa para detectar lluvia y granizo usando ZHH, ZDR, y LDR. ZHH mide la reflectividad en la polarización horizontal a horizontal. Representa la energía reflejada en la misma polarización del pulso enviado. Se usa para describir la intensidad de la señal reflejada horizontalmente, proporcionando información sobre las partículas [3].

ZDR es la diferencia en reflectividad entre polarizaciones horizontal y vertical. Se calcula entre ZHH y ZVV o ZHV. Caracteriza la forma y tipo de partículas. Valores positivos indican partículas grandes y esféricas, negativos indican alargadas o asimétricas [3].

LDR mide la linealidad de la polarización de la señal reflejada. Indica cuánto se ha modificado la polarización original por interacción con partículas. Valores bajos indican cambios significativos en la polarización original, posiblemente por partículas no esféricas o estructuras aleatorias [3].

Las funciones de pertenencia para ZHH, ZDR, y LDR definen el límite entre las regiones de lluvia y granizo y se formulan en base a sus valores típicos. Para valores particu-

lares de ZHH, ZDR, y LDR, las funciones de pertenencia determinan el grado al que una observación pertenece a cada uno de los conjuntos difusos apropiados, es decir, lluvia y granizo [22].

Un clasificador difuso inicialmente calcula los grados de pertenencia a cada conjunto (ya sea lluvia o granizo) en función de las funciones de pertenencia. Para la clase "lluvia" se define una función de membresía  $mLluvia(Z_h)$  que representa la probabilidad condicional  $Prob[eco \text{ es lluvia} \mid Z_h = x]$ , mientras que  $mGranizo(Z_h)$  corresponde a  $Prob[eco \text{ es granizo} \mid Z_h = x]$ , como se ilustra en los recuadros a la izquierda de la figura. Por ejemplo, un valor de  $Z_h$  de 35 dBZ implicaría una probabilidad condicional de 0.9 para la clase lluvia y menor de 0.1 para granizo. De manera análoga, se definen funciones de membresía para los observables  $Z_{dr}$  y  $LDR$ . Es decir, las funciones de membresía cuantifican los grados de pertenencia a clases difusas dadas las observaciones de entrada [22].

El siguiente paso es ponderar los 6 valores de membresía anteriores de acuerdo con la "confianza" de la medición de cada una de las variables ( $Z_h$ ,  $Z_{dr}$ ,  $LDR$ ). A continuación, los valores de agregación (Q) para las clases de lluvia y granizo se calculan sumando los tres valores de  $P_{lluvia}$  y los tres valores de  $P_{granizo}$  anteriores. La clasificación final de lluvia o granizo se basa en la clase que tiene el valor máximo (Q). El último paso se llama "defuzzificación" que es el resultado de la clasificación que solo se permite una clase de hidrometeoro [9, 22].

#### 4. Diseño del modelo

Diversas técnicas pueden ser empleadas para la implementación de un clasificador, como la lógica difusa, redes neuronales, máquinas de vectores de soporte, árboles de decisión, Naive Bayes, entre otras. Sin embargo, sin importar la técnica utilizada, es esencial identificar una o varias características que posibiliten la distinción entre las señales meteorológicas y las señales de desorden [24]. Afortunadamente, las señales de desorden suelen exhibir estas características distintivas, también conocidas como "firmas polarimétricas". No obstante, es importante tener en cuenta que estas firmas pueden variar en cierta medida en función de la región geográfica o el tipo de clima en el que se ubique el radar [12]. Este aspecto implica que la construcción de un clasificador sea un proceso empírico que requiere de datos de entrenamiento previamente categorizados por un experto. Estos datos son analizados con relación a diversas variables con el propósito de identificar los aspectos que permitan clasificar los diferentes tipos de fenómenos con precisión [25].

De manera general, los pasos necesarios para desarrollar un clasificador son los siguientes (ver Figura 3):

- Recopilar datos provenientes de diversos eventos climatológicos registrados por el radar.

- Realizar una depuración de los datos para valores faltantes, redundantes o atípicos.
- A partir de las variables polarimétricas suministradas por el radar generar nuevas variables como la textura.
- A través de un experto en la materia realizar la clasificación de cada celda de datos, distinguiendo entre eventos de precipitación y desorden de tierra.
- Seleccionar las variables que pueden resultar más eficientes en el proceso de clasificación de cada tipo de fenómeno, ya sea precipitación o desorden de tierra.
- Implementar la técnica de clasificación seleccionada, un clasificador difuso en este caso.
- Ajustar los parámetros del clasificador a través del algoritmo de descenso del gradiente a fin de optimizar su rendimiento.
- Evaluar el desempeño del clasificador tanto con los datos de entrenamiento como con los datos de prueba.

Una vez completados los dos primeros pasos, se procede a llevar a cabo un análisis de los datos recopilados por el radar, los cuales han sido previamente categorizados por un experto en la materia. Este análisis implica un estudio estadístico de las variables polarimétricas relacionadas con cada uno de los eventos clasificados. Resulta esencial la creación de nuevas variables, distintas de las que originalmente captura el radar, dado que numerosos fenómenos exhiben características diferenciadoras que solo pueden ser identificadas a través de su variabilidad espacial o temporal [7]. Estas nuevas variables se relacionan con la desviación estándar o la textura calculada en una cuadrícula cuyo centro es la celda bajo análisis.

plano en una serie de celdas. Cada una de estas celdas está asociada con un conjunto de medidas que constituyen las variables polarimétricas fundamentales, entre las que se incluyen la reflectividad (Z), la reflectividad diferencial (Zdr), la fase de propagación diferencial ( $\Phi_{dp}$ ), la fase diferencial específica (Kdp), el coeficiente de correlación ( $\rho_{co}$ ) y la velocidad Doppler ( $V_r$ ) [3]. Sin embargo, el comportamiento individual de estas variables no resulta suficiente para la implementación de un clasificador eficiente. Para abordar esta limitación, se recurre a la observación de nuevas variables generadas mediante el cálculo de la textura o la desviación estándar de las mencionadas variables polarimétricas, como Z, Zdr,  $\Phi_{dp}$ , entre otras [23]. Estas nuevas variables, derivadas de la variabilidad espacial de los datos, desempeñan un papel fundamental en la identificación de ecos de desorden o saturación, ya que su comportamiento diferencial puede proporcionar valiosa información para el proceso de clasificación.

En meteorología, la textura se calcula a partir de datos de radar para obtener información detallada sobre la estructura y composición de las partículas en la precipitación y otros eventos atmosféricos. La textura se refiere a la variación espacial de una propiedad de un objeto o fenómeno meteorológico. El cálculo de la textura mediante datos de radar es una herramienta útil en meteorología para estudiar en detalle las propiedades e irregularidades de los fenómenos atmosféricos por lo que permite detectar la presencia de eventos atmosféricos anómalos.

La textura se calcula a través de la siguiente expresión:

$$\text{Text}(y_{a,b}) = \sqrt{\frac{\sum_{i=-(m-1)/2}^{(m-1)/2} \sum_{j=-(n-1)/2}^{(n-1)/2} (y_{a,b} - y_{a+i,b+j})^2}{mn}} \quad (1)$$

Donde a y b representan el azimut y el rango de la compuerta. La textura de  $tZDR$ ,  $t\Phi_{dp}$  se calcula en una cuadrícula que consta de tres celdas en dirección de azimut (m) y tres en dirección de rango (n) centrados en la compuerta. Es decir, se calcula la textura para la celda central en función de las 8 celdas vecinas. Dado que el radar de Corozal tiene una resolución azimutal de  $1^\circ$  y una resolución de rango de 450 m, la textura es representativa de un área de  $3^\circ \times 1350$  m. Esta área aumenta con el rango; por lo tanto, la variable de textura tendrá una dependencia de rango.

La dependencia del rango implica que el cálculo de la textura debe ser corregido ya que las celdas más alejadas del radar al ser más grandes afectan la interpretación de la variabilidad espacial de una variable polarimétrica. Esto se debe a que la resolución espacial disminuye a medida que nos alejamos del radar, y las celdas más alejadas cubren un área mayor en comparación con las celdas cercanas al radar [23]. Inicialmente se realiza una regresión lineal de los valores

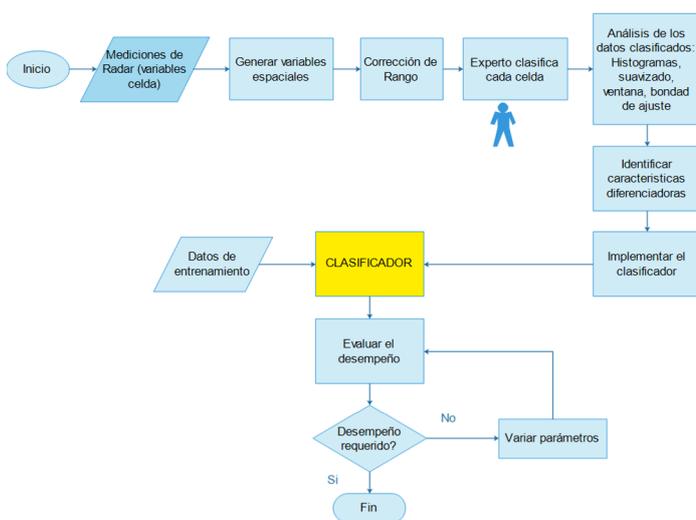


Figura 3: Diagrama de flujo del modelo de construcción de un clasificador

Para comprender con mayor detalle este proceso, es importante destacar que, durante un barrido, el radar divide un

de textura contra distancia, este proceso permite determinar que la pendiente en las distancias más cercanas es menor que la de las distancias lejanas. Mediante un proceso de interpolación se corrigen los cálculos de la textura de las celdas lejanas.

$$\sum_{i=1}^n (Y_i - aX_i - b)^2 \quad (2)$$

Donde:  $Y_i$  es el valor observado de la variable dependiente en la  $i$ -ésima observación.  $X_i$  es el valor observado de la variable independiente en la  $i$ -ésima observación.  $n$  es el número total de observaciones. Los valores óptimos de “a” (pendiente) y “b” (intersección en el eje y) que minimizan esta suma se calculan mediante:

$$a = \frac{n \sum_{i=1}^n (X_i Y_i) - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n (X_i^2) - \left( \sum_{i=1}^n X_i \right)^2}, \quad (3)$$

$$b = \frac{\sum_{i=1}^n Y_i - a \sum_{i=1}^n X_i}{n}$$

El comportamiento de las diversas variables polarimétricas se puede analizar con precisión mediante la construcción de histogramas, los cuales permiten visualizar la distribución de frecuencias relativas de estas variables. Los histogramas se convierten en una herramienta esencial para determinar cuáles de estas variables son más efectivas en el proceso de clasificación. Por ejemplo, si se desea evaluar cómo se comporta la desviación estándar de  $\Phi dp$  en la diferenciación entre precipitación y el eco de tierra, se procede a crear un histograma que incluye todos los valores de esta variable presentes en las celdas clasificadas como precipitación [12].

Posteriormente, se compara este histograma con otro que contiene todos los valores de la misma variable en las celdas clasificadas como eco de tierra. La medida del área de superposición entre ambos histogramas se relaciona de manera inversa con la eficacia de la variable en la distinción entre los dos fenómenos, lo que proporciona una valiosa guía para el proceso de clasificación [26].

El proceso de identificación del área de superposición se puede hacer calculando la función de densidad después de realizar un proceso de interpolación y suavizado de los datos del histograma. Esta estrategia resulta más eficaz debido a que implica una etapa de interpolación y suavizado, lo que conlleva a la eliminación de valores atípicos, la mitigación de los efectos del ruido y, en última instancia, a una mejor comprensión del comportamiento de las diversas variables [24].

A partir de los histogramas de cada variable polarimétrica se calculan las funciones densidad de para cada clase (es

decir, precipitación, desorden del suelo, y ecos de aire claro) utilizando la estimación densidad kernel Gaussiana [27].

$$\hat{f}(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_i - x}{\sigma}\right)^2\right) \quad (4)$$

Una vez obtenidas las funciones de densidad es posible llevar a cabo una prueba de bondad de ajuste para determinar si la variable en cuestión sigue una distribución de probabilidad conocida. En casos en los que la variable no se ajuste a ninguna distribución preestablecida, aún es factible trabajar exclusivamente con las curvas de distribución de frecuencias obtenidas, lo que proporciona información valiosa para el proceso de análisis y clasificación [24].

A partir de las funciones de distribución, en el caso de que las variables superen las pruebas de bondad de ajuste, o bien utilizando funciones de densidad de los histogramas, es posible determinar las funciones de membresía que alimentarán el clasificador.

Una función de membresía que cuantifica la “probabilidad condicional” de que el eco pertenece a una clase particular dados los valores de las observaciones de entrada [9]. La función de membresía se puede obtener usando Bayes donde  $T$  es señal de eco de tierra,  $L$  representa la precipitación o lluvia y  $x$  es el valor de la variable polarimétrica [26].

$$P(L|x) = \frac{P(L)P(x|L)}{P(L)P(x|L) + P(T)P(x|T) + P(A)P(x|A)} \quad (5)$$

No obstante, es importante destacar que no es un requisito obligatorio calcular estas funciones de membresía, ya que el clasificador puede funcionar de manera igualmente efectiva si se emplean las curvas de distribución en su lugar [23]. Esta afirmación se respalda con resultados observados que muestran una similitud notable al implementar tanto curvas de distribución de probabilidad como funciones de membresía en el clasificador desarrollado por Rico-Ramirez & Cluckie [24].

La implementación del clasificador se fundamenta en las variables previamente identificadas como las más apropiadas para catalogar de manera efectiva las mediciones obtenidas a través del radar. Posteriormente, una vez que el clasificador ha sido desarrollado, ya sea de tipo Bayesiano, difuso o basado en una red neuronal, se procede a llevar a cabo la fase de entrenamiento.

Los datos se organizan en una tabla donde cada fila representa una celda del radar y cada columna una variable. Las columnas incluyen valores de variables polarimétricas, las nuevas variables generadas a partir del cálculo de la textura y una etiqueta categórica indicando si la celda contiene precipitación o eco. La tabla se divide en dos conjuntos de datos, uno para entrenar el modelo y otro para evaluar su desempeño. Durante el entrenamiento, el modelo aprende a predecir la etiqueta de clase a partir de los valores de las variables. Posteriormente, se hace pasar el conjunto de prueba por el modelo entrenado para obtener predicciones y se comparan con las etiquetas reales para calcular la tasa de error. Esta tasa de

error en entrenamiento y prueba cuantifica la capacidad del modelo para generalizar a nuevos datos. Entre menor sea el error en ambos conjuntos, mejor es el desempeño del modelo [28].

En el entrenamiento del clasificador difuso se usa el algoritmo de descenso del gradiente para minimizar la función de coste o error que compara la diferencia entre la distribución de los valores predichos por el clasificador y la distribución los valores reales o etiquetas de cada celda [9]. El valor predicho por el clasificador difuso se obtiene a partir de la siguiente expresión:

$$Q_i = \max \left( \sum_j P_i(j) W_j \right) \quad (6)$$

Donde  $j$  es el valor de la variable polarimétrica. El tipo de fenómeno (precipitación, eco de tierra o eco de aire) se representa con  $i$ .  $P_i(j)$  es el valor de pertenencia a cada uno de los 3 fenómenos calculado con base en la función de membresía o función de densidad de cada variable polarimétrica.

Se calcula el error usando la función de entropía cruzada donde:  $L$  es la función de costo o error.  $y_i$  es la etiqueta de clase para la  $i$ -ésima muestra.  $Q_i$  es el valor predicho por el clasificador la  $i$ -ésima muestra [29].

$$L = - \sum_i (y_i \log Q_i) \quad (7)$$

Una vez calculado el error se usa el descenso del gradiente para ajustar los valores de  $W_j$ . En cada iteración, calcula el gradiente de la función de costo con respecto a los parámetros del clasificador difuso. Esto implica derivar la función de costo con respecto a cada parámetro del clasificador  $W_j$ . La actualización de los parámetros se realiza multiplicando el gradiente por una tasa de aprendizaje (*learning rate*) y restando el resultado a los parámetros [30].

$$W_{jn} = W_{ja} - \eta \frac{dL}{dW_j} \quad (8)$$

Donde  $W_{jn}$  es el nuevo valor del parámetro.  $W_{ja}$  es el antiguo valor del parámetro.  $\eta$  es la tasa de aprendizaje (*learning rate*) y  $dL/dW_j$  es el gradiente de la función de costo en función de cada parámetro. Dado que no se dispone de un conjunto de datos de radares previamente analizados y clasificados por un experto en el país, se opta por suplir las funciones de dicho experto mediante la implementación de dos clasificadores basados en los documentos desarrollados por Gourley [23], y Rico-Ramirez [24]. Estos clasificadores se utilizan como referencia para entrenar y ajustar el clasificador en desarrollo, de manera que pueda realizar clasificaciones precisas en las condiciones específicas de interés.

## 5. Implementación del modelo

### 5.1. Datos de trabajo

Los datos empleados se originan en un radar de banda C situado en Corozal, Sucre. Este radar opera a una frecuencia de 5.624 GHz y tiene una frecuencia de repetición PRF (*Pulse Repetition Frequency*) de 500 Hz. Su emplazamiento se encuentra en la cabecera del Aeropuerto Las Brujas, con coordenadas geográficas de 9.331 de latitud, -75.28 de longitud y una altitud de 143 metros sobre el

nivel del mar. Cada PPI (*Plan Position Indicator*) se compone de 664 celdas, cada una de 450 metros de longitud, lo que permite un alcance efectivo de 298.8 kilómetros. Dentro de cada celda, el radar proporciona mediciones para seis variables polarimétricas: Reflectividad Z, Reflectividad Diferencial Zdr, Fase de Propagación Diferencial  $\Phi_{dp}$ , Fase Diferencial Específica Kdp, Coeficiente de Correlación  $\rho_{co}$  y Velocidad Doppler Vr.

### 5.2. Implementación del clasificador “experto”

En Matlab, se desarrolló un clasificador basado en el artículo titulado “A Fuzzy Logic Algorithm for the Separation of Precipitating from Nonprecipitating Echoes Using Polarimetric Radar Observations” [23]. Este algoritmo tiene la capacidad de asignar cada celda a uno de tres tipos de fenómenos: Precipitación, ecos de tierra (*Ground Clutter*) y ecos de aire claro (*Clear Air*). Para llevar a cabo esta clasificación, el algoritmo utiliza tres variables fundamentales: la Textura de Reflectividad Diferencial tZDR, la Textura de la Fase de Propagación Diferencial t $\Phi_{dp}$  y el Coeficiente de Correlación  $\rho_{hv}$  [23].

Los autores derivaron funciones de densidad de probabilidad (Figura 4) a partir de los datos, las funciones caracterizan el comportamiento de las variables asociadas a los fenómenos de precipitación, desorden de tierra y aire claro (*Clear Air*). Estas curvas probabilísticas sirvieron de base para la implementación de un clasificador difuso en Matlab.

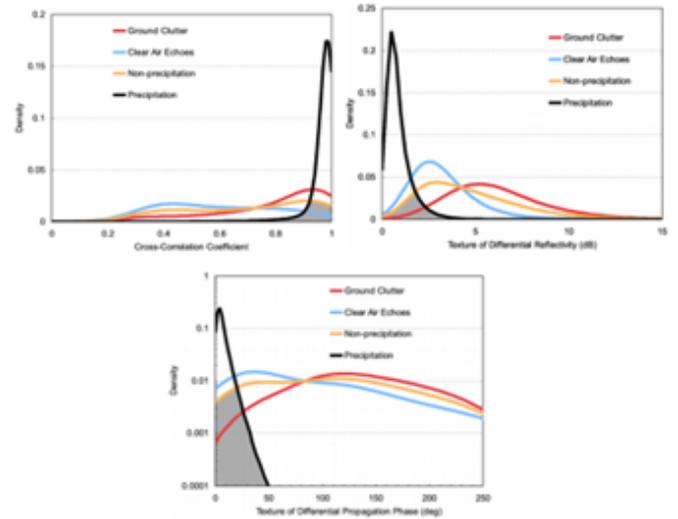


Figura 4: Funciones de densidad para las variables textura de reflectividad diferencial tZdr, textura de la fase de propagación diferencial t $\Phi_{dp}$  y el coeficiente de correlación  $\rho$  [23].

### 5.3. Corrección de rango

A partir de las seis variables registradas por el radar, se generan nuevas variables de carácter espacial, específicamente, la textura. El cálculo de esta textura se llevó a cabo en una cuadrícula de dimensiones 3x3, con la celda analizada ubicada en su centro. Dado que a medida que nos alejamos del radar, las celdas cubren áreas cada vez mayores, las cuadrículas utilizadas para el cálculo se tornan progresivamente más extensas [23]. Por lo tanto, fue

necesario aplicar un factor de corrección que tuviera en cuenta esta diferencia en las áreas. La figura 5 ilustra la relación entre el alcance y la textura de reflectividad diferencial (tZDR); cada punto representa el promedio de las texturas a una distancia específica del radar [23]. Para realizar la corrección de los datos, se efectuó una substracción de cada valor calculado, considerando un factor que varía según el grado de aumento de la textura en función de la distancia. La obtención de este grado de crecimiento se realizó mediante la aplicación de una regresión lineal a los datos [23].

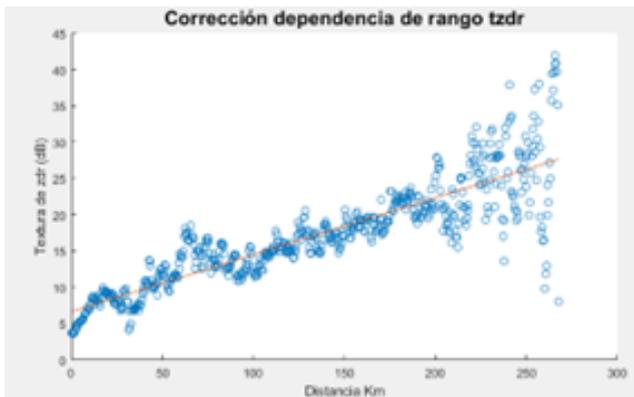


Figura 5: Dependencia del rango de tZdr (Elaboración propia).

La corrección de rango es un proceso esencialmente empírico, donde el factor de corrección se encuentra intrínsecamente ligado a las particularidades del radar utilizado.

#### 5.4. Resultados de clasificación del experto

Se aplicó el denominado clasificador “experto” a los datos correspondientes a un evento de precipitación en Corozal. Este clasificador asignó a cada celda una etiqueta que indica si en la celda se tiene precipitación (1), desorden de tierra (2) o aire claro (3). Los resultados se representan gráficamente en la Figura 6. De las 79,099 celdas evaluadas, 76,606 se categorizaron como precipitación, 373 como eco de tierra (*ground clutter*) y 2,120 como ecos de aire claro (*clair air*).

#### 5.5. Análisis de información: Identificación de las mejores variables

La teoría nos indica que el desorden de tierra presenta notables fluctuaciones espaciales [10], lo que motiva el cálculo de texturas para cada variable en una ventana de dimensiones 3x3. Los datos de cada variable se filtran en grupos correspondientes al número de categorías de clasificación existentes, y se genera un histograma para cada grupo.

La Figura 7 muestra las sunciones de densidad de  $t\Phi_{dp}$ , ZDR y  $\rho_{hv}$ , que representan el comportamiento de cada variable en los casos de precipitación, *ground clutter* o *clair air*.  $t\Phi_{dp}$  se revela como un eficaz clasificador, dado que existe una mínima superposición entre las curvas de precipitación y desorden.

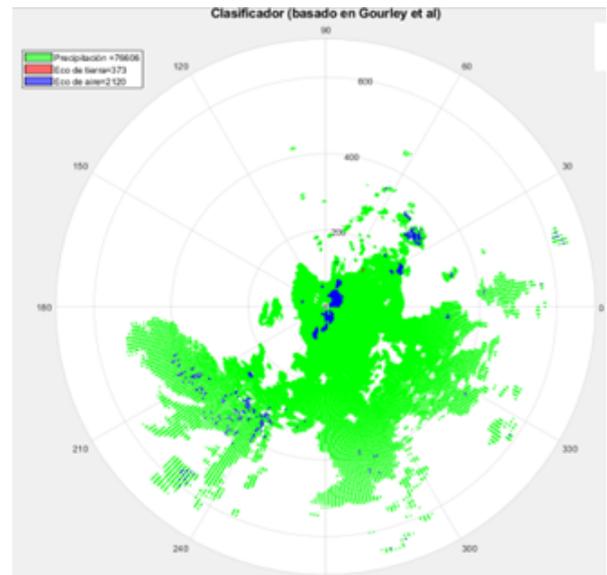


Figura 6: Resultados del clasificador “Experto” implementado. (Elaboración propia).

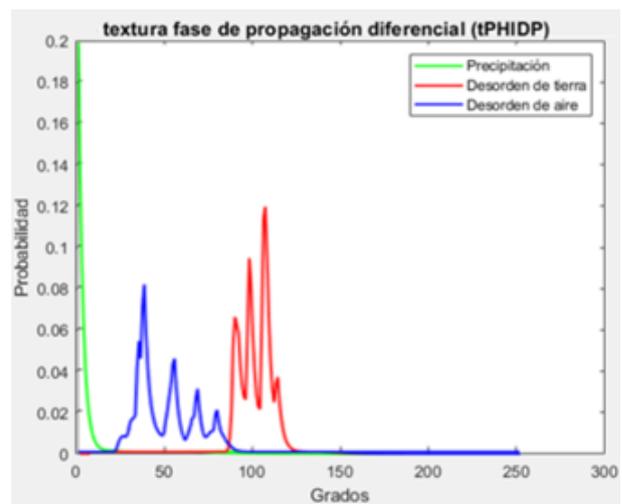


Figura 7: Textura de  $t\Phi_{dp}$  para cada clase de fenómeno (Precipitación, desorden de tierra, desorden de aire). Elaboración propia.

La Figura 8 destaca que la textura de ZDR también se posiciona como un excelente clasificador. Por otro lado, la Figura 9 revela que los valores de  $\rho_{hv}$  superiores a 0.95 corresponden a precipitación, mientras que aquellos situados entre 0.74 y 0.95 indican eco de aire, y valores inferiores a 0.74 representan eco de tierra.

En resumen, los histogramas de  $t\Phi_{dp}$ , tZdr y  $\rho_{hv}$  sugieren que estas variables funcionarán como clasificadores efectivos. Sin embargo, los histogramas de las texturas de las demás variables exhiben amplias zonas de superposición, razón por lo cual fueron descartadas.

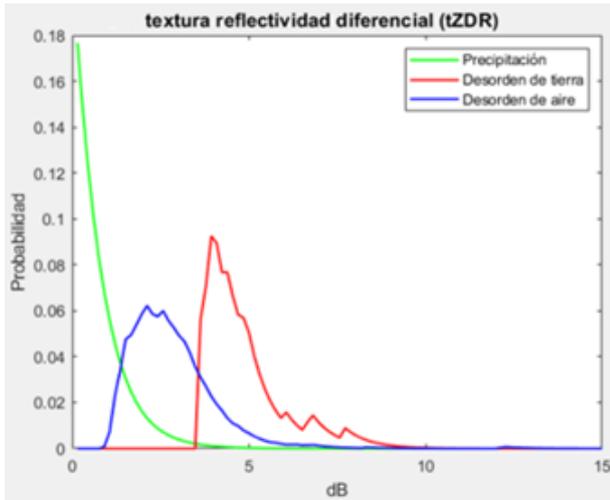


Figura 8: Textura de tzdr. Elaboración propia.

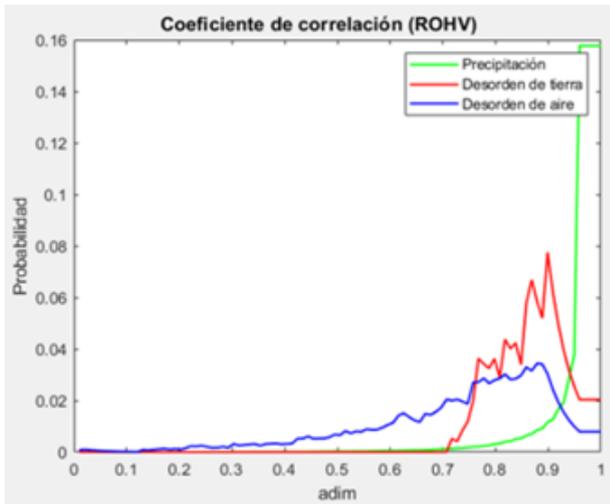


Figura 9: Coeficiente de correlación copolar  $\rho_{hv}$ . Elaboración propia

### 5.6. Suavizado de los datos

En los histogramas obtenidos, se observaron frecuencias relativas nulas para diversos valores de la variable. No obstante, esta ausencia de frecuencia relativa no indica necesariamente que la variable sea igual a cero en esos puntos; simplemente refleja que en los datos de entrenamiento, la variable no asumió esos valores. Para calcular los valores intermedios, se empleó la ecuación de la recta que se extiende entre dos mediciones consecutivas. Además del suavizado, se aplicó una ventana de tamaño 5 a los datos. Esto significa que el valor de la variable en un punto se reemplazó por el promedio de dicho punto, los dos puntos previos y los dos puntos subsiguientes. Esta estrategia contribuye a mitigar tanto el ruido como la influencia de valores atípicos.

### 5.7. Pruebas de bondad de ajuste

Se llevaron a cabo pruebas de bondad de ajuste chi cuadrado con diversas distribuciones, y en todos los casos, los resultados arro-

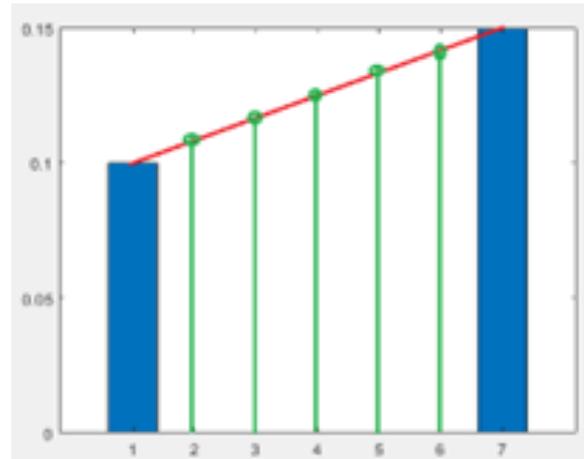


Figura 10: Suavizado de los datos, elaboración propia.

jaron resultados negativos. No obstante, este hallazgo no reviste una gran relevancia, ya que aún es posible implementar el clasificador utilizando las distribuciones identificadas o las funciones de membresía derivadas de ellas.

### 5.8. Implementación del clasificador difuso y pruebas de desempeño

Las funciones de densidad sirvieron como punto de partida para la construcción del clasificador difuso. Una vez implementado, se llevaron a cabo pruebas de rendimiento.

En la primera prueba, los parámetros de agregación o ponderación se establecieron en 1 para  $t\Phi_{dp}$ , 1 para  $tZdr$  y 1 para  $\rho_{hv}$ , con los siguientes resultados: 98.48 % para precipitación, 100 % para desorden de tierra y 85.71 % para eco de aire. Estos porcentajes representan la coincidencia de clasificación entre el clasificador desarrollado y el clasificador experto inicial.

Posteriormente, se emprendió la tarea de encontrar los valores óptimos de los parámetros para maximizar la eficiencia. Se empleó el algoritmo de descenso del gradiente, que en este contexto busca maximizar los porcentajes de coincidencia. Se determinó que las ponderaciones que generaban el mayor nivel de eficiencia eran las siguientes: 0.442 para  $t\Phi_{dp}$ , 0.564 para  $tZdr$  y 0.456 para  $\rho_{hv}$ . Estas ponderaciones condujeron a tasas de aciertos de 99.62 % para precipitación, 100 % para desorden de tierra y 87.89 % para eco de aire.

### 6. Validación del modelo

En principio, un porcentaje de coincidencias en la clasificación de más del 98 % indica la validez del modelo propuesto. Sin embargo, es importante considerar que este alto nivel de eficiencia se logró cuando el clasificador desarrollado etiquetó los mismos datos con los que fue entrenado. La verdadera validación del modelo se alcanza cuando se aplica a datos distintos de los utilizados durante el entrenamiento. Para este propósito, se recopiló información de 5 eventos del radar de Corozal, lo que resultó en un archivo con datos de 431,879 celdas. Cada celda fue etiquetada mediante el clasificador “experto”, y los resultados fueron los siguientes:

417,962 celdas de precipitación (96.78%), 572 celdas de eco de tierra (0.13%) y 13,345 celdas de eco de aire (3.09%).

La información se dividió en dos grupos: el 70% se destinó al entrenamiento del clasificador desarrollado, y el 30% restante se reservó para validar su rendimiento. Los datos de entrenamiento se seleccionaron de forma aleatoria, respetando las proporciones del conjunto total de datos. En otras palabras, de los 302,315 datos de entrenamiento, 292,580 correspondieron a precipitación, 393 a ecos de tierra y 9,341 a ecos de aire.

Para medir la eficiencia, se empleó el parámetro del error. El error se calculó como el número total de celdas clasificadas de manera diferente entre el "experto" y el nuevo clasificador, dividido por el total de celdas clasificadas.

$$\text{error} = \frac{\text{Total de no coincidencias}}{\text{Total de celdas clasificadas}} \quad (9)$$

Los pesos óptimos de las variables utilizadas ( $t\phi dp$ ,  $tZdr$  y  $\rho_{hv}$ ) se determinaron mediante la implementación del algoritmo de descenso del gradiente. Durante el entrenamiento, se alcanzó la máxima eficiencia con las siguientes ponderaciones: 0.418 para  $t\phi dp$ , 0.605 para  $tZdr$  y 0.452 para  $\rho_{hv}$ , lo que resultó en tasas de aciertos de 99.614% para precipitación, 98.51% para desorden de tierra y 87.65% para eco de aire, con un error de 0.00478. Al emplear estas ponderaciones en la validación, se obtuvieron los siguientes resultados: coincidencias de 99.625% para precipitación, 95.40% para desorden de tierra y 87.26% para eco de aire, con un error de 0.00473.

Es relevante destacar que el error de validación resultó ser incluso mejor que el error de entrenamiento, lo cual refuerza la validez del modelo y descarta la posibilidad de sobre-entrenamiento (*overfitting*).

Tabla 1: Matriz de confusión: validación

	Precipitación Predicha	Desorden de Tierra Predicho	Desorden de Aire Predicho	Total Real
Precipitación Real	127999	41	441	128481
Desorden de Tierra Real	3	83	1	87
Desorden de Aire Real	77	50	870	997
Total Predicho	128079	174	1312	129565

### 7. Corrección de la clasificación

La eficiencia de la clasificación puede mejorar significativamente si aprovechamos el hecho de que el desorden de tierra generalmente se encuentra concentrado en los primeros dos *Plan Position Indicators* (PPI), lo que implica que su extensión vertical es limitada. Según lo indicado en [31] y [32]: "Tanto en la propagación media (MP) como en las condiciones AP, los ecos de saturación solo afectan las inclinaciones más bajas del radar. Por lo tanto, la parte superior del eco (es decir, la extensión vertical de los ecos del radar) en las regiones desordenadas no afectadas por la precipitación se limita a las elevaciones más bajas."

Por lo tanto, podemos utilizar esta limitación en la extensión vertical del desorden para mejorar la clasificación. Si una celda se etiqueta como eco de tierra o aire en el primer PPI, comparamos esta clasificación con la misma celda en el segundo PPI. En caso de que la celda superior se clasifique como precipitación, podemos corregir la etiqueta de la celda inferior a precipitación, ya que es poco probable que el eco de la celda en el primer PPI se extienda a la misma celda en el segundo PPI.

Si la celda en el segundo PPI también se clasifica como eco de tierra o aire, podemos reemplazar esta etiqueta por la de la misma celda en el tercer PPI si su clasificación difiere de las dos clasificaciones anteriores. Si en los tres PPIs la celda se clasifica como eco de tierra, entonces examinamos las 8 celdas circundantes. Si más de 4 de estas celdas adyacentes se clasifican como precipitación, corregimos la clasificación de la celda en cuestión. La figura 11 ilustra la metodología utilizada para la corrección de los datos proporcionados por el nuevo clasificador.

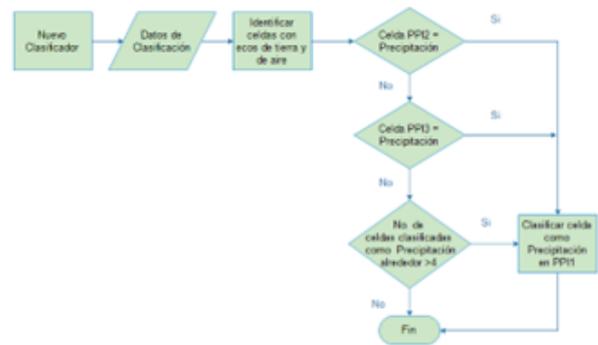


Figura 11: Modelo de corrección de datos de clasificación. Elaboración propia.

Gracias a la corrección implementada, se pudo reducir el número de celdas etiquetadas como eco de tierra de 432 a tan solo 15 celdas, mientras que las celdas clasificadas como eco de aire disminuyeron de 2248 a 353 celdas.

### 8. Aplicación desarrollada

La aplicación desarrollada ofrece la capacidad de construir un clasificador a partir del archivo netcdf proveniente del radar, realizando una serie de acciones clave:

1. Extracción del primer Plan Position Indicator (PPI) y ajuste de las 6 variables polarimétricas medidas por el radar.
2. Cálculo de nuevas variables espaciales, incluyendo textura y desviación estándar.
3. Corrección de rango de las variables espaciales.
4. Aplicación del clasificador experto implementado al PPI-1, que etiqueta cada celda.
5. Utilización de los datos de entrenamiento proporcionados por el experto para construir histogramas individuales para cada variable.
6. Suavizado de los datos obtenidos a partir de los histogramas.
7. Generación de funciones de membresía o pertenencia a partir de los datos suavizados provenientes de los histogramas.

8. Clasificación de las mediciones del PPI-1 utilizando el nuevo clasificador difuso.
9. Comparación de los resultados del experto con los del clasificador desarrollado para determinar su eficiencia.
10. Ajuste de las ponderaciones hasta alcanzar la eficiencia “máxima”.
11. Corrección de la clasificación basada en los datos de los PPIs superiores o de las celdas adyacentes.

## 9. Conclusiones

En este trabajo, se ha evidenciado la alta variabilidad espacial del desorden. Variables como la textura o la desviación estándar se han revelado como discriminadores efectivos para diferenciar entre la precipitación y las señales no meteorológicas. Asimismo, se ha aprovechado la limitada extensión vertical de los ecos de tierra para corregir la reflectividad y mejorar la clasificación.

Las variables espaciales muestran una clara dependencia del rango, lo que hace necesaria su corrección. Las pruebas realizadas han demostrado que omitir esta corrección tiene un impacto significativo en la eficiencia del clasificador.

En combinación con la interpolación, se recomienda aplicar un filtro que reemplace cada valor por el promedio de los valores dentro de una ventana de tamaño predefinido. Estos dos procesos, además de aumentar la eficiencia del clasificador, ayudan a eliminar valores atípicos y a reducir los efectos del ruido.

La comparación de los histogramas de una variable para cada fenómeno (precipitación, desorden) ha revelado que el área de superposición de los histogramas está inversamente relacionada con la capacidad de la variable para distinguir diferentes tipos de fenómenos meteorológicos. Este criterio se utiliza para seleccionar las mejores variables al desarrollar un clasificador.

En este trabajo, se presenta una metodología que permite la construcción de un clasificador difuso. Los errores de entrenamiento y de validación obtenidos para este clasificador fueron bajos y prácticamente iguales.

La construcción de un clasificador adaptado a las condiciones de la región donde opera el radar es un proceso empírico que requiere datos de entrenamiento previamente clasificados por un experto. La revisión de trabajos realizados por diversos autores en diferentes países de Europa revela que en general se utilizan las mismas variables, aunque los umbrales pueden variar ligeramente. Estas diferencias sutiles en los umbrales de las variables permiten que los resultados se ajusten a las condiciones específicas de la región donde se encuentra el radar. Sin embargo, este estudio también subraya la sensibilidad del etiquetado de un clasificador ante pequeñas variaciones en las funciones de distribución, lo que puede llevar a clasificaciones muy diferentes. Por lo tanto, construir un clasificador basado en investigaciones realizadas en regiones con condiciones climáticas distintas a Colombia resulta inapropiado. Es fundamental llevar a cabo estudios que generen datos de entrenamiento válidos en Colombia para construir clasificadores que se adapten a las condiciones climatológicas específicas de las diferentes regiones del país.

## 10. Trabajo futuro

Para investigaciones futuras, se propone desarrollar un modelo para identificar y suprimir el desorden meteorológico. Esto puede lograrse aprovechando características no exploradas en este trabajo, como el espectro estrecho y la variabilidad temporal de las señales radar. El fundamento es que el desorden tiene un tiempo de correlación considerablemente mayor al de los fenómenos atmosféricos, permitiendo distinguirlos. Otra línea de trabajo es implementar la clasificación Precipitación/Eco con algoritmos de aprendizaje automático avanzados como máquinas de vectores de soporte o redes neuronales. Estos modelos podrían explotar eficientemente las relaciones no lineales en los datos polarimétricos para mejorar el desempeño.

## 11. Agradecimientos

A la Aeronáutica Civil, la entidad que proporcionó los datos del radar de Coroza, que constituyen el componente fundamental de este proyecto.

## Referencias

- [1] Raubert Robert M. and Nesbitt Stephen W. “Radar Meteorology, A First Course”. Wiley, 2014.
- [2] Florenci. Rey Benadero, Fundación Española para la Ciencia y la Tecnología., y Semana de la Ciencia y la Tecnología (2004. Madrid), Meteorología y climatología: unidad didáctica. Fundación Española para la Ciencia y la Tecnología, 2004.
- [3] R. J. Doviak and R. D. Palmer, “Radar: Polarimetric Doppler Weather Radar,” in *Encyclopedia of Atmospheric Sciences: Second Edition*, Elsevier Inc., 2015, pp. 444–454. doi: 10.1016/B978-0-12-382225-3.00329-7.
- [4] M. Gabella and R. Notarpietro, “Ground clutter characterization and elimination in mountainous terrain,” *Proceedings of ERAD*, pp. 305–311, 2002, [Online]. Available: [http://www.einfalt.de/literature/Gabel\\_Groun.pdf](http://www.einfalt.de/literature/Gabel_Groun.pdf)
- [5] M. A. (Mark A. ) Richards, J. Scheer, and W. A. Holm, “Principles of modern radar”. Volume I, Basic principles. SciTech Pub, 2010.
- [6] J. C. Hubbert, M. Dixon, S. M. Ellis, and G. Meymaris, “Weather radar ground clutter. Part I: Identification, modeling, and simulation,” *J Atmos Ocean Technol*, vol. 26, no. 7, pp. 1165–1180, 2009.
- [7] M. H. Golbon-Haghighi, G. Zhang, Y. Li, and R. J. Doviak, “Detection of ground clutter from Weather radar using a dual-polarization and dual-scan method,” *Atmosphere (Basel)*, vol. 7, no. 6, pp. 1–11, 2016, doi: 10.3390/atmos7060083.
- [8] M. R. Kumjian, “Principles and Applications of Dual-Polarization Weather Radar. Part I: Description of the Polarimetric Radar Variables,” *J Operat Meteorol*, vol. 1, no. 19, pp. 226–242, 2013, doi: 10.15191/nwajom.2013.0119.
- [9] R. H. V.N. Bringi, M. Thurai, *Dual-Polarization Weather Radar Handbook*, 2nd ed. Neuss, Germany: Gematronik GmbH, Neuss, Germany, 2007.
- [10] D. Charalampidis, T. Kasparis, and W. L. Jones, “Removal of nonprecipitation echoes in weather radar using multifractals and intensity,” *IEEE Transactions on Geoscience and*

- Remote Sensing, vol. 40, no. 5, pp. 1121–1131, 2002, doi: 10.1109/TGRS.2002.1010899.
- [11] D. Moisseev, C. Unal, H. Russchenberg, and L. Ligthart, “A new method to separate ground clutter and atmospheric reflections in the case of similar Doppler velocities,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 239–246, 2002, doi: 10.1109/36.992778.
- [12] M. Berenguer, D. Sempere-Torres, C. Corral, and R. Sánchez-Diezma, “A fuzzy logic technique for identifying nonprecipitating echoes in radar scans,” *J Atmos Ocean Technol*, vol. 23, no. 9, pp. 1157–1180, 2006, doi: 10.1175/JTECH1914.1.
- [13] “Artificial Intelligence Methods in the Environmental Sciences.”
- [14] Peckol James K, “Introduction to Fuzzy Logic,” 2021.
- [15] Girraj Singh, “A Review: Fuzzy Logic and Its Application,” 2014.
- [16] L. A. Zadeh, “fuzzy logic,” 1988.
- [17] L. A. Zadeh, “Fuzzy Sets,” *University of California, Berkeley*, vol. 8, pp. 338–353, 1965.
- [18] Richard J. Oentaryo, “Automated driving based on self organizing GenSoYager neuro-fuzzy system,” *QuantumBlack, McKinsey & Company*.
- [19] D. Reina, “Fundamentos de Matemática Difusa,” 2008.
- [20] E. Trillas and L. Eciolaza, “Studies in Fuzziness and Soft Computing Fuzzy Logic An Introductory Course for Engineering Students.” [Online]. Available: <http://www.springer.com/series/2941>
- [21] A. Bonarini, “An Introduction to Learning Fuzzy Classifier Systems”. doi: Bonarini, A. An Introduction to Learning Fuzzy Classifier Systems. *Learning Classifier Systems*, pp. 83–104, 2000.
- [22] J. Vivekanandan, D. S. Zrníc, S. M. Ellis, R. Oye, A. V Ryzhkov, and J. Straka, “Cloud Microphysics Retrieval Using S-Band Dual-Polarization Radar Measurements”, pp. 381–388, 1999.
- [23] J. J. Gourley, P. Tabary, and J. Parent du Chatelet, “A fuzzy logic algorithm for the separation of precipitating from non-precipitating echoes using polarimetric radar observations,” *J Atmos Ocean Technol*, vol. 24, no. 8, pp. 1439–1451, 2007, doi: 10.1175/JTECH2035.1.
- [24] M. A. Rico-Ramirez and I. D. Cluckie, “Classification of ground clutter and anomalous propagation using dual-polarization weather radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 1892–1904, 2008, doi: 10.1109/TGRS.2008.916979.
- [25] M. Hedir and B. Haddad, “Ground echoes filtering using the completed local binary pattern and the support vector machine,” *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)*, pp. 1–4, 2015, doi: 10.1109/Multi-Temp.2015.7245777.
- [26] J. Nicol and A. Illingworth, “Techniques for improving ground clutter identification,” *Proc. Symp. Weather ...*, no. April, pp. 1–6, 2011, [Online]. Available: [http://cree.rdg.ac.uk/~radar/publications/WRaH\\_025\\_Nicol\\_rev.pdf](http://cree.rdg.ac.uk/~radar/publications/WRaH_025_Nicol_rev.pdf)
- [27] Jaroslav Drapala, “Kernel Density Estimator explained step by step,” *Towards Data Science*.
- [28] Jason Brownlee, “Train-Test Split for Evaluating Machine Learning Algorithms,” *Machine Learning Mastery*.
- [29] Jason Brownlee, “A Gentle Introduction to Cross-Entropy for Machine Learning,” *Machine Learning Mastery*.
- [30] N. Ketkar, “Stochastic Gradient Descent,” in *Deep Learning with Python*, Apress, pp. 113–132, 2017.
- [31] M. Steiner and J. A. Smith, “Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data,” *J Atmos Ocean Technol*, vol. 19, no. 5, pp. 673–686, 2002.
- [32] M. Grecu and W. F. Krajewski, “An Efficient Methodology for Detection of Anomalous Propagation Echoes in Radar Reflectivity Data Using Neural Networks,” 2000.