

Clasificación de imágenes de radar usando una red neuronal convolucional

Nota Técnica SMN 2024-173

Paula Dunayevich¹, Federico Cutraro¹, Martin Rugna¹, Maximiliano Sacco¹ y Luciano Vidal¹

¹ *Dirección de Productos de Modelación Ambiental y de Sensores Remotos*

Junio 2024

Información sobre Copyright

Este reporte ha sido producido por empleados del Servicio Meteorológico Nacional con el fin de documentar sus actividades de investigación y desarrollo. El presente trabajo ha tenido cierto nivel de revisión por otros miembros de la institución, pero ninguno de los resultados o juicios expresados aquí presuponen un aval implícito o explícito del Servicio Meteorológico Nacional.

La información aquí presentada puede ser reproducida a condición que la fuente sea adecuadamente citada.

Resumen

Las observaciones meteorológicas, especialmente aquellas obtenidas mediante radares, desempeñan un papel fundamental en la comprensión y el pronóstico de los fenómenos atmosféricos. Este tipo de tecnología resulta crucial a la hora de monitorear los sistemas meteorológicos que pueden tener asociados fenómenos severos. Los radares tienen la capacidad de detectar diversas señales, como las generadas por aves, insectos, murciélagos y el terreno, agrupadas bajo la categoría de “ecos no meteorológicos”. Estas señales son útiles para determinar regiones de interés meteorológico, pero otras aplicaciones pueden considerarlas como ruido y por esta razón resulta de interés distinguirlas de los ecos meteorológicos. Es por eso que en este trabajo nos proponemos utilizar una red neuronal convolucional para segmentar y clasificar imágenes de radar en ecos meteorológicos, ecos no meteorológicos y N/D (No Dato). Para realizar una evaluación inicial del modelo, utilizamos un dataset que contiene imágenes pertenecientes a 5 radares del sistema NEXRAD ubicados en una región de Estados Unidos que presenta similitudes meteorológicas con la República Argentina. Los primeros resultados indican que la elección de una red neuronal convolucional con una arquitectura U-Net fue acertada a la hora de abordar este tipo de problema, tal y como indican el accuracy y f1-score obtenidos para el modelo. Asimismo, sugieren que nuestros esfuerzos futuros deben concentrarse en ampliar el dataset de manera tal de obtener una mayor representación de píxeles clasificados como ecos meteorológicos y no meteorológicos, y en disminuir la clasificación de ecos no meteorológicos como N/D.

Abstract

Meteorological observations, especially those obtained by radar, play a fundamental role in the understanding and forecasting of atmospheric phenomena. This type of technology is crucial when monitoring weather systems that may have severe phenomena associated with them. Radars have the ability to detect various types of signals such as those generated by birds, insects, bats and terrain, grouped under the category "non-meteorological echoes". These signals are useful for determining regions of meteorological interest, but other applications may consider them as noise and for this reason it is of interest to distinguish them from meteorological echoes. That is why in this work we propose to use a convolutional neural network to segment and classify radar images into meteorological echoes, non-meteorological echoes and N/D (No Data). To perform an initial evaluation of the model, we use a dataset containing images belonging to 5 radars of the NEXRAD system located in a region of the United States that presents meteorological similarities with the Argentine Republic. The first results indicate that the choice of a convolutional neural network with a U-Net architecture was the right one to address this type of problem, as indicated by the accuracy and f1-score obtained for the model. They also suggest that our future efforts should concentrate on enlarging the dataset in order to obtain a larger representation of pixels classified as meteorological and non-meteorological echoes, and on decreasing the classification of non-meteorological echoes as N/D.

Palabras clave: radares meteorológicos, clasificación, segmentación, aprendizaje supervisado, red neuronal convolucional

Citar como:

Dunayevich, P., F. Cutraro, M. Rugna, M. Sacco y L. Vidal, 2024: Clasificación de imágenes de radar usando una red neuronal convolucional. Nota Técnica SMN 2024-173.

1. INTRODUCCIÓN

Las observaciones meteorológicas son de vital importancia para entender y prever los fenómenos atmosféricos. La perspectiva ofrecida por la tecnología de radar resulta crucial a la hora de realizar el diagnóstico y monitoreo de los sistemas que generan precipitación, así como de los posibles fenómenos severos asociados. Por esta razón, el radar se ha convertido en un instrumento clave en el pronóstico meteorológico a corto plazo (Fabry, 2018).

Los radares (del acrónimo en inglés de RAdio Detection And Ranging) son instrumentos que emiten señales de ondas electromagnéticas (radio o microondas) (Fabry, 2018). Al encontrarse con un objeto, como puede ser algún tipo de precipitación, parte de la señal se refleja nuevamente hacia la antena. El radar analiza la señal para determinar la intensidad y ubicación de la precipitación. Asimismo, el radar puede determinar su velocidad y dirección.

Las señales relacionadas con eventos meteorológicos se conocen como ecos meteorológicos (Fabry, 2018). Estos pueden dividirse en dos categorías principales: targets meteorológicos precipitantes (como lluvia, llovizna, nieve, granizo, etc) y targets meteorológicos no precipitantes (que incluyen nubes de hielo y nubes de agua entre otros). Por otro lado, los radares tienen la capacidad de detectar diversas señales no vinculadas con eventos meteorológicos, como las generadas por aves, insectos, murciélagos y ecos provenientes del terreno, entre otros. Estas señales se agrupan bajo la categoría de ecos no meteorológicos. Algunos ecos no meteorológicos son útiles desde el punto de vista meteorológico. Sin embargo, determinados tipos de ecos no meteorológicos en determinadas situaciones pueden ser considerados como interferencia e introducir ruido e incertidumbre en el análisis y deben ser filtrados para interpretar correctamente los datos de radar.

Se han utilizado distintas técnicas para tratar de filtrar este tipo de señales. Rico-Ramirez y Cluckie (2008) evaluaron el desempeño de clasificadores basados en lógica difusa y Bayes para detectar ecos de terreno usando radares meteorológicos; Berenguer y otros (2006) presentaron un algoritmo basado en lógica difusa que utiliza datos volumétricos para producir un valor que cuantifica la posibilidad de que cada celda sea afectada por el eco de terreno; Suárez y otros (2019) aplicaron cuatro métodos de clasificación supervisada (Modelo lineal, Modelo Cuadrático, Modelo Logístico y Modelo de Bayes Naive) para discriminar ecos meteorológicos de ecos no meteorológicos usando datos de radar; por mencionar algunos.

Las metodologías empleadas en estos trabajos exhibieron un muy buen rendimiento en cuanto a la tarea de clasificación propuesta. No obstante, consideramos que la aplicación de técnicas de inteligencia artificial, capaces de capturar la estructura espacial de los datos y de incorporar un extenso conjunto de variables en el análisis, podría mejorar aún más la clasificación de ecos de radar. Es por eso que en este trabajo nos proponemos utilizar técnicas de inteligencia artificial para clasificar ecos meteorológicos y no meteorológicos utilizando datos de radar.

Los modelos de aprendizaje automático (machine learning, en inglés), especialmente las redes neuronales convolucionales (CNN, por sus siglas en inglés) inspiradas en el reconocimiento de imágenes, tienen un gran potencial para aprender patrones y características complejas directamente de datos de radar. Entre una gran variedad de ejemplos, podemos mencionar los siguientes trabajos: Geiss y Hardin (2020) usaron CNNs para incrementar artificialmente la resolución de escaneos PPI de NEXRAD; Han y otros (2020) desarrollaron una red neuronal convolucional para el pronóstico de tormentas convectivas utilizando datos tridimensionales de radares meteorológicos, transformando el problema en una clasificación binaria; Albu y otros (2022) desarrollaron una red neuronal convolucional para el pronóstico del tiempo basada en la predicción de productos de radar, es decir, mapeando las mediciones de radar del pasado a las mediciones de radar que se registran en el futuro; Zhang y otros (2020) combinaron una red neuronal convolucional con memoria a corto y largo plazo bidireccional para predecir lluvia usando mapas de ecos de radar; Ayzel y otros (2020) desarrollaron una red neuronal convolucional inspirada en las familias de modelos de

aprendizaje profundo U-Net y SegNet para el pronóstico de precipitaciones basado en radar; [Pavlik y otros \(2022\)](#) usaron una red neuronal convolucional 3D con arquitectura U-Net para realizar un pronóstico de precipitaciones usando datos de radar; [Lu y Kumar \(2019\)](#) usaron una red neuronal convolucional para clasificar hidrometeoros en cuatro categorías: Cristales de Hielo, Nieve Seca, Lluvia Ligera y/o Moderada y Gotas Grandes (Lluvia).

Luego, en este trabajo nos proponemos utilizar técnicas de inteligencia artificial para clasificar ecos meteorológicos y no meteorológicos utilizando datos de radar. En particular, decidimos implementar una red neuronal convolucional para clasificar imágenes de radar en ecos meteorológicos, no meteorológicos y N/D (No Data) usando como datos de entrada 6 variables: la reflectividad, el ancho espectral, la velocidad Doppler, la reflectividad diferencial, el coeficiente de correlación copolar y el diferencial de fase específico. En este trabajo mostramos resultados preliminares obtenidos utilizando un dataset reducido con el fin de realizar una evaluación inicial del modelo elegido. La Sección 2 describe los datos utilizados para entrenar la red neuronal; la Sección 3, la metodología adoptada en este trabajo; la Sección 4, los resultados obtenidos hasta el momento; y la Sección 5, las conclusiones del trabajo.

2. DATOS

Con el propósito de llevar a cabo una evaluación preliminar del modelo, elaboramos un dataset que contiene 40 imágenes pertenecientes a 5 radares que forman parte del sistema Next Generation Weather Radar (NEXRAD) [[NEXRAD Homepage](#), última vez accedido 2024/05/31], una red de radares meteorológicos Doppler de banda S de los Estados Unidos. Los radares seleccionados (Fig. 1), KABX, KDMX, KFTG, KFWS y KRIW, se encuentran en una región de Estados Unidos que presenta similitudes meteorológicas con la República Argentina, de manera tal que el aprendizaje de la red neuronal sea extrapolable a datos locales de la red SINARAME ([Giménez y otros, 2021](#)).

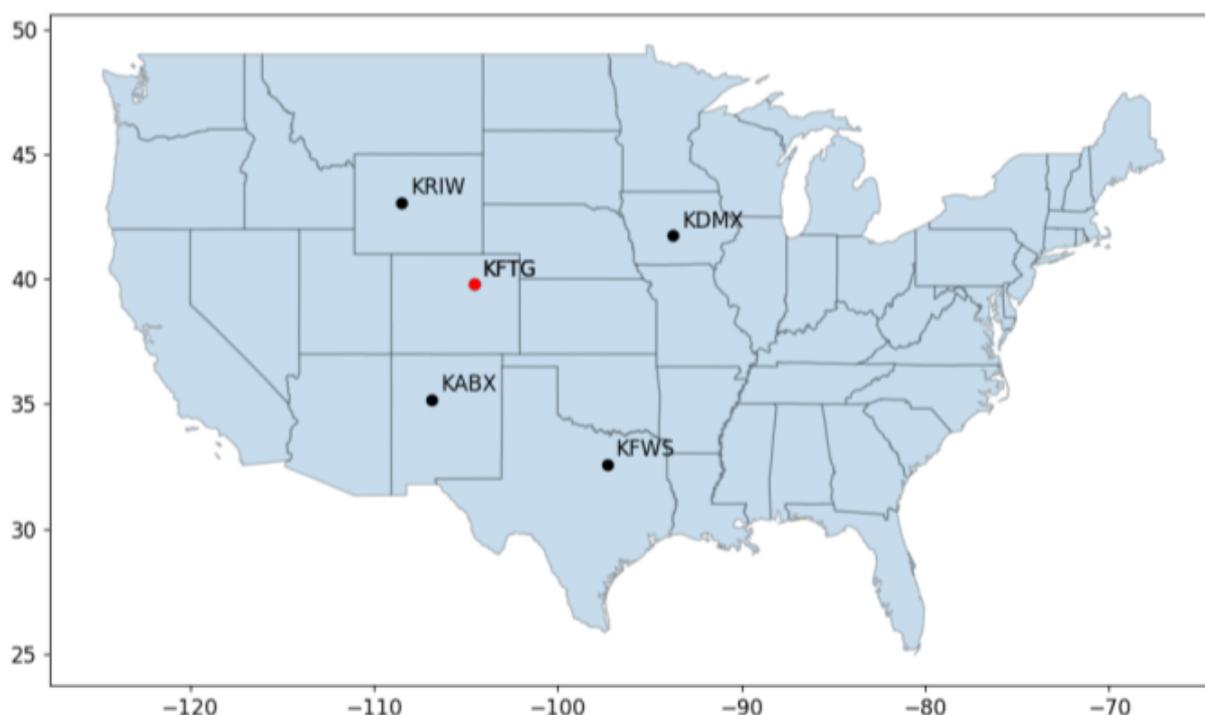


Figura 1: Radares de la red NEXRAD que forman parte del dataset original. En negro, los radares utilizados para entrenar la CNN. En rojo, el radar utilizados para testear la CNN.

Seleccionamos un evento por año para cada radar, para 4 fechas distintas (correspondientes a los años 2020, 2021, 2022 y 2023). En este contexto, el término "evento" no se refiere a un episodio meteorológico específico completo (que podría estar representado por decenas o centenas de imágenes consecutivas), sino a una cantidad fija de imágenes representativas de dicho episodio. Los tipos de eventos incluidos en nuestro conjunto de datos son granizo, nieve y lluvia. Para cada evento, elegimos 3 imágenes, pero utilizamos solo la primera y la última para entrenar la red neuronal.

La Figura 2 muestra un esquema del funcionamiento de los radares meteorológicos en general. El objetivo del radar es detectar targets meteorológicos, como la lluvia o el granizo, que reflejan parte de la energía de las señales de radar emitidas. Para ello, la antena de un radar meteorológico emite pulsos de microondas que parten de una determinada altura sobre el nivel del suelo (ha). Cuando el modo de exploración del radar es el denominado PPI (por las siglas en inglés de Plan Position Indicator), el radar mantiene fijo el ángulo de elevación (ángulo del haz con respecto al suelo) pero varía el ángulo azimuth (ángulo del haz con respecto al norte), es decir, escanea a una determinada altura recorriendo los 360° correspondientes para luego cambiar la altura de elevación y repetir el escaneo. El ancho del haz aumenta con la distancia al radar (R , por rango o range, en inglés), lo cual influye en la resolución espacial del mismo, afectando su capacidad para detectar blancos con precisión. En función del ángulo de elevación, los blancos detectados se encuentran a una altura respecto del nivel del suelo (H) que aumenta con la distancia al radar. Por último, el ancho del pulso (h) es la duración del pulso de microondas, que determina la sensibilidad y resolución del radar.

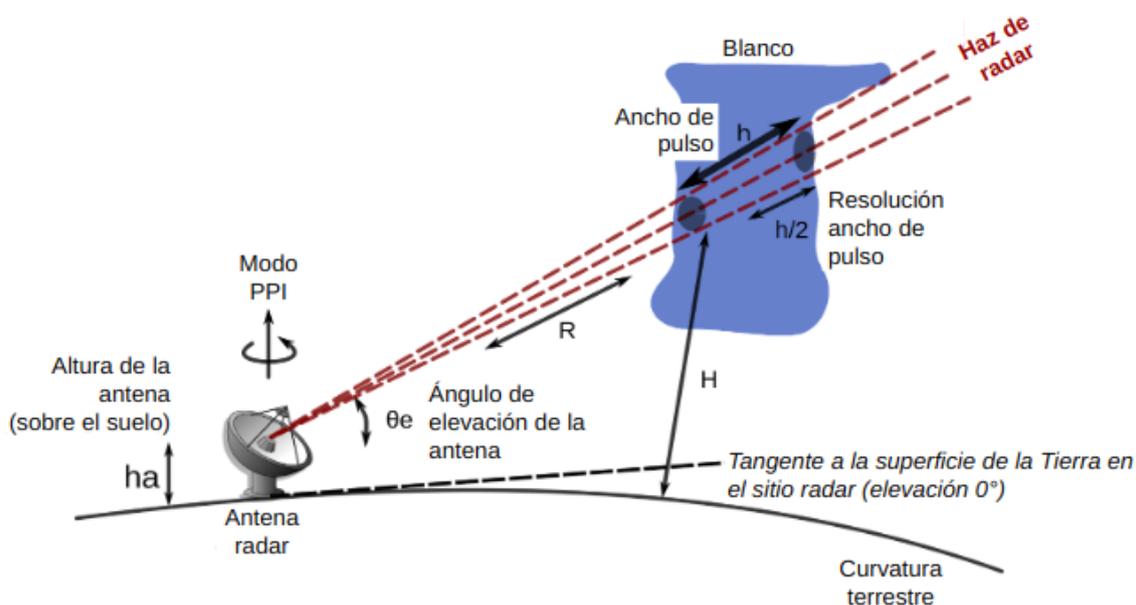


Figura 2: Esquema del funcionamiento de un radar meteorológico (Fuente: <https://commons.wikimedia.org/w/index.php?curid=4177376>).

Las variables con las que vamos a trabajar son la reflectividad (Digital Base Reflectivity), el ancho espectral (Base Spectrum Width), la velocidad Doppler (Digital Base Velocity), la reflectividad diferencial (Differential Reflectivity), el coeficiente de correlación copolar (Correlation Coefficient) y el diferencial de fase específico (Specific Differential Phase) como input y la clasificación de hidrometeoros (Hydrometeor Classification) modificada (ver sección Metodología) como output de la red neuronal. El ángulo de elevación aproximado en todos los casos es 0.5° . A continuación, damos una breve descripción de las variables:

1. **Reflectividad:** La reflectividad (Z) se define como el retorno de potencia promedio en el canal horizontal de todos los objetivos en el volumen de resolución y posee unidades de "dBZ". La reflectividad depende del diámetro y la concentración de la precipitación escaneada. En otras palabras, gotas grandes y/o un gran número de gotas darán lugar a una reflectividad alta.
2. **Reflectividad diferencial:** La reflectividad diferencial, o ZDR, es la diferencia entre los factores de reflectividad horizontal y vertical en unidades "dB".
3. **Coefficiente de correlación copolar:** El coeficiente de correlación copolar (CC) indica qué tan similarmente se comportan los pulsos polarizados horizontal y verticalmente de pulso a pulso dentro de un volumen de resolución. Esta variable no posee unidades.
4. **Diferencial de fase específico:** El diferencial de fase (PhiDP) de la cual deriva el diferencial de fase específico (KDP) es una medida que describe la diferencia entre las fases de los pulsos de microondas polarizados horizontalmente y verticalmente. KDP es la derivada con respecto al rango de PhiDP y tiene unidades de grados/km.
5. **Velocidad Doppler:** La velocidad doppler (V) se define como la velocidad radial media ponderada por la potencia de todos los objetivos en un volumen de resolución y posee unidades de kts (nudo o knots en inglés, siendo un nudo una milla náutica por hora, que es aproximadamente 1.852 kilómetros por hora). Sólo mide la componente radial del viento. En otras palabras, no es una medida real del viento a menos que esté orientado radialmente.
6. **Ancho espectral:** El ancho del espectro se define como la dispersión de la velocidad dentro de un volumen de resolución. Es decir, el ancho del espectro muestra la variabilidad de los movimientos individuales de los objetivos dentro de un volumen de resolución. De la misma manera que V, tiene unidades de kts.
7. **Clasificación de hidrometeoros:** El algoritmo de clasificación de hidrometeoros o HCA, se propone determinar la clasificación más probable de todos los bins que el radar detecta utilizando datos base y variables polarimétricas. Las distintas categorías de clasificación de ecos son: Below Threshold (ND), Biological (BI), Anomalous Propagation/Group Clutter (GC), Ice Crystals (IC), Dry Snow (DS), Wet Snow (WS), Light and/or Moderate Rain (RA), Heavy Rain (HR), Big Drops (rain) (BD), Graupel (GR), Hail, possibly with rain (HA), Large Hail (LH), Giant Hail (GH), Unknown Classification (UK) y Range Folded (RH).

3. METODOLOGÍA

El proceso de trabajo se puede dividir en cuatro grandes etapas, las cuales se describen detalladamente a continuación: [3.1 Elección del tipo y arquitectura de la red neuronal](#), [3.2 Preprocesamiento de datos](#), [3.3 Optimización de hiperparámetros](#), y [3.4 Entrenamiento de la red neuronal](#).

3.1. Elección del tipo y arquitectura de la red neuronal

Para realizar la clasificación de imágenes de radar empleamos un enfoque de aprendizaje supervisado, es decir, trabajamos con un modelo entrenado utilizando un conjunto de datos etiquetado. Dado que el objetivo principal de la red neuronal incluye en primer lugar la segmentación de imágenes, hemos optado por utilizar una red neuronal convolucional para llevar a cabo esta tarea ([Ghanbari y otros, 2021](#)).

La arquitectura que empleamos como punto de partida en este trabajo es la U-Net ([Ronneberger y otros, 2015](#)), una red neuronal convolucional originalmente diseñada para realizar tareas de segmentación de imágenes biomédicas que tuvo buen desempeño habiendo sido entrenada con un número bajo de datos. La arquitectura consta de un camino de contracción para capturar el contexto y un camino de expansión simétrico que permite una localización precisa. La implementación de la U-Net está basada en el código desarrollado por Vidushi Bhatia (<https://github.com/VidushiBhatia/U-Net-Implementation>), implementado utilizando TensorFlow ([Abadi y otros, 2015](#)).

3.2. Preprocesamiento de datos

Como mencionamos anteriormente, vamos a usar como input de la red neuronal la reflectividad, el ancho espectral, la velocidad Doppler, la reflectividad diferencial, el coeficiente de correlación copolar y el diferencial de fase específico. Como etiquetas, utilizaremos la clasificación de hidrometeoros generada por un algoritmo computacional, la cual asigna a cada píxel una categoría de hidrometeoros entre las disponibles para los radares de la red NEXRAD. Sin embargo, dado que este trabajo pretende realizar una clasificación en 3 categorías (eco meteorológico, eco no meteorológico y N/D), las categorías de clasificación usado por la red NEXRAD fueron modificadas: dentro de ecos meteorológicos (M) quedan incluidas las categorías Ice Crystals, Dry Snow, Wet Snow, Light and/or Moderate Rain, Heavy Rain, Big Drops (rain), Graupel, Hail, possibly with rain, Large Hail y Giant Hail; dentro de ecos no meteorológicos (NM) se encuentran las clases Biological y Anomalous Propagation/Group Clutter; incluidos en N/D se identifican Below Threshold, Unknown Classification y Range Folded.

El preprocesamiento de los datos, tanto los empleados como input como los que forman parte del output de la red neuronal, consistió en recortar la imágenes de manera tal de que tengan un rango de 150 km (para evitar la isoterma de cero grados); de ser necesario cambiar artificialmente la resolución de los datos, ya que los arrays de las distintas variables deben tener el mismo tamaño para entrenar la red neuronal; y reordenar los datos según el azimuth. Por último, identificar y filtrar valores atípicos para la variable Specific Differential Phase y reemplazar outliers y FillValues por N/Ds para todas las variables.

3.3. Optimización de hiperparámetros

En esta etapa, se analizan diferentes combinaciones de hiperparámetros para encontrar los que den lugar a la red neuronal que mejor ajuste los datos de la red NEXRAD. Luego de dividir el conjunto de datos en entrenamiento (train set) y testeo (testing set) (usando los datos pertenecientes al radar KFTG para el conjunto de evaluación) buscamos los hiperparámetros óptimos del modelo empleando una estrategia llamada validación cruzada en k -iteraciones (*k-fold cross-validation*, en inglés) ([Raschka, 2020](#)), que consiste en iterar sobre un conjunto de datos (del entrenamiento o train set) k -veces. En cada iteración, dividimos el conjunto de datos en k -partes: una parte se utiliza para la validación y las $k-1$ partes restantes se utilizan como conjunto de entrenamiento (train set). Al finalizar, se calcula el rendimiento de la validación cruzada como la media aritmética de las k -estimaciones de rendimiento de los conjuntos de validación.

En este caso, utilizamos una variante en particular de esta técnica denominada *object-based spatial leave-one-out cross-validation* (validación cruzada espacial basada en objetos dejando un objeto fuera) que se encuentra detallada en el trabajo de [Karasiak y otros \(2021\)](#). Esta técnica consiste en dejar en cada iteración todos los datos correspondientes a un radar (objeto) como datos de validación y utilizar los datos del resto de los radares como conjunto de entrenamiento. Nótese que este es un caso particular de *k-fold cross-validation* en el cual $k = n$, siendo

n el número de objetos (en este trabajo, radares) del dataset. De esta manera logramos evitar que la autocorrelación espacial de los datos dé lugar a una falta de independencia de los conjuntos de entrenamiento y validación.

En cada iteración, es necesario normalizar los datos del input para llevar los valores de las distintas variables a una escala similar (Géron, 2019), de manera tal que todas contribuyan de forma equitativa al modelo; reemplazar los N/Ds por los valores que corresponda (-1 para los datos input y 2 para los datos output), ya que el modelo no acepta valores faltantes; suavizar los datos del output para tratar de representar más fielmente la realidad (de lo contrario los datos del output presentarán píxeles aislados o grupos de pocos píxeles de alguna de las categorías de clasificación que posiblemente constituyan errores en la clasificación original); y aplicar one-hot-encoding de manera tal de transformar las variables categóricas (ecos meteorológicos, ecos no meteorológicos y valores desconocidos) en variables numéricas. Luego, para cada conjunto de hiperparámetros que deseamos evaluar, compilamos la red neuronal y la entrenamos guardando la precisión (más conocida como accuracy, en inglés) que se calcula como el porcentaje de datos clasificados correctamente del total de datos y el valor de la función de costo (loss, en inglés) del train y del validation set.

Para testear el código en este trabajo vamos a fijar los siguientes hiperparámetros: la probabilidad asociada a la técnica de regularización dropout (0.3), el número de filtros en las capas convolucionales (32 para la primer convolución) y la cantidad de épocas utilizadas durante el entrenamiento de la red neuronal (6 como máximo). Por otro lado, vamos a testear dos valores para el tamaño de lote o *batch size* (2 y 4) y dos para la tasa de aprendizaje o learning rate (0.01 y 0.001). En trabajos futuros, planeamos ampliar el conjunto de datos y explorar una gama más amplia de hiperparámetros.

En la Figura 3 podemos ver el resultado de la optimización de hiperparámetros, en particular, el accuracy de la training y validation data correspondiente a cada una de las 6 épocas para cada conjunto de hiperparámetros y para cada iteración (*fold*). Las curvas que se interrumpen antes de alcanzar las 6 épocas lo hicieron debido a la aplicación de la técnica de early stopping (que detiene el entrenamiento cuando una métrica determinada, en este caso el loss del validation set, ha dejado de mejorar).

Para cada conjunto de hiperparámetros, calculamos el promedio y el desvío estándar del accuracy del train y del validation set de las 4 iteraciones correspondientes a la última época. Seleccionamos el mejor conjunto de hiperparámetros como el que posea el Validation accuracy promedio más alto.

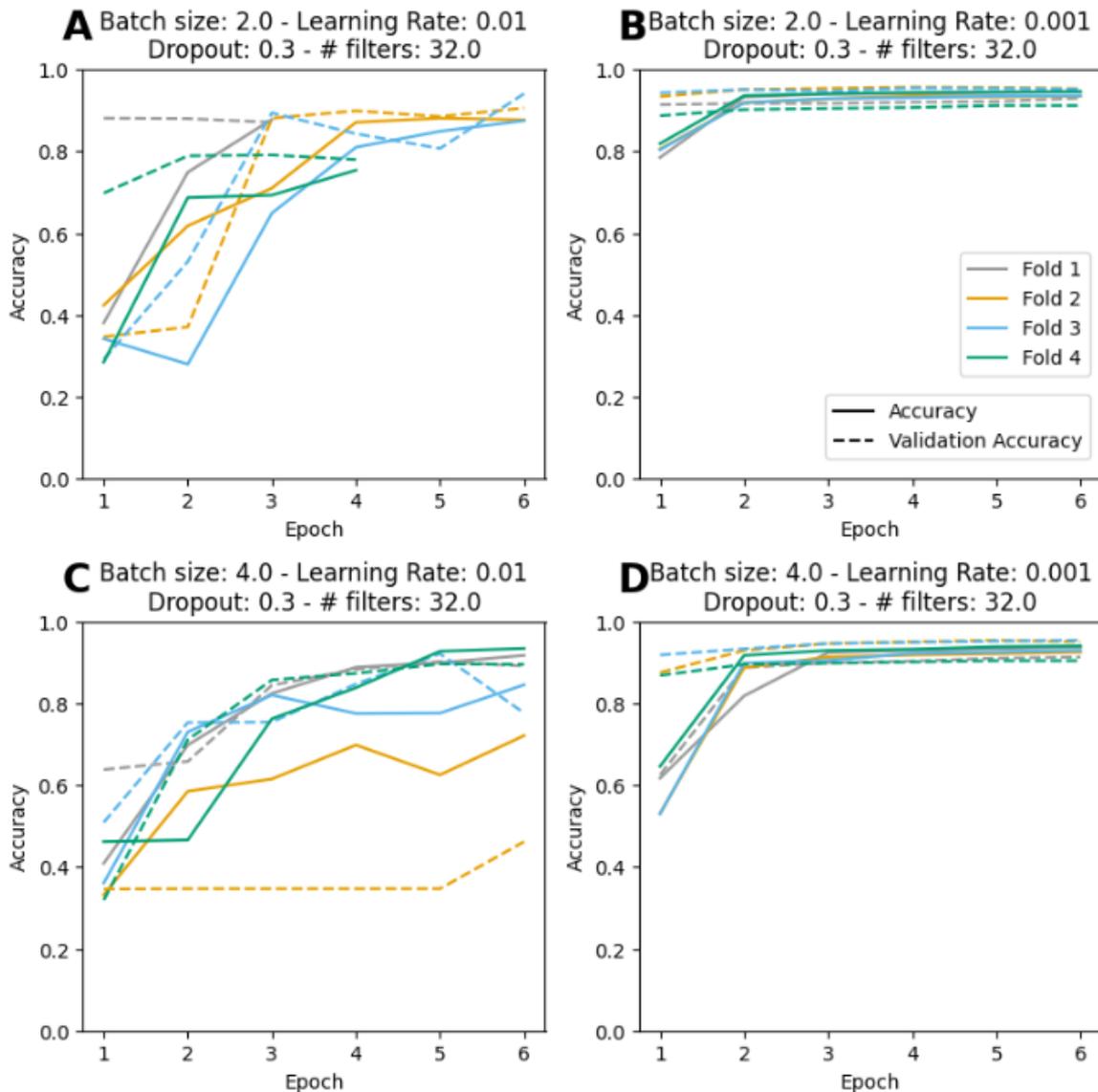


Figura 3: Accuracy del train (“Accuracy”) y del validation set (“Validation accuracy”) vs época para cada conjunto de hiperparámetros, coloreado por fold (iteración). En todos los casos la probabilidad de Dropout es 0.3 y el número de filtros inicial es 32. (A) y (B) Batch size = 2, (C) y (D) Batch size = 4. (A) y (C) Learning rate = 0.01, (B) y (D) Learning rate = 0.001.

3.4. Entrenamiento de la red neuronal

Por último, entrenamos la red neuronal con los hiperparámetros óptimos encontrados en la etapa anterior: Batch size: 2, Learning rate: 0.001, Dropout: 0.3 y cantidad de filtros: 32. Para el entrenamiento, utilizamos el train set completo, es decir, el train y el validation set del *k-fold cross-validation*, y 16 épocas en lugar de 6. Por último, utilizamos el test set para evaluar el modelo e inferir su capacidad de generalización a datos nuevos.

4. RESULTADOS

Luego del entrenamiento del modelo con los hiperparámetros que dieron lugar al mejor ajuste, obtuvimos un Accuracy de 0.945 (para el conjunto de entrenamiento) y un Validation accuracy de 0.936 (para el conjunto de evaluación). En la Figura 4 podemos ver el accuracy del train y test set para 9 de las 16 épocas con las que decidimos entrenar el modelo (el entrenamiento se detuvo antes de llegar a las 16 épocas debido al *early stopping*). En ambos casos, el accuracy aumenta con las épocas hasta estabilizarse en un valor similar para el train y test set.

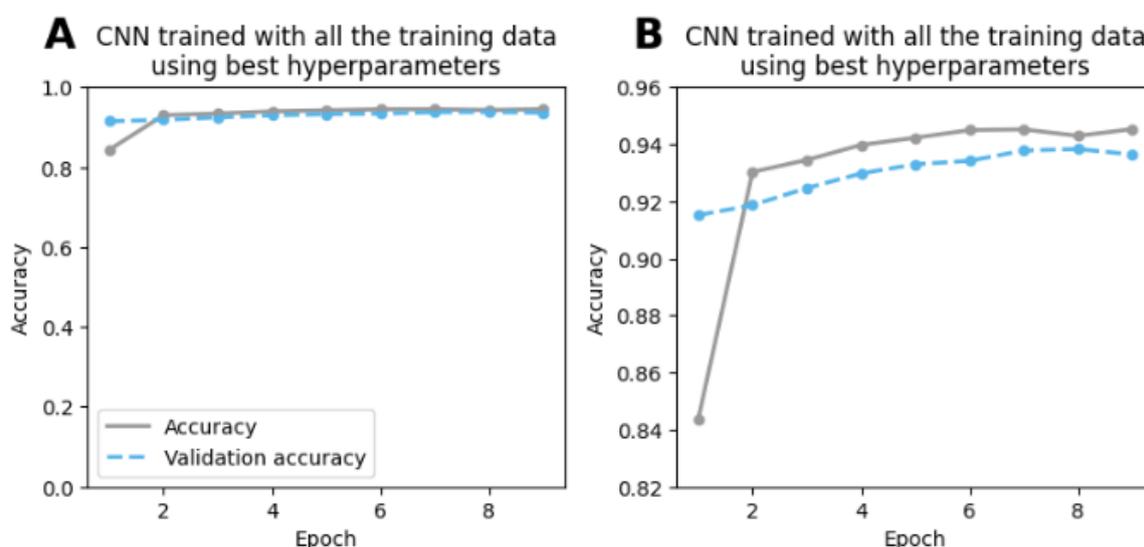


Figura 4: Accuracy (train set) y Validation accuracy (test set) para el modelo entrenado con los hiperparámetros que dan lugar al mejor ajuste de los datos. En (A) se muestran los datos y en (B) se amplía el gráfico para visualizar mejor la diferencia entre el train y el test set.

Para tener más información respecto al desempeño de la red neuronal, analizamos otras 3 métricas comúnmente utilizadas para evaluar clasificadores: la precisión, la sensibilidad y el f1-score (Géron, 2019). La precisión, también conocida como "success ratio", se calcula como:

$$\frac{TP}{TP + FP} \quad (1)$$

siendo TP el número de verdaderos positivos (por *true positives*, en inglés) y FP el número de falsos positivos (por *false positives*, en inglés). Este valor indica el porcentaje de datos clasificados correctamente en una cierta categoría en función del total de datos clasificados como pertenecientes a dicha categoría. La sensibilidad (o *recall*, en inglés), también conocida como "probabilidad de detección", se calcula como:

$$\frac{TP}{TP + FN} \quad (2)$$

siendo FN el número de falsos negativos (por *false negatives*, en inglés). Es por eso que el *recall* indica del total de píxeles que efectivamente pertenecen a una cierta categoría, cuántos fueron identificados como parte de ésta. Por último, el *f1-score*, es la media armónica de la precisión y la sensibilidad, es decir:

$$\frac{\textit{precision} \times \textit{sensibilidad}}{\textit{presicion} + \textit{sensibilidad}} \quad (3)$$

La Tabla 1 muestra estas métricas para el modelo en su conjunto, obtenidas utilizando scikit-learn (Pedregosa y otros, 2011).

Tabla 1: Precisión, recall y f1-score para el modelo, según el reporte de clasificación de scikit-learn. Macro indica la media de la métrica de cada una de las clases, y weighted la media ponderada de la métrica por clase.

	Precision	Recall	f1-score
macro	0.93	0.93	0.93
weighted	0.94	0.94	0.94

Podemos ver que las métricas son similares entre sí, es decir que el modelo es equilibrado en términos de su capacidad para identificar falsos positivos y falsos negativos.

Para comprender mejor el rendimiento del modelo en cuanto a la clasificación de cada categoría individualmente, calculamos las mismas métricas pero esta vez para cada una de las categorías de clasificación por separado (Tabla 2). Esto nos permitirá comprender si existe alguna variación significativa en su capacidad de identificar falsos positivos y falsos negativos en diferentes contextos.

Tabla 2: Precisión, recall y f1-score para las 3 categorías de clasificación del modelo (NM: ecos no meteorológicos, M: ecos meteorológicos y N/D: No dato) según el reporte de clasificación de scikit-learn.

	Precision	Recall	f1-score
NM	0.93	0.88	0.90
M	0.93	0.93	0.93
N/D	0.95	0.96	0.95

En la Tabla 2 podemos ver que los píxeles clasificados como N/D en general son efectivamente píxeles N/D (precisión de 0.95). Los datos clasificados como NM y M tienen un porcentaje levemente mayor de píxeles que fueron clasificados como FP. Por otro lado, observamos que el recall es más alto para la categoría N/D, seguido de la categoría M, y más bajo para la categoría NM. Es decir, del total de píxeles NM, hay una mayor proporción que fue clasificada como parte de otra categoría (FN). Esto da lugar a un valor del indicador f1-score mayor para la categoría N/D y menor para la categoría NM.

Para saber en qué categoría está clasificando el modelo erróneamente los píxeles NM, analizamos la matriz de confusión -conocida en otras aplicaciones como tabla de contingencia- para los datos del test set que está compuesto por los 8 datos correspondientes a 4 eventos para el radar KFTG, con dos imágenes para cada evento (Fig. 5). La matriz de confusión es una tabla que indica para cada categoría (True label) el porcentaje de píxeles asignados a cada categoría por el clasificador (Predicted label).

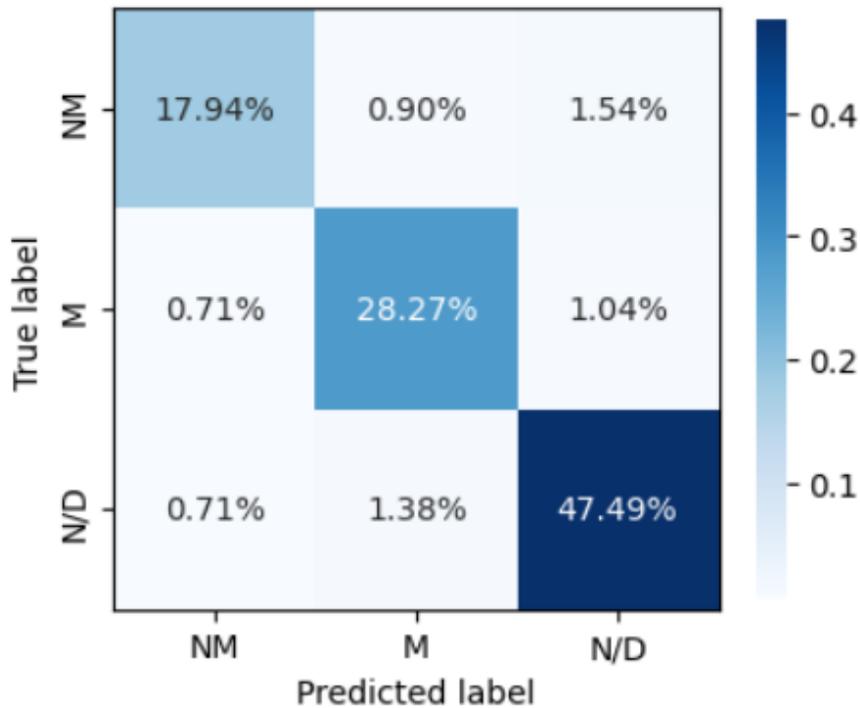


Figura 5: Matriz de confusión para los datos del test set.

Como podemos ver en la Figura 5, los píxeles bien clasificados (diagonal) representan un 93.7% del total de los píxeles. Los píxeles clasificados como M o NM pero que originalmente eran N/D, representan el 2.09% de los píxeles. Los píxeles mal clasificados, es decir los correspondientes a ecos meteorológicos clasificados como ecos no meteorológicos o a la inversa, representan el 1.61 % de los píxeles. Por último, los píxeles clasificados como N/D que originalmente eran M o NM, representan el 2.58 % de los píxeles totales.

Por otro lado, y en consonancia con lo discutido anteriormente, podemos ver que la categoría de error mayoritario del clasificador consiste en clasificar píxeles NM como N/D. Por último, con el fin de visualizar el tipo de errores cometidos por el clasificador en mayor detalle, seleccionamos 3 ejemplos del test set y graficamos los píxeles que presentaron diferencias en cuanto a su etiqueta real y predicha por el modelo (Fig. 6).

En la Figura 6 podemos ver que las diferencias que existen entre la clasificación de los datos hecha por el algoritmo del sistema NEXRAD y la predicha por el modelo resulta especialmente evidente en las zonas límite entre 2 categorías de clasificación.

5. CONCLUSIONES

En este apartado, nos interesa discutir sobre el desempeño de la red neuronal y su capacidad predictiva, siempre teniendo en cuenta que este estudio constituye un trabajo preliminar ya que no hemos empleado el conjunto de datos completo para entrenar la red neuronal ni hemos explorado el rango deseado de hiperparámetros durante su entrenamiento.

En primer lugar, queremos destacar que el desempeño de la red neuronal es adecuado para los usuarios, obtenién-

dose un Validation accuracy de 0.936, y un f1-score de 0.94 o 0.93 dependiendo de si los datos se ponderan por clase o no. Sin embargo, creemos que es posible mejorar el rendimiento general de la red neuronal.

Por un lado, es importante discutir sobre la distribución de las distintas categorías de clasificación presentes en el dataset. A pesar de que cada imagen de radar constituye en su conjunto un ejemplo de entrenamiento, realizaremos este análisis calculando el porcentaje total de píxeles correspondiente a cada categoría de clasificación. De esta manera, obtenemos un 25.04 % de píxeles correspondientes a la categoría NM, un 21.23 % a la categoría M y un 53.72 % a la categoría N/D. Las categorías NM y M están representadas aproximadamente a la mitad respecto de la categoría N/D. Que el conjunto de datos esté equilibrado puede o no ser importante para garantizar que el modelo pueda identificar con precisión todas las categorías. Esto depende del modelo, del tipo de datos y de la cantidad de datos con los que contamos. Analizando los valores presentados en la Tabla 2, encontramos que tanto la precisión como el recall y por lo tanto también el f1-score son mayores para la categoría N/D, que resulta mayoritaria en el dataset respecto de las categorías NM y M. Por lo tanto, es posible que cuando las categorías NM y M estén mejor representadas en el dataset, aumente la capacidad del modelo de discriminarlas. Esto resulta especialmente importante para disminuir la clasificación de ecos meteorológicos como no meteorológicos y a la inversa, lo cual posee una relevancia particular en el contexto de este proyecto.

Por otro lado, como mencionamos anteriormente, el recall de la categoría NM es significativamente menor que el de las otras dos categorías, lo cual indica que esta categoría tiene una mayor proporción de píxeles clasificados como falsos negativos. Al mirar la matriz de confusión notamos que el tipo de error que comete el clasificador en mayor medida es el de clasificar píxeles NM como N/D. Es por eso que nuestro trabajo futuro debe concentrarse particularmente en tratar de mejorar la identificación de ecos NM.

Un punto fundamental para el buen desempeño de la red neuronal y una correcta interpretación de los resultados es que el train y el test set provengan dentro de lo posible de la misma distribución. En este caso, el conjunto de entrenamiento posee un 26.21 % de píxeles NM, un 19.04 % de píxeles M y un 54.76 % píxeles N/D, mientras que el conjunto de evaluación posee un 20.38 % de píxeles NM, un 30.03 % de píxeles M y un 49.59 % píxeles N/D. Podemos ver que la categoría M está siendo entrenada con un número menor de píxeles que el resto, y a su vez con un menor porcentaje de píxeles respecto de la misma categoría en el test set. Cuando trabajemos con el dataset completo, trataremos de revertir esta situación. Es importante tener en cuenta que si el test set no proviene de la misma distribución que el train set, podríamos subestimar o sobreestimar el desempeño de la red neuronal.

Al analizar datos del test set de manera individual, notamos que muchos errores se producen en las áreas donde diferentes categorías de clasificación se encuentran en contacto. Es probable que este fenómeno también se presente de haber realizado la clasificación un experto en lugar de un algoritmo computacional, por lo que no podemos esperar que este tipo de error desaparezca por completo, sino que más bien se reduzca.

Como mencionamos anteriormente, el clasificador identificó un 2.09 % de los píxeles que originalmente habían sido clasificados como N/D por el algoritmo de NEXRAD como pertenecientes a las clases M o NM. Esta clasificación puede o no ser correcta. Es decir, es posible que el modelo logre identificar píxeles que el algoritmo de clasificación utilizado para clasificar los datos del sistema NEXRAD no haya podido clasificar. Por otro lado, los píxeles erróneamente clasificados como M o NM (originalmente clasificados como NM y M respectivamente), sumados a los píxeles clasificados como N/D que originalmente eran considerados M o NM, representan en conjunto el 4.19 % de los datos. Sobre estos dos últimos tipos de errores debemos prestar especial atención en trabajos futuros.

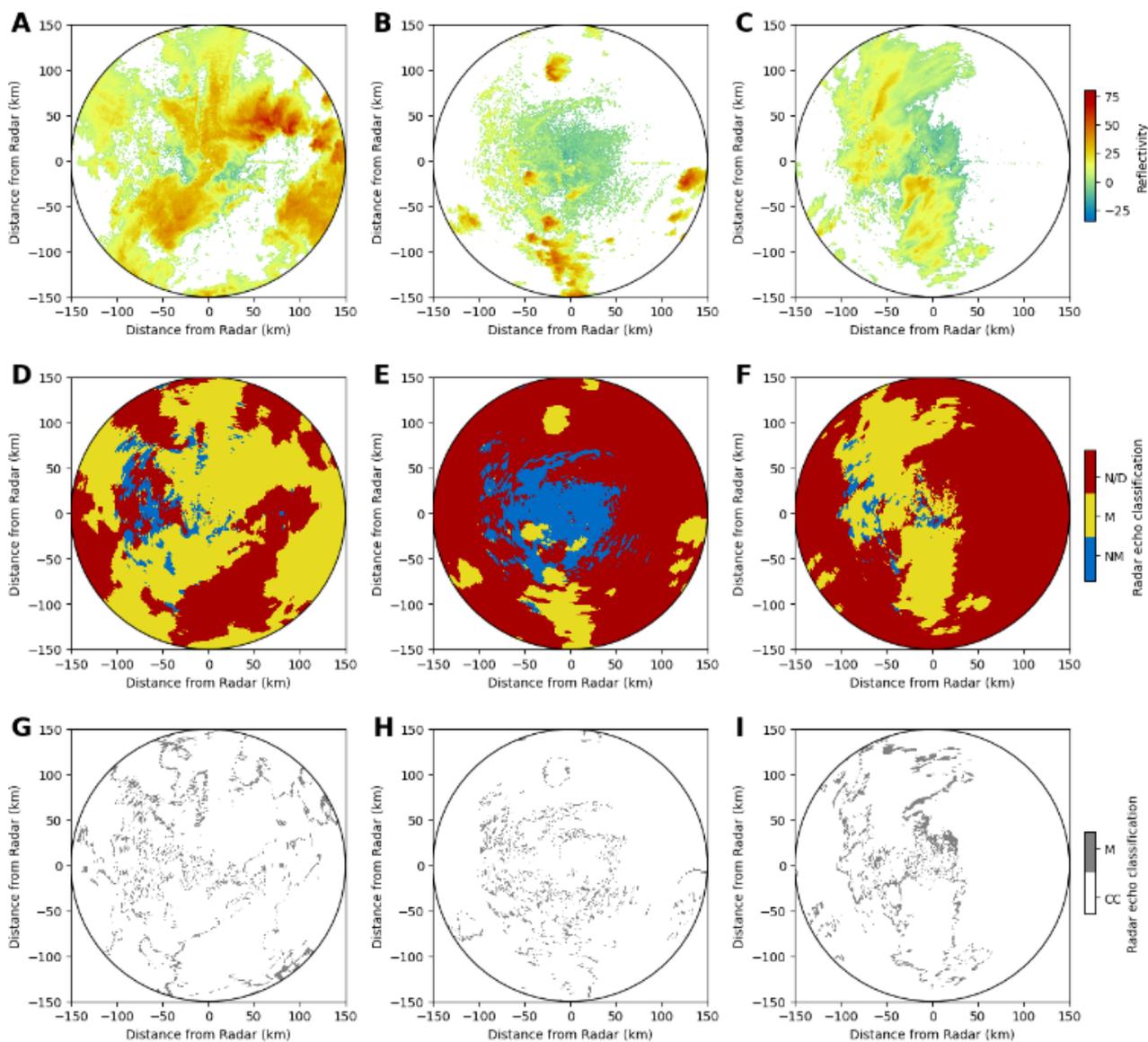


Figura 6: Diferencias en cuanto a la clasificación de hidrometeoros entre los datos suavizados y los datos predichos por el modelo para 3 eventos distintos (uno por columna) correspondientes al radar KFTG (test set). Se muestra la reflectividad (A, B y C), la clasificación de los datos luego del suavizado (D, E y F), y la diferencia respecto de la clasificación de los datos luego del suavizado y la clasificación predicha por el modelo (G, H y I). N/D = No Data, M = ecos meteorológicos, NM = ecos no meteorológicos. En blanco se indica que la clasificación es la misma según los datos y el modelo (CC = Correct Classification). En negro, que el modelo clasificó un pixel erroneamente (M = Misclassification).

REFERENCIAS

- Abadi, M., y coautores, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. URL <https://www.tensorflow.org/>, software available from tensorflow.org.
- Albu, A.-I., G. Czibula, A. Mihai, I. G. Czibula, S. Burcea, y A. Mezghani, 2022: Nextnow: A convolutional deep learning model for the prediction of weather radar data for nowcasting purposes. *Remote Sensing*, **14 (16)**, doi: 10.3390/rs14163890.
- Ayzel, G., T. Scheffer, y M. Heistermann, 2020: Rainnet v1.0: a convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, **13 (6)**, 2631–2644, doi:10.5194/gmd-13-2631-2020.
- Berenguer, M., D. Sempere-Torres, C. Corral, y R. Sánchez-Diezma, 2006: A fuzzy logic technique for identifying nonprecipitating echoes in radar scans. *Journal of Atmospheric and Oceanic Technology*, **23 (9)**, 1157–1180, doi:10.1175/JTECH1914.1.
- Fabry, F., 2018: *Radar Meteorology*. Cambridge University Press.
- Geiss, A., y J. C. Hardin, 2020: Radar super resolution using a deep convolutional neural network. *Journal of Atmospheric and Oceanic Technology*, **37 (12)**, 2197–2207, doi:10.1175/JTECH-D-20-0074.1.
- Ghanbari, H., M. Mahdianpari, S. Homayouni, y F. Mohammadimanesh, 2021: A meta-analysis of convolutional neural networks for remote sensing applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 3602–3613, doi:10.1109/JSTARS.2021.3065569.
- Giménez, D., L. Vidal, M. Rugna, R. de Elía, y L. Giordano, 2021: Climatología de ecos no meteorológicos para la red de radares SINARAME. *Nota Técnica SMN 2021-103*, URL <http://hdl.handle.net/20.500.12160/1685>.
- Géron, A., 2019: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.
- Han, L., J. Sun, y W. Zhang, 2020: Convolutional neural network for convective storm nowcasting using 3-d doppler weather radar data. *IEEE Transactions on Geoscience and Remote Sensing*, **58 (2)**, 1487–1495, doi:10.1109/TGRS.2019.2948070.
- Karasiak, N., J.-F. Dejoux, C. Monteil, y D. Sheeren, 2021: Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, doi:10.1007/s10994-021-05972-1, URL <https://hal.inrae.fr/hal-03238347>.
- Lu, Y., y J. Kumar, 2019: Convolutional neural networks for hydrometeor classification using dual polarization doppler radars.
- Pavlík, P., V. Rozinajová, y A. B. Ezzeddine, 2022: Radar-based volumetric precipitation nowcasting: A 3d convolutional neural network with u-net architecture. *EUR Workshop Proceedings*, **(3207)**, 65–72.
- Pedregosa, F., y coautores, 2011: Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, **12 (85)**, 2825–2830, URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- Raschka, S., 2020: Model evaluation, model selection, and algorithm selection in machine learning. [1811.12808](https://arxiv.org/abs/1811.12808).

- Rico-Ramirez, M. A., y I. D. Cluckie, 2008: Classification of ground clutter and anomalous propagation using dual-polarization weather radar. *IEEE Transactions on Geoscience and Remote Sensing*, **46 (7)**, 1892–1904, doi:10.1109/TGRS.2008.916979.
- Ronneberger, O., P. Fischer, y T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, y A. F. Frangi, Eds., Springer International Publishing, Cham, 234–241.
- Suárez, S. R., M. Sued, L. Vidal, P. Salio, D. Rodriguez, S. Nesbitt, y Y. G. Skabar, 2019: Técnicas de clasificación supervisada para la discriminación entre ecos meteorológicos y no meteorológicos usando informacion de un radar de banda c. *Meteorológica*, **44**, 45–65.
- Zhang, C.-J., H.-Y. Wang, J. Zeng, L.-M. Ma, y L. Guan, 2020: Tiny-rainnet: a deep convolutional neural network with bi-directional long short-term memory model for short-term rainfall prediction. *Meteorological Applications*, **27 (5)**, e1956, doi:<https://doi.org/10.1002/met.1956>.

Instrucciones para publicar Notas Técnicas

En el SMN existieron y existen una importante cantidad de publicaciones periódicas dedicadas a informar a usuarios distintos aspectos de las actividades del servicio, en general asociados con observaciones o pronósticos meteorológicos.

Existe no obstante abundante material escrito de carácter técnico que no tiene un vehículo de comunicación adecuado ya que no se acomoda a las publicaciones arriba mencionadas ni es apropiado para revistas científicas. Este material, sin embargo, es fundamental para plasmar las actividades y desarrollos de la institución y que esta dé cuenta de su producción técnica. Es importante que las actividades de la institución puedan ser comprendidas con solo acercarse a sus diferentes publicaciones y la longitud de los documentos no debe ser un limitante.

Los interesados en transformar sus trabajos en Notas Técnicas pueden comunicarse con Ramón de Elía (rde-elia@smn.gob.ar), Luciano Vidal (lvidal@smn.gob.ar) o Martín Rugna (mrugna@smn.gob.ar) de la Dirección Nacional de Ciencia e Innovación en Productos y Servicios, para obtener la plantilla WORD o los archivos Tex que sirven de modelo para la escritura de la Nota Técnica. Una vez armado el documento deben enviarlo en formato PDF a los correos antes mencionados. Antes del envío final los autores deben informarse del número de serie que le corresponde a su trabajo e incluirlo en la portada.

La versión digital de la Nota Técnica quedará publicada en el Repositorio Digital del Servicio Meteorológico Nacional. Cualquier consulta o duda al respecto, comunicarse con Melisa Acevedo (macevedo@smn.gob.ar).