



Nowcasting de precipitación basado en técnicas de aprendizaje profundo de máquinas e imágenes de radar: Caso de estudio, Valle del Aburrá

Julio Simón Pedro Bohórquez Hurtado

Modalidad de Práctica

Trabajo de Grado

Esteban Silva Villa

PhD. Astrofísica

John Alejandro Martínez

PhD. Hidrometeorología

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Ambiental

Medellín

2015

Cita

(Bohórquez Hurtado, 2025)

Referencia

Bohórquez Hurtado, (2025). *Nowcasting de precipitación basado en técnicas de aprendizaje profundo de máquinas e imágenes de radar: Caso de estudio, Valle del Aburrá*. Universidad de Antioquia, Medellín.

Estilo APA 7 (2020)



Grupo de investigación FCom (Física y Astrofísica Computacional)



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenidos

Resumen	6
Abstract	6
1 Introducción	7
2 Objetivos	11
2.1 Objetivo General	11
2.2 Objetivos específicos	11
3 Marco Teórico	12
3.1 Precipitación y Radares meteorológicos	12
3.2 Nowcasting y modelos de pronóstico convencionales	14
3.3 Aprendizaje profundo de máquinas y Nowcasting	16
4 Metodología	19
4.1 Recolección y selección de Datos	19
4.2 Preprocesamiento	21
4.2.1 Concatenación de imágenes de radar y limpieza preliminar de datos	21
4.2.2 Reducción dimensional espacial y remuestreo	22
4.2.3 Cálculo de secuencias y adecuación de formato	24
4.3 Entrenamiento y predicciones	26
4.3.1 Arquitectura del modelo	27
4.3.2 Entrenamiento por cada lote	30
4.4 Evaluación	30
4.4.1 Evaluación categórica del pronóstico de precipitación	31
4.4.2 Evaluación cuantitativa continua del pronóstico de precipitación	32
5 Análisis de Resultados	33
5.1 Evolución de pronóstico en el tiempo	34
5.2 Primer paso temporal por secuencia: métricas y casos destacados	36
5.3 Evolución de la función de pérdida en el entrenamiento	39
5.4 Otros Experimentos	40
6 Conclusiones	44
Referencias	47

Lista de tablas

Tabla 1	Matriz de contingencia 2×2 con el fin de ilustrar las evaluaciones categóricas del pronóstico de precipitación	31
Tabla 2	Comparación de métricas entre <i>ConvLSTM</i> y <i>Persistencia</i> con respecto a las observaciones en los pasos de tiempo pronosticados. Los mejores resultados para cada métrica por cada paso de tiempo están resaltados en negritas, para el modelo que obtuvo mejores resultados).	34
Tabla 3	Tabla de Comparación del desempeño de predicción de 6 minutos ($t + 1$) para todas las secuencias en el conjunto de prueba, donde el modelo que mejor desempeño tuvo se encuentra en negrilla , y para cada métrica se encuentra el valor en verde para el mejor valor por métrica de ConvLSTM y en rojo para el peor	38

Lista de figuras

Figura 1	Radar Operacional Meteorológico del Sistema de Alertas Tempranas del Valle del Aburrá (SIATA), ubicado en Santa Elena (6.19, -75.52) (SIATA, 2025)	7
Figura 2	Mapa de Reflectividad Sobre Horizontal a 1° de elevación, donde se puede observar la ubicación del radar, los municipios del Área Metropolitana del Valle del Aburrá (AMVA); y la cobertura máxima del radar con este grado de inclinación (120km)	9
Figura 3	(a). Características de la precipitación estratiforme donde se ve cómo del tope de la nube se encuentran de partículas de hielo que se disponen para su agregación y condensación para formar precipitación, se puede observar que es una nube estratiforme enlongada que presenta precipitación. (b). Características de precipitación convectiva, en donde se nota el desarrollo vertical de la nube a lo largo del tiempo (<i>t</i>). Figura y explicación tomada de (Houze Jr, 2014), capítulo 6.	12
Figura 4	Medidas de un objetivo desde un sensor flotante, donde se observa el <i>rango</i> (slant range), el <i>azimut</i> (δ) y la <i>elevación</i> (α) (Shroder, 2013).	13
Figura 5	Tipos de escaneo y presentación de información del radar meteorológico (Rauber & Nesbitt, 2018) .	14
Figura 6	Diagrama de la capacidad de predicción medida por el puntaje simétrico de dependencia extrema (<i>SEDS</i> , por sus siglas en inglés), el cual muestra 4 tipos de modelo para realizar Nowcasting, donde la extrapolación y el combinado son los únicos que tienen puntaje antes de la primera hora (negro y rojo), y los (<i>NWP</i>) en verde y azul, mostrando mejor rendimiento el corregido (corrección de intensidad y posición).(Sun et al., 2014)	16
Figura 7	Morfología general de una red neuronal, en la que las flechas que conectan cada neurona (representadas por círculos) corresponden a los pesos que se ajustan durante el proceso de entrenamiento. Las funciones de activación y las sumas ponderadas conforman las capas que reciben las salidas de la capa anterior, y finalmente se obtiene una salida de la red neuronal. Esta salida se evalúa mediante una función de pérdida que calcula el error y lo propaga hacia atrás para ajustar los parámetros. Tomada de (Paaß, s.f.).	17
Figura 8	Diagrama de las fases del proceso de pronóstico de precipitación a corto plazo	19
Figura 9	Diagrama detallado del proceso de recolección de datos	20
Figura 10	Proceso de recorte aplicado a una imagen de radar. A la izquierda se observa la imagen original del radar, con una resolución de 1728×1728 píxeles, correspondiente a un rango aproximado de 120 km. A la derecha se muestra la región recortada, centrada en la ubicación del radar, con un área de 0.5° de ancho tanto en latitud como en longitud,	21
Figura 11	Porcentaje de cantidad de imágenes de resolución temporal de 5 minutos para cada Mes	22
Figura 12	Histograma de la distribución de resoluciones para el conjunto de datos en el año 2020	23
Figura 13	Diagrama detallado del proceso de preprocesamiento	24
Figura 14	Métodos y eventos de análisis de reducción dimensional, donde cada fila representa la misma imagen de radar y las columnas son las reducciones dimensionales	25
Figura 15	Métodos y eventos de análisis de reducción dimensional, donde cada fila representa la misma imagen de radar y las columnas son las reducciones dimensionales	26
Figura 16	Diagrama detallado del proceso de entrenamiento del modelo, donde se puede diferenciar la forma de entrenar por lote, las secuencias utilizadas y la forma en que se mide el error e inicia el ciclo de propagación y cambio de gradientes.	27
Figura 17	Imagen de una celda de ConvLSTM donde se pueden ver las compuertas de entrada, de olvido, de celda, de salida y el estado oculto i_t, f_t, C_t, o_t, H_t . En esta imagen se puede ver el flujo de los datos dentro de una neurona. (C. Shi et al., 2022)	28
Figura 18	Diagrama de la evolución temporal de las métricas de rendimiento en todos los pasos del tiempo de pronóstico. Se puede observar en ésta el decaimiento del desempeño de ambos modelos con los pasos de tiempo	33
Figura 19	Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la mejor predicción en términos de <i>CSI</i> y <i>POD</i> , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3	35
Figura 20	Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la peor predicción en términos de <i>POD</i> , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3	36

Figura 21	Mapas de reflectividad para los pasos de tiempo ($t+5$) de la peor predicción en términos de <i>RMSE</i> y la mejor en términos de <i>FAR</i> , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3	37
Figura 22	Mapas de reflectividad para los pasos de tiempo ($t+5$) de la mejor predicción en términos de <i>RMSE</i> y peor en términos de <i>FAR</i> , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3	39
Figura 23	Mapas de reflectividad para los pasos de tiempo ($t+5$) de la mediana predicción en términos de <i>CSI</i> , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3	40
Figura 24	Evolución de la función de pérdida a lo largo del entrenamiento y del conjunto de validación a lo largo de las épocas	41
Figura 25	Experimento de entrenamiento con 50 épocas, utilizando <i>CPU</i> , en el conjunto de datos en el que los resultados se basaron (01-11-2020 - 17-11-2020) sin utilizar conjunto de validación.	42
Figura 26	Experimento de entrenamiento con 25 épocas, utilizando <i>CPU</i> para JJA.	42

Resumen

El pronóstico a muy corto plazo (*nowcasting*) cumple un rol crucial en la predicción meteorológica, donde el radar meteorológico constituye un instrumento clave para la observación de fenómenos atmosféricos. Sin embargo, los modelos tradicionales presentan desafíos importantes: los enfoques basados en extrapolación tienen dificultades para representar procesos no lineales, mientras que los modelos numéricos requieren un uso intensivo de recursos computacionales. En respuesta a estas limitaciones, este trabajo implementa técnicas de aprendizaje profundo de máquinas en el contexto del nowcasting, aprovechando su capacidad para modelar relaciones no lineales y su mayor eficiencia computacional. En particular, se entrenó un modelo basado en redes neuronales ConvLSTM, cuya estructura permite capturar dinámicas espacio-temporales, con el objetivo de predecir 30 minutos en el futuro (a intervalos de 6 minutos) a partir de una secuencia de 29 imágenes de radar consecutivas. Los resultados muestran que el modelo alcanza un buen desempeño en el primer paso de tiempo en comparación con el modelo de persistencia, lo cual se explica por su entrenamiento enfocado en la predicción de la imagen siguiente; no obstante, el rendimiento decrece conforme avanza el horizonte temporal, lo que sugiere la necesidad de ajustar la estrategia de entrenamiento para incluir múltiples pasos de predicción en el proceso de optimización.

Palabras Clave: nowcasting, precipitación, radar meteorológico, aprendizaje profundo de máquinas, ConvLSTM, pronóstico a muy corto plazo

Abstract

Nowcasting, or very short-term weather forecasting, plays a crucial role in meteorology, with weather radar serving as a key observation instrument. However, traditional models face significant challenges: extrapolation-based approaches struggle to capture nonlinear atmospheric processes, while numerical models demand high computational resources. To address these limitations, this study implements deep learning techniques for nowcasting task, leveraging their ability to model nonlinear relationships and their greater computational efficiency. Specifically, a ConvLSTM-based neural network model was trained to capture spatiotemporal dynamics, aiming to forecast 30 minutes into the future (at 6 minute intervals) using a sequence of 29 consecutive radar images as input. The results show that the model performs well for the first prediction step when compared to the persistence model, which is explained under its training strategy focused on predicting the next image in the sequence; however, performance deteriorates as the forecast horizon increases, suggesting the need for an improved training approach that includes multiple future steps in the optimization process.

Keywords: nowcasting, precipitation, weather radar, deep learning, ConvLSTM, very short-term weather forecast

1 Introducción

Es importante conocer el estado de la atmósfera y sobre todo, el estado de la precipitación. Esto se debe a que la mayoría de las actividades humanas dependen directa o indirectamente del recurso hídrico para su desarrollo (Sepulveda Berrío, 2016). Las nubes suceden principalmente en las escalas convectivas (0.2km - 20km) y mesoescalar (20km - 200km) (Houze Jr, 2014); estas escalas presentan dificultad para ser observadas y medidas, debido a que tienden a ser muy pequeñas para ser visualizadas con la suficiente resolución por plataformas satelitales, y muy grandes para ser detectadas de manera local como las estaciones meteorológicas (Houze Jr, 2014). Por esto, los sensores remotos como los radares meteorológicos son de suma importancia en la observación de estos fenómenos, puesto que permiten captar información sobre los hidrometeoros y su movimiento (Houze Jr, 2014; X. Shi et al., 2017).

Gracias a la información que los radares pueden obtener sobre el estado actual de la atmósfera y su evolución en el espacio y en el tiempo, hace que sean considerados instrumentos con una gran aplicabilidad en las ciencias de la atmósfera, desde la caracterización cualitativa y cuantitativa de precipitación hasta la detección de amenazas hidrometeorológicas y el pronóstico del tiempo. (Rauber & Nesbitt, 2018; Sepulveda Berrío, 2016).

En la actualidad, Colombia cuenta con 10 radares distribuidos sobre el territorio nacional (IDEAM, 2025). Nuestro objeto de estudio será el radar meteorológico ubicado en el Valle del Aburrá y operado por el Sistema de Alertas Temprana (SIATA), debido a que es un radar Doppler, de doble polarización, que nos permite obtener la reflectividad horizontal, ver figura 1 (Sepulveda Berrío, 2016).



Figura 1: Radar Operacional Meteorológico del Sistema de Alertas Tempranas del Valle del Aburrá (SIATA), ubicado en Santa Elena (6.19, -75.52) (SIATA, 2025)

El SIATA, y en general los sistemas de alerta temprana, se encargan de alertar en tiempo real a los organismos de gestión de riesgos y a la comunidad en respuesta a la posible ocurrencia de un fenómeno natural que pueda generar un estado de emergencia (SIATA, 2025). Por lo tanto, el pronóstico del tiempo atmosférico, y en especial la predicción a mediano y corto plazo, es esencial para la detección temprana de eventos extremos que puedan ocasionar desastres naturales y con ellos posiblemente pérdidas humanas y materiales. (Prudden et al., 2020).

La descripción detallada del tiempo atmosférico en el presente con un detalle local y su pronóstico de máximo 6 horas en el futuro es llamado *nowcasting* (X. Shi et al., 2017). El nowcasting es normalmente aplicado al tiempo atmosférico que ocurre en la mesoescala o en la escala local y depende de sistemas de observaciones que brinden información constante y en tiempo real (Wang et al., 2017). Estos sistemas pueden ser alimentados por observaciones de satélites, de radiosondas, estaciones en tierra, perfiladores de vientos y radares, siendo éstos últimos los más importantes para el nowcasting y particularmente para los fenómenos convectivos (Wang et al., 2017).

En la actualidad, el nowcasting basado en imágenes de radar puede ser categorizado en dos clases (Prudden et al., 2020; X. Shi et al., 2015; Xiong et al., 2024). En primera instancia se tienen los métodos basados en NWP (*Numerical Weather Prediction*, por sus siglas en inglés). Esta primera clase de modelos resuelve ecuaciones diferenciales parciales sujetas a las leyes de la termodinámica y de la dinámica que rigen la física de la atmósfera (e.g., L. Chen et al., 2020). En segundo lugar, se tienen los métodos basados en la extrapolación de los ecos de radar, los cuales adoptan técnicas de visión por computadora, como flujo óptico, para predecir el mapa con los nuevos ecos de radar como se presenta en la figura 2 (se invita al lector a ver también X. Shi et al., 2017).

Un factor que no es resuelto por los métodos de extrapolación es la predicción de la evolución lagrangiana en los sistemas convectivos, en la cual se consideran los cambios de velocidad y posición de las partículas, estando esta relacionada con la iniciación convectiva de dichos sistemas. Esta limitación representa la principal fuente de error en el nowcasting del desarrollo de estructuras de precipitación convectiva (Prudden et al., 2020). Además, la capacidad predictiva de estos modelos asociada al tamaño y organización de la precipitación, tiende a degradarse rápidamente a medida que se incrementa el horizonte temporal (Sun et al., 2014), debido a su inhabilidad de modelar las interacciones asociadas al estado de la atmósfera (Prudden et al., 2020). Los modelos NWP mitigan, en cierta medida, estas limitaciones mediante el modelado probabilístico de la incertidumbre, así como la incorporación de diagnósticos adicionales como la energía convectiva potencial (CAPE, por sus siglas en inglés) y la inhibición convectiva (Atencia et al., 2010; Prudden et al., 2020; Sun et al., 2014).

Ambos tipos de modelos presentan limitaciones en el contexto del nowcasting: los modelos de extrapolación carecen de la capacidad para capturar relaciones no lineales en los datos (Agrawal et al., 2019), mientras que los modelos NWP presentan restricciones para generar inferencias útiles en horizontes de muy corto plazo (X. Shi et al., 2017). Como respuesta a estas limitaciones, se han desarrollado sistemas que combinan las ventajas de ambas técnicas (extrapolación de imágenes de radar y modelos NWP). Un ejemplo de esta estrategia es el sistema implementado por el servicio meteorológico de Cataluña (España) (Atencia et al., 2010), en el cual se asignan pesos a las predicciones generadas por cada tipo de modelo, de forma que los métodos de advección tienen mayor peso en las primeras horas (0–2 horas), mientras que los modelos numéricos cobran mayor relevancia a partir de la segunda hora (Atencia et al., 2010).

Si bien realizar una combinación entre ambas técnicas tiene como objetivo minimizar el error en las primeras dos horas, presenta el desafío de representar la no linealidad de los procesos en esa franja de tiempo debido a que el modelo a utilizar es basado en extrapolación (Atencia et al., 2010). Así, las estrategias basadas en datos, que emplean modelos de aprendizaje de máquinas y modelos de aprendizaje profundo de máquinas, pueden ofrecer oportunidades para captar relaciones espacio-temporales que representen esta no linealidad (Düben et al., 2021; Prudden et al., 2020; Sønderby et al., 2020)

La inteligencia artificial y el aprendizaje automático han experimentado avances significativos en los últimos años, proporcionando capacidades avanzadas para el estudio del sistema terrestre. Estas tecnologías ofrecen un gran potencial en áreas como la observación, el análisis, la predicción y la comprensión de fenómenos naturales (National Academies of Sciences & Medicine, 2022). En el ámbito del nowcasting, las redes neuronales ConvLSTM se han destacado como tecnologías pioneras (L. Chen et al., 2020; X. Shi et al., 2015). Estas redes combinan arquitecturas recurrentes que capturan relaciones temporales, con arquitecturas convolucionales que identifican patrones espaciales. Esta integración permite abordar la complejidad y no linealidad de los procesos atmosféricos (X. Shi et al., 2015).

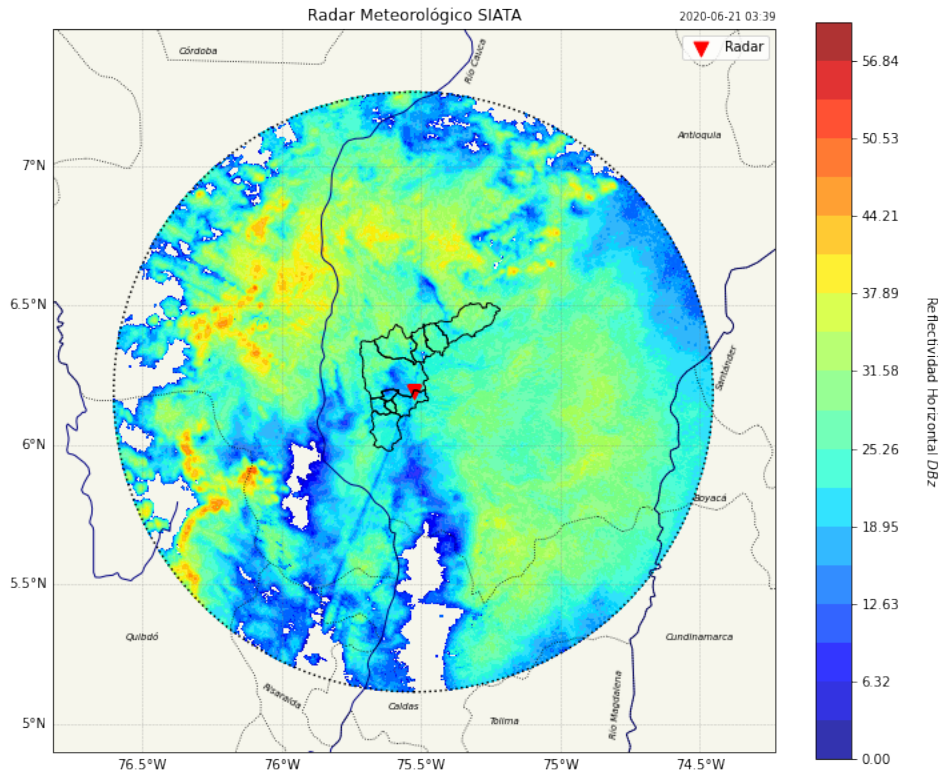


Figura 2: Mapa de Reflectividad Sobre Horizontal a 1° de elevación, donde se puede observar la ubicación del radar, los municipios del Área Metropolitana del Valle del Aburrá (AMVA); y la cobertura máxima del radar con este grado de inclinación (120km)

Por otro lado, se ha investigado el uso de redes neuronales generativas, como las redes neuronales adversariales (*GAN*, por sus siglas en inglés), que entrenan dos modelos en competencia entre sí (Ravuri et al., 2021). Este enfoque involucra un sistema generador, que crea imágenes a partir de ruido en los datos, y un sistema discriminador, que en algunos casos está informado por principios físicos (Yin et al., 2024). El discriminador evalúa si una imagen fue producida por el generador o si pertenece a los datos reales (Tian et al., 2020). Esta dinámica competitiva entre ambos sistemas contribuye a mejorar la calidad y el realismo de las imágenes generadas.

En el contexto colombiano, no se ha encontrado evidencia de estudios previos que utilicen imágenes de radar para el nowcasting. Sin embargo, se ha investigado el uso de radares para la identificación de sistemas convectivos y su distribución espacial y temporal en relación con ciertas cuencas hidrográficas de Antioquia (Velásquez, 2022). Además, se han explorado aplicaciones para la estimación cuantitativa de precipitación (*QPE*, por sus siglas en inglés), enfocándose en la adaptación y optimización de metodologías clásicas (Sepulveda Berrío, 2016), así como en la caracterización de la microfísica de la lluvia (Ladino Rincón, 2018).

Por otro lado, en cuanto al uso de aprendizaje de máquinas para el pronóstico de precipitación, se ha demostrado

la posibilidad de hacer pronósticos de anomalías de precipitación de 5 meses en adelante en el suroeste colombiano, mediante el uso de redes neuronales artificiales y estaciones pluviométricas (Canchala et al., 2020). A su vez, en Boyacá, se empleó el uso de redes neuronales recurrentes con Memoria a Largo y Corto Plazo (LSTM, por sus siglas en inglés) para realizar predicciones de precipitación de hasta 16 meses mediante imágenes satelitales, mostrando superioridad al capturar variaciones estacionales (Niño Medina et al., 2024).

Teniendo en cuenta lo anterior, el presente estudio utiliza datos del radar meteorológico del SIATA, para el periodo entre 2018 y 2021, con el objetivo de determinar un periodo de tiempo óptimo para implementar un modelo de nowcasting de precipitación basado en datos. El propósito es contribuir a la exploración del nowcasting de precipitación en el país mediante un modelo que pueda predecir eficazmente al menos los primeros 30 minutos a partir de una secuencia de imágenes. Además, se busca que el modelo tenga una eficiencia operativa y un uso de recursos óptimo, con el fin de servir como alternativa a los modelos de extrapolación y flujo óptico.

Inicialmente, se recolectaron imágenes de radar del SIATA cada cinco minutos durante el periodo mencionado. Estas imágenes fueron recortadas para definir un área de interés específica sobre el Valle del Aburrá. Posteriormente, se seleccionó el periodo de tres meses, entre septiembre y noviembre de 2020, para asegurar la inclusión de eventos de precipitación en el conjunto de datos.

Con este conjunto de datos, se llevó a cabo una etapa de preprocesamiento en la que se redujo la dimensionalidad espacial de cada imagen mediante algoritmos de interpolación. Además, las imágenes fueron transformadas de forma que adquirieran el formato adecuado para entrenar el modelo de aprendizaje profundo. El modelo seleccionado se basó en la arquitectura del trabajo pionero en el uso de algoritmos de aprendizaje profundo, utilizando redes neuronales ConvLSTM. Estas redes han demostrado un desempeño competitivo al compararse con modelos tradicionales de flujo óptico (L. Chen et al., 2020; Demetrakopoulos, 2023; X. Shi et al., 2015), lo que las hace idóneas para un estudio exploratorio en el ámbito del nowcasting.

Finalmente, se entrenó un modelo basado en redes neuronales ConvLSTM con el objetivo de realizar predicciones de reflectividad con un tiempo máximo de 30 minutos. Para evaluar el desempeño del modelo se utilizó un conjunto de datos de prueba previamente seleccionado, y se aplicaron métricas como el error cuadrático medio (*RMSE*, por sus siglas en inglés). Adicionalmente, se analizaron las capacidades del modelo para detectar reflectividad mediante indicadores como la probabilidad de detección (*POD*, por sus siglas en inglés) y el índice crítico de éxito (*CSI*, por sus siglas en inglés).

2 Objetivos

2.1 Objetivo General

Implementar un modelo basado en datos para el pronóstico de precipitación a muy corto plazo (0 a 30 minutos), utilizando imágenes de radar y redes neuronales ConvLSTM.

2.2 Objetivos específicos

- Recolectar y preprocesar imágenes de radar con el propósito de crear un conjunto de datos adecuado para el modelo.
- Evaluar estrategias de preprocesamiento con el fin de escoger la más adecuada para la preparación del conjunto de datos
- Configurar los hiperparámetros del modelo de aprendizaje profundo de máquinas y entrenar el modelo basado en redes neuronales ConvLSTM.
- Evaluar el desempeño del modelo con medidas de rendimiento cuantitativas comparando predicciones del modelo con un conjunto de datos de prueba.
- Analizar los resultados obtenidos, para comprender las debilidades y fortalezas del modelo para el pronóstico de precipitación a corto plazo.

3 Marco Teórico

3.1 Precipitación y Radares meteorológicos

La precipitación se clasifica comúnmente en dos tipos principales: estratiforme y convectiva (Ladino Rincón, 2018). La precipitación estratiforme se asocia típicamente con nubes del tipo nimboestratos, mientras que la convectiva se origina en nubes de desarrollo vertical como los cúmulos y cumulonimbos como muestra la figura 3 (Houze Jr, 2014). La precipitación estratiforme se caracteriza por presentar movimientos ascendentes de aire dentro de la nube, los cuales son menores a la velocidad de caída de los cristales de hielo. Ésta puede dividirse en dos subtipos, dependiendo de si las masas de aire en general se encuentran ascendiendo o descendiendo: precipitación estratiforme activa e inactiva, respectivamente (ver Houze Jr, 2014, para más detalles).

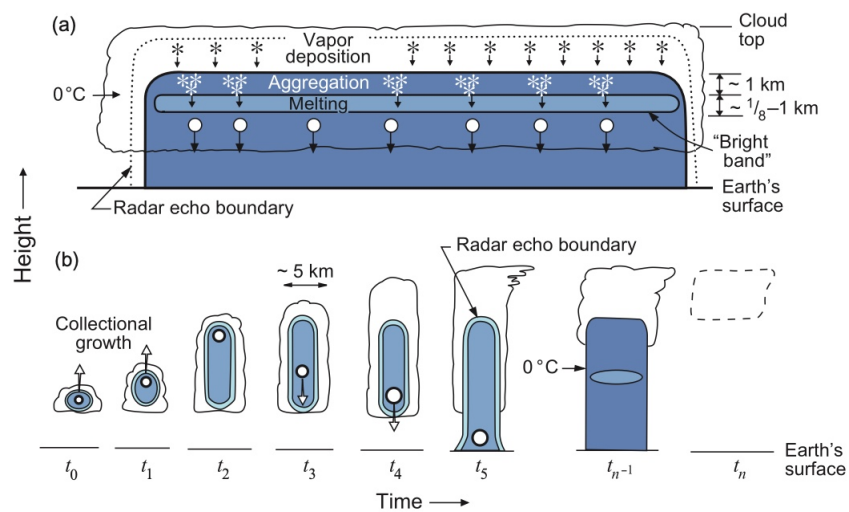


Figura 3: (a). Características de la precipitación estratiforme donde se ve cómo del tope de la nube se encuentran de partículas de hielo que se disponen para su agregación y condensación para formar precipitación, se puede observar que es una nube estratiforme alargada que presenta precipitación. (b). Características de precipitación convectiva, en donde se nota el desarrollo vertical de la nube a lo largo del tiempo (t). Figura y explicación tomada de (Houze Jr, 2014), capítulo 6.

Por otro lado, la precipitación convectiva se caracteriza por su rápida formación, impulsada por fuertes movimientos ascendentes de aire que favorecen procesos como la acreción de gotas, lo cual permite la condensación de grandes cantidades de agua (Houze Jr, 2014). Este tipo de precipitación puede generar columnas convectivas que alcanzan alturas de hasta 15 km (Sepulveda Berrío, 2016).

Los tipos de precipitación mencionados anteriormente suelen desarrollarse dentro de sistemas convectivos de mesoescala (Houze Jr, 2014), los cuales son estructuras atmosféricas complejas formadas por un conjunto de tormentas que abarcan áreas de precipitación continua de al menos 100 km en una de sus dimensiones (Houze Jr, 2014; Paul Markowski, 2010). Estos sistemas pueden evolucionar a partir de convección aislada o generarse de manera inmediata mediante procesos de iniciación convectiva (Paul Markowski, 2010).

Los radares son los instrumentos meteorológicos más importantes para observar los sistemas convectivos y de mesoescala (Wang et al., 2017). Estos mecanismos son sensores remotos activos que funcionan emitiendo pulsos de radiación electromagnética, generalmente en el rango de las microondas. Estas ondas atraviesan la atmósfera y, al

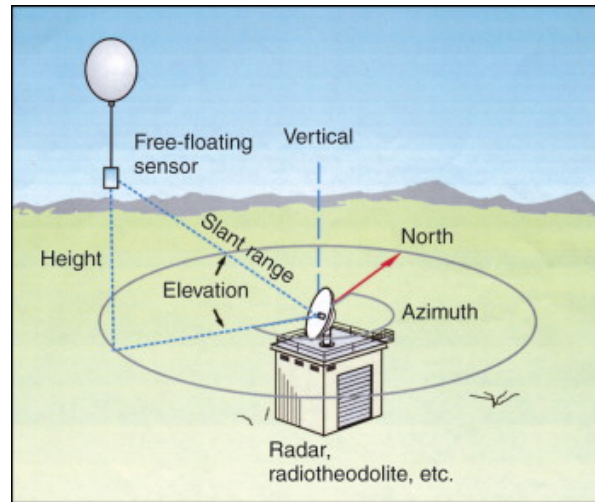


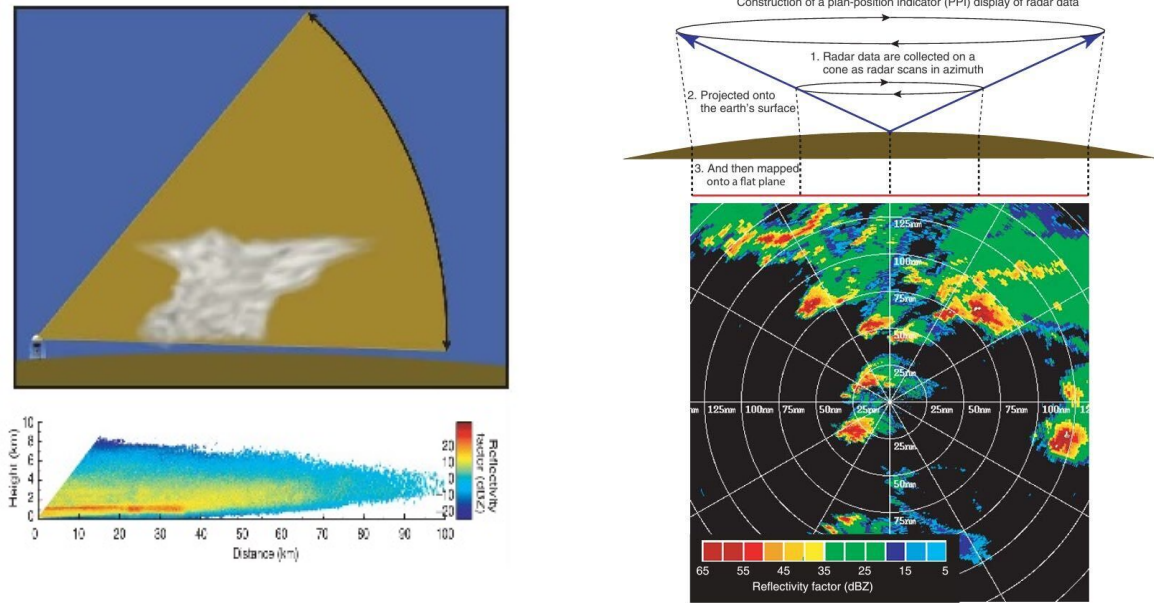
Figura 4: Medidas de un objetivo desde un sensor flotante, donde se observa el *rango* (slant range), el *azimut* (δ) y la *elevación* (α) (Shroder, 2013).

interactuar con partículas o distintos medios, se dispersan y parte de su energía se refleja de vuelta al radar, donde es interceptada e interpretada para inferir propiedades de la atmósfera (Ladino Rincón, 2018). A diferencia de estos sensores activos, los sensores remotos pasivos, como ciertos instrumentos satelitales, no emiten radiación propia, sino que dependen de las emisiones naturales de los objetos, como la radiación térmica o la energía solar reflejada (Rauber & Nesbitt, 2018).

Cuando la energía electromagnética es recibida nuevamente por el radar, éste es capaz de medir la potencia de la señal retornada, así como variables que permiten ubicar el objeto en el espacio, tales como el azimut, el rango y la elevación (Figura 4). La potencia reflejada se convierte en lo que se conoce como el factor de reflectividad, que constituye la variable más relevante en los radares meteorológicos, ya que permite estimar condiciones atmosféricas como la precipitación (Rauber & Nesbitt, 2018). En el caso de los radares Doppler, además de estas mediciones, se registra la frecuencia de la onda retornada; con base en el efecto Doppler, se puede determinar si los objetos se acercan o se alejan del radar, lo que facilita la interpretación del movimiento de las estructuras atmosféricas (Doviak & Zrníc, 1972).

Los datos obtenidos por los radares se recolectan en las coordenadas esféricas mencionadas anteriormente (*rango*, *elevación*, *azimuth*), y gracias a esto se pueden realizar distintas presentaciones de los datos, como lo son los PPI (*Plan Position Indicator*), y los RHI (*Range Height Indicator*) (Rauber & Nesbitt, 2018). Para los RHI, el escaneo del radar cuenta con un azimut δ fijo y un escaneo en el cual se cambia su ángulo de elevación α (Figura 5 a). Este tipo de escaneo es muy útil para entender las estructuras verticales de los sistemas atmosféricos (Rauber & Nesbitt, 2018). Por otro lado, el escaneo PPI, utilizado en este trabajo, presenta los datos de una manera tal que el ángulo de elevación se mantiene fijo y el escaneo se hace en los 360° grados de azimuth, proyectándose posteriormente en el plano (Rauber & Nesbitt, 2018) (Figura 5 b). Este tipo de escaneo es importante para obtener información completa a nivel espacial sobre las distintas estructuras que puedan existir en un área.

El factor de reflectividad Z , es la variable más relevante medida por el radar meteorológico (Rauber & Nesbitt, 2018). Se define como la suma de todos los diámetros de las gotas de agua elevado a la sexta potencia D_j^6 , dividido el volumen de la región medida (V_c) como lo indica la ecuación 1 (Doviak & Zrníc, 1972). Por lo tanto, sus medidas están dadas por mm^6/mm^3 . Sin embargo, se representa en unidades logarítmicas debido a que puede cambiar en varios órdenes de magnitud, tomando la notación de decibeles dB con respecto al nivel de referencia de $1mm^6/m^3$,



(a) *RHI*: Escaneo vertical cambiando el ángulo de elevación de la antena, útil para observar el desarrollo vertical de un sistema

(b) *PPI*: Escaneo horizontal manteniendo el ángulo de elevación fijo, y proyectando hacia el plano, útiles para observar la distribución espacial de los sistemas atmosféricos

Figura 5: Tipos de escaneo y presentación de información del radar meteorológico (Rauber & Nesbitt, 2018)

representando así el factor de reflectividad como *dBZ* (Rauber & Nesbitt, 2018). El valor de la reflectividad puede variar entre 0 *dBZ* y 65 *dBZ*, sin tener en consideración los tamaños de hidrometeoros menores a 1mm.

$$Z = \frac{\sum_j D_j^6}{V_c} \quad (1)$$

Ya que los radares no miden la precipitación de forma directa, su estimación cuantitativa se basa en diversos métodos convencionales que buscan relacionar la reflectividad registrada por el radar con la intensidad de precipitación medida en superficie, conocidas como relaciones *Z-R* (Rauber & Nesbitt, 2018; Sepulveda Berrío, 2016). La precipitación estratiforme puede identificarse en los *PPI*, dado que se manifiesta como extensas áreas horizontales con valores de reflectividad medios o bajos (Houze Jr, 2014; Sepulveda Berrío, 2016). A diferencia de los sistemas convectivos, los cuales se caracterizan por núcleos compactos de alta reflectividad y menor extensión espacial (Houze Jr, 2014; Sepulveda Berrío, 2016).

3.2 Nowcasting y modelos de pronóstico convencionales

El nowcasting corresponde a la predicción del estado atmosférico en muy el corto plazo con detalle local, abarcando desde el momento presente hasta un máximo de seis horas en el futuro (Yin et al., 2024). Por lo tanto, el nowcasting es aplicado generalmente al tiempo atmosférico que ocurre en la mesoescala (Wang et al., 2017), con el propósito de alimentar sistemas de alerta temprana que prestan apoyo a diversos sectores, tales como la gestión del riesgo,

la administración de recursos energéticos, el control aéreo y marítimo, así como los sistemas de respuesta ante emergencias por deslizamientos (Ravuri et al., 2021).

Gracias a esto, el radar meteorológico se consolida como el sistema de observación primario y de mayor confiabilidad para emitir advertencias sobre fenómenos convectivos extremos mediante el nowcasting (Wang et al., 2017). En ausencia de este instrumento, en la mayoría de los casos solo sería posible emitir un aviso o una vigilancia en términos operacionales, debido a la menor precisión y resolución temporal de otros métodos de observación (Wang et al., 2017). Luego de esto, los modelos tradicionales de pronóstico a muy corto plazo que se apoyan en imágenes de radar pueden clasificarse en dos grandes grupos: aquellos que hacen parte de los *NWP*, y aquellos basados en la extrapolación de ecos de radar (Prudden et al., 2020; Wang et al., 2017; Xiong et al., 2024).

Los modelos de extrapolación de ecos de radar se basan en la suposición de persistencia Lagrangiana, la cual asume que el estado de cada parcela de aire permanece constante en el tiempo y que todos los cambios observados se deben únicamente al proceso de advección, es decir, al desplazamiento del flujo atmosférico (Prudden et al., 2020). Muchos de estos algoritmos de flujo óptico comienzan por identificar las tormentas como objetos dentro del escaneo del radar, para posteriormente estimar el movimiento de dichos objetos mediante el análisis del desplazamiento entre imágenes sucesivas (Wang et al., 2017).

De esta manera, podemos decir que los algoritmos de extrapolación de ecos de radar se dividen en dos tipos: el primero estima el campo de advección y lo utiliza para propagar las observaciones hacia el futuro y realizar predicciones; mientras que el segundo identifica las tormentas como objetos individuales y se enfoca principalmente en las celdas convectivas, sin emplear el campo de advección completo (Prudden et al., 2020).

Si bien los modelos de extrapolación de ecos de radar son los más utilizados para periodos de pronóstico muy cortos (Mandapaka et al., 2012), su habilidad predictiva decae exponencialmente con el tiempo, además presentan limitaciones importantes para anticipar la iniciación convectiva, así como el decaimiento de las tormentas (Mandapaka et al., 2012; Sun et al., 2014).

Los modelos numéricos de tiempo atmosférico (*NWP*) son métodos de pronóstico que emplean ecuaciones que describen la termodinámica y las dinámicas de la atmósfera, utilizando las ecuaciones gobernantes de la dinámica de los fluidos, la parametrización de procesos físicos, y condiciones de frontera (Coiffier, 2011). Las técnicas implementadas tienen como objetivo resolver ecuaciones que describen el comportamiento de la atmósfera y, de esta manera, calcular los valores futuros de sus características a partir de condiciones iniciales conocidas mediante observaciones. Estos sistemas de ecuaciones no lineales, en general, no pueden resolverse de forma analítica (Coiffier, 2011).

Tales modelos tienen la capacidad de predecir la evolución de los sistemas atmosféricos, incluyendo la iniciación convectiva, el desarrollo y el decaimiento de las tormentas (Prudden et al., 2020). Sin embargo, presentan una menor habilidad para realizar pronósticos en periodos de tiempo muy cortos (en contraste con los métodos de extrapolación de radar) debido a los procesos de calentamiento (*spin-up*) y asimilación de observaciones (Mandapaka et al., 2012), siendo el periodo típico de calentamiento de un modelo de convección permitiva de 3h a 6h en "iniciación fría" (Sun et al., 2014).

Si bien esto se ha intentado solucionar mediante la asimilación de datos de radar de alta resolución y la implementación de ciclos de actualización rápida que permiten asimilar observaciones cada hora, con el fin de utilizar los *NWP* bajo una "iniciación caliente" (Sun et al., 2014), los modelos de extrapolación de radar continúan siendo más eficientes computacionalmente y ofrecen una buena capacidad de predicción en los primeros momentos del pronóstico (Atencia et al., 2010; Mandapaka et al., 2012).

Se evidencia un contraste entre ambas clases de modelos, ya que cada una tiene sus respectivas ventajas y desventajas. No obstante, en términos de habilidad predictiva, éstas están estrechamente ligadas a la evolución temporal del pronóstico: en los momentos iniciales, los modelos de extrapolación muestran un mejor desempeño, mientras que a medida que avanza el tiempo, los modelos de predicción numérica (*NWP*) tienden a superar su capacidad (Figura 6).

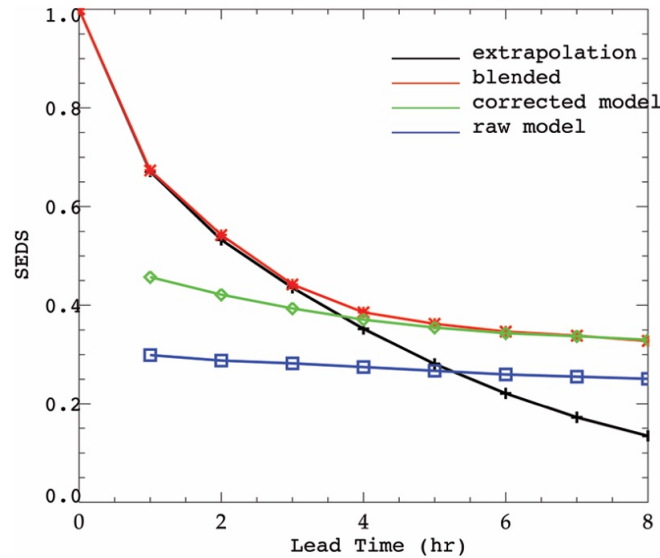


Figura 6: Diagrama de la capacidad de predicción medida por el puntaje simétrico de dependencia extrema (*SEDS*, por sus siglas en inglés), el cual muestra 4 tipos de modelo para realizar Nowcasting, donde la extrapolación y el combinado son los únicos que tienen puntaje antes de la primera hora (negro y rojo), y los (*NWP*) en verde y azul, mostrando mejor rendimiento el corregido (corrección de intensidad y posición). (Sun et al., 2014)

Por esta razón se han hecho combinaciones de ambos tipos de modelos, en los cuales se les da mayor peso a las predicciones realizadas por modelos de interpolación de radar en los primeros momentos (Sun et al., 2014), generando así una transición más suave entre los dos modelos, lo que permite poder aprovechar la capacidad de pronóstico de ambos (Figura 6).

Un ejemplo representativo de esta combinación se implementó en el Servicio Meteorológico de Cataluña, donde se mejoró la estimación cuantitativa de la precipitación en las primeras horas del pronóstico. En este enfoque, se calculó el campo de advección a partir de imágenes de radar y posteriormente se combinó, de manera espacial, con el campo de precipitación estimado por el modelo *NWP*, asignando pesos a cada uno (Atencia et al., 2010).

3.3 Aprendizaje profundo de máquinas y Nowcasting

La Inteligencia artificial es uno de los campos más nuevos de la ciencia y la ingeniería, la cual busca desarrollar sistemas que puedan imitar la inteligencia humana y sus procesos cognitivos y racionales, habilitando así capacidades de procesar lenguaje natural, razonamiento automático, visión por computadora, entre otros (Russell & Norvig, 2010).

El aprendizaje de máquinas (*Machine Learning*), es el campo más relevante de la inteligencia artificial para las ciencias de la tierra (Düben et al., 2021) que permite a los sistemas aprender patrones en los datos, y mejorar automáticamente sin tener que ser programados explícitamente. Por lo tanto es una aproximación basada en datos que utiliza estos patrones aprendidos para pronosticar, generalizar y realizar diferentes actividades (Russell & Norvig, 2010).

El sistema Tierra presenta una alta dificultad para ser simulado, ya que se considera un sistema complejo, caótico y no lineal. Por lo tanto, el aprendizaje de máquinas puede desempeñar un papel importante en las ciencias de la

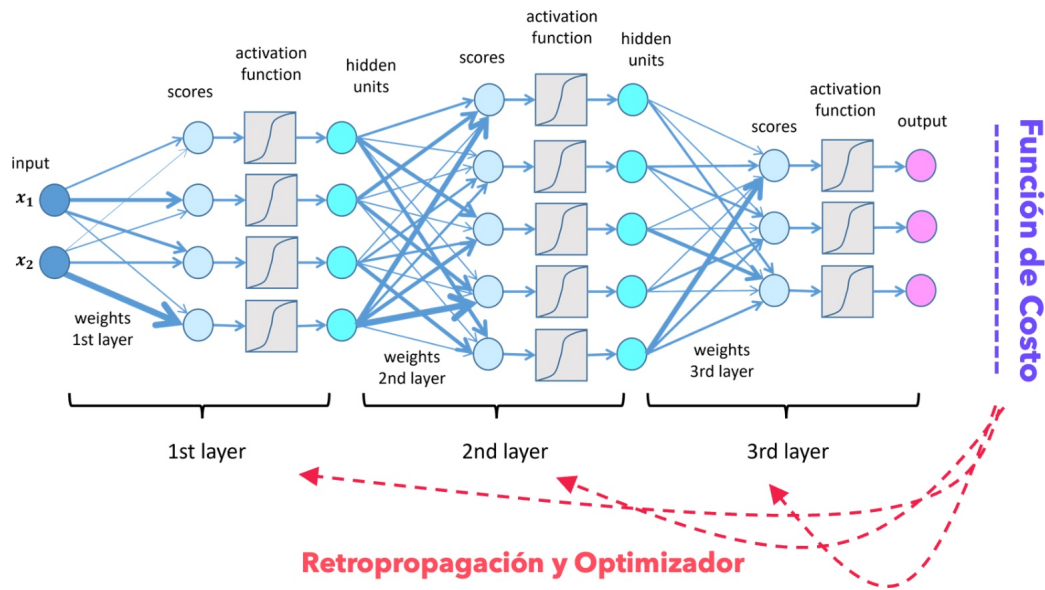


Figura 7: Morfología general de una red neuronal, en la que las flechas que conectan cada neurona (representadas por círculos) corresponden a los pesos que se ajustan durante el proceso de entrenamiento. Las funciones de activación y las sumas ponderadas conforman las capas que reciben las salidas de la capa anterior, y finalmente se obtiene una salida de la red neuronal. Esta salida se evalúa mediante una función de pérdida que calcula el error y lo propaga hacia atrás para ajustar los parámetros. Tomada de (Paaß, s.f.).

Tierra debido a su capacidad para extraer información a partir de los datos (National Academies of Sciences & Medicine, 2022). Además, puede contribuir a ampliar el entendimiento del sistema mediante la cuantificación de la incertidumbre, acelerar las simulaciones al emular componentes de los modelos, y mejorar dichos modelos corrigiendo sus sesgos (National Academies of Sciences & Medicine, 2022).

Incluso están empezando a emerger sistemas completos, de extremo a extremo, basados en datos, capaces de utilizar solo el 8% de los datos de entrada y tres órdenes de magnitud menos recursos computacionales que los *NWP*, logrando un rendimiento superior al de estos, como el sistema Aardvark (Vaughan et al., 2024). A su vez, el Centro Europeo para el Pronóstico a Mediano y Largo Plazo (*ECMWF*, por sus siglas en inglés) lanzó el sistema *AIFS*, el cual produce pronósticos que se desempeñan muy bien en variables de la alta atmósfera, como en el seguimiento de ciclones tropicales (Lang et al., 2024).

El proceso de entrenamiento de un algoritmo de aprendizaje de máquinas comprende el conjunto de prácticas mediante las cuales el sistema identifica patrones en los datos y “aprende” a partir de ellos (Géron, 2019). Este proceso puede dar lugar a distintos tipos de algoritmos, como los supervisados y no supervisados, cuya diferencia radica en la existencia o no de un valor real u observación conocido como etiqueta (Géron, 2019). Esta etiqueta permite al sistema ajustar el error en función del valor observado; por ejemplo, en un ejercicio de clasificación de tipos de nubes a partir de imágenes, etiquetar las imágenes de entrenamiento con valores cualitativos, como el tipo de nube, constituye un caso de aprendizaje supervisado.

El aprendizaje profundo de máquinas (*Deep Learning*) es el campo del aprendizaje de máquinas que estudia las redes neuronales profundas, inspiradas en las redes neuronales humanas, que representan capas de unidades de procesamiento que se comunican entre sí (Géron, 2019). A las neuronas llegan datos de entrada y se hacen sumas ponderadas con estos, incluyendo pesos y sesgos, los cuales son los parámetros que se ajustan con el proceso de

entrenamiento de la red neuronal (Pérez Velasco, 2020). Luego, mediante la función de activación (una función no lineal), estas sumas ponderadas pasan a la siguiente neurona. Las funciones de activación son las principales responsables de permitir que la red aprenda procesos no lineales (Géron, 2019) (*Figura 7*).

Para llevar a cabo el proceso de entrenamiento y ajustar los parámetros previamente mencionados, las redes neuronales utilizan funciones de costo, que miden el error entre la predicción realizada y las observaciones; algoritmos de retropropagación, que calculan la contribución al error de cada conexión en la red; y algoritmos de descenso de gradiente u optimizadores, que buscan minimizar el error de la función de costo (Géron, 2019) (*Figura 7*). La elección y configuración de estos componentes se realiza mediante hiperparámetros, los cuales se definen antes del entrenamiento y no son ajustados por el modelo, a diferencia de los pesos y sesgos (*llamados parámetros*). Este proceso ocurre para todo el conjunto de datos, y en la mayoría de ocasiones, controlado por un hiperparámetro llamado época (Géron, 2019).

El aprendizaje profundo de máquinas ha tenido éxito en muchos campos como el reconocimiento de imágenes, la segmentación de texto, e incluso en este momento en el procesamiento de lenguaje natural (Xiong et al., 2024). Debido a que son soluciones basadas en datos, la calidad y la cantidad de los datos es esencial, permitiéndonos así modelar el problema de nowcasting desde el aprendizaje profundo de máquinas, ya que satisface estos requerimientos de datos (X. Shi et al., 2015).

El uso de redes neuronales en el nowcasting ha sido variado en los últimos años, empleando distintas arquitecturas que buscan modelar el problema desde diferentes perspectivas. Dado que el nowcasting basado en datos de radar es un problema espacio-temporal, las redes neuronales convolucionales (*CNN*) juegan un papel crucial, ya que son ampliamente utilizadas en el reconocimiento de imágenes (X. Shi et al., 2015). Estas redes están compuestas por capas convolucionales que no están completamente conectadas entre sí, lo que permite extraer atributos de importancia mediante operaciones convolucionales y sub - muestreos (Géron, 2019).

La arquitectura U-net, basada en este tipo de redes, puede alcanzar un desempeño superior al de modelos de *NWP* como *HRRR* en los primeros instantes del pronóstico (Agrawal et al., 2019). Además, existen modelos de línea base, basados únicamente en capas convolucionales, que presentan una eficiencia comparable con los modelos de estado del arte de flujo óptico en los primeros instantes del pronóstico (Ayzel et al., 2019). Un acercamiento similar se empleó en el modelo de Redes Neuronales Convolucionales Lagrangianas (*L-CNN*), en el cual se separa la estimación del crecimiento y decaimiento de la precipitación del cálculo del campo de advección, logrando así que *L-CNN* representara mejor estos procesos en comparación con el modelo de extrapolación base *LINDA* (Ritvanen et al., 2023).

Por otro lado, las redes neuronales recurrentes (*RNN*) son ampliamente utilizadas para procesar datos secuenciales, ya que permiten transferir información a lo largo de la secuencia al usar la salida de una neurona como entrada para la siguiente. Esto facilita el manejo, tanto de las características inherentes a los datos, como de las dependencias en la secuencia (Géron, 2019). Una variación de las redes neuronales recurrentes especial para manejar series de tiempo, son las redes neuronales Long Short Term Memory (*LSTM*), las cuales permiten representar patrones estacionales y además manejar series de tiempo complejas de término largo (Niño Medina et al., 2024).

Puesto que el problema del nowcasting es espacio-temporal, la arquitectura pionera en modelarlo de esta forma es la *ConvLSTM*, propuesta en 2015 (Xiong et al., 2024). En esta primera aproximación a esta arquitectura, se logró un mejor desempeño durante las primeras 6 horas y en todos los pasos de tiempo, en comparación con el algoritmo operacional basado en flujo óptico *ROVER* (X. Shi et al., 2015); gracias a su buen desempeño, han surgido numerosos trabajos posteriores que buscan mejorar su habilidad de pronóstico, generando así avances y nuevas variantes (Jianzhu et al., 2024). Una variación importante en la arquitectura basada en redes neuronales *ConvLSTM*, que transfiere información entre los pasos de tiempo, logró un desempeño a nivel del estado del arte superando modelos basados en flujo óptico como *CORTEC* y versiones anteriores basadas en *ConvLSTM* (L. Chen et al., 2020).

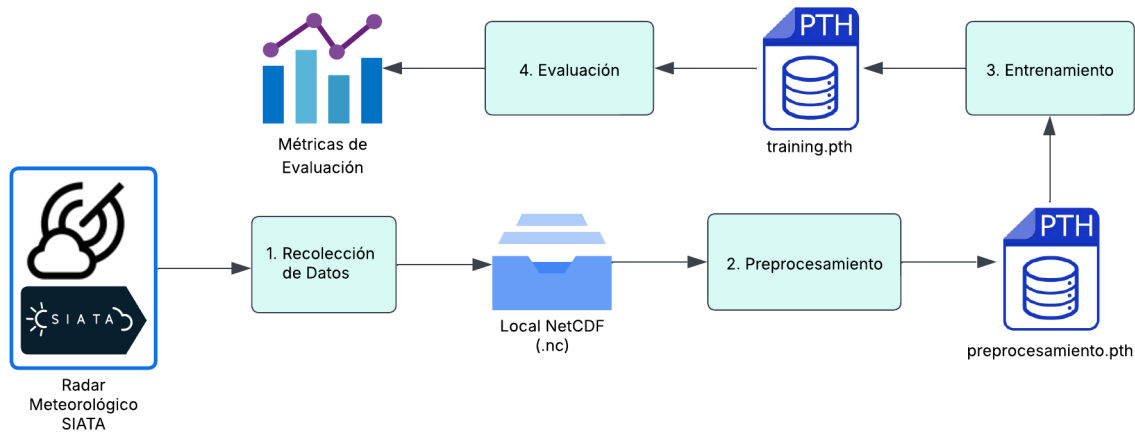


Figura 8: Diagrama de las fases del proceso de pronóstico de precipitación a corto plazo

4 Metodología

Con base en los objetivos planteados anteriormente, se definió una estrategia de trabajo compuesta por cuatro fases secuenciales, donde cada una de ellas fue diseñada con el fin de producir un componente intermedio. Esto permitió no solo garantizar la reproductibilidad y la experimentación de cada fase, sino también proporcionar el insumo necesario para la siguiente etapa del proceso (*Figura 8*)

Como punto de partida, se llevó a cabo la fase de recolección de datos, en la cual se obtuvieron las imágenes del radar meteorológico del SIATA, almacenándolas de forma local en su formato original. Posteriormente, en la etapa de preprocesamiento, se construyó el conjunto de datos utilizando como insumo las imágenes crudas recopiladas en la fase anterior, y generando una versión procesada en el formato adecuado para el entrenamiento del modelo. La tercera fase correspondió al entrenamiento, en la cual se aplicaron técnicas de aprendizaje profundo para entrenar el modelo ConvLSTM, obteniendo como resultado un archivo *.pth* que contiene los pesos del modelo. Finalmente, dicho archivo, junto con los datos preprocesados, fue empleado en la fase de evaluación, donde se examinó el desempeño del modelo utilizando criterios tanto cualitativos como cuantitativos, con el propósito de estimar su eficacia en el nowcasting.

4.1 Recolección y selección de Datos

Esta fase inicia con la conexión al repositorio de las imágenes de radar mediante el protocolo de transferencia de archivos (*FTP*), provisto por el SIATA. Posteriormente, mediante la descarga y descompresión del archivo, se obtuvieron las imágenes de radar en formato NetCDF (*.nc*). Las imágenes de radar se recolectaron de manera automática mediante la ejecución de un código en Python desarrollado en el proyecto, con el fin de recolectar la mayor cantidad de datos posible, tomando como entrada el año a descargar

Para este trabajo se obtuvieron los datos correspondientes a los años 2018, 2019 y 2020, teniendo así un total de *283.329* imágenes de radar. Además de las imágenes, el código se encargó de generar archivos de metadatos por cada año, recolectando la fecha de cada imagen (cada una representa un paso en el tiempo), su tamaño en *KB* y la cantidad de píxeles en la dimensión de latitud y longitud con el fin de realizar análisis preliminares (*Figura 9*).

Cada imagen cruda cuenta con un tamaño en píxeles de *1728 x 1728*, cubriendo el rango máximo del radar me-

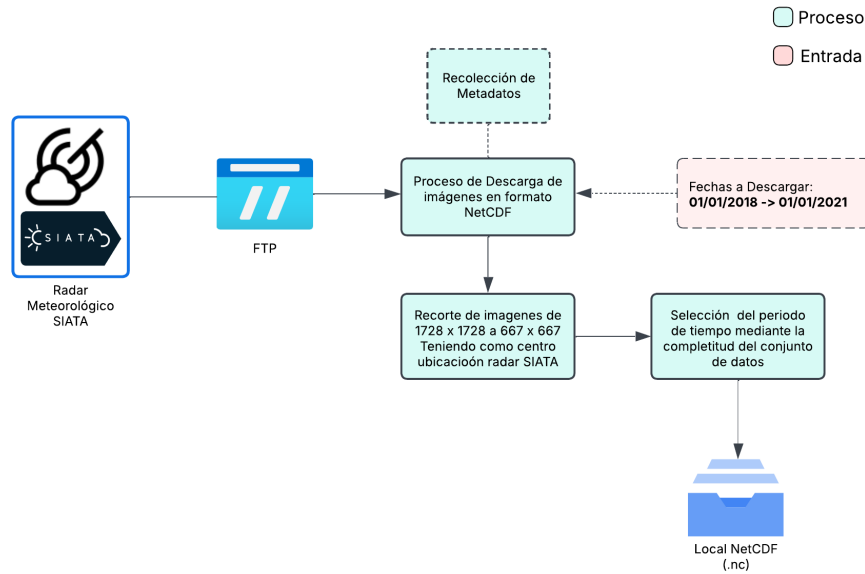


Figura 9: Diagrama detallado del proceso de recolección de datos

teorológico, aproximadamente 120 km, con una elevación de 1° . La variable medida por píxel corresponde al factor de reflectividad horizontal en dBZ . Antes de almacenar cada imagen de manera local, éstas se recortan tomando como centro de referencia la ubicación del radar y una región de 0.5° de ancho en latitud y longitud, obteniendo así imágenes de 667×667 píxeles, como se muestra en la figura 10.

Si bien la recolección de datos se realizó para tres años, se decidió escoger un periodo inicial de tres meses para la realización del experimento, debido a limitaciones en los recursos computacionales (como el uso de memoria RAM y unidad de procesamiento gráfico GPU) asociadas al tamaño de las imágenes, la complejidad del modelo y el volumen total de datos. Este periodo fue seleccionado con base en un análisis de completitud del conjunto, priorizando secuencias con una latencia máxima de 6 minutos entre imágenes, además de consideraciones relacionadas con las dinámicas atmosféricas de la zona.

El radar meteorológico del SIATA, en condiciones ideales entrega sus datos cada 5 minutos (SIATA, 2025). Por lo tanto, el primer análisis de completitud se realizó calculando el porcentaje de imágenes con una frecuencia de 5 minutos para cada mes de los tres años. Se encontró que los años 2019 y 2020 presentaban una cantidad similar de datos con esta resolución temporal mínima (Figura 11). Sin embargo, se eligió el 2020 debido a la ocurrencia del fenómeno de La Niña (CPC NOAA Team, 2025), ya que éste suele estar asociado con un incremento de precipitación para la zona (Arias et al., 2021).

Tras seleccionar el año 2020 como base del conjunto de datos para el ejercicio de nowcasting, se construyó un histograma mensual con el fin de contar la cantidad de imágenes disponibles en distintos intervalos de resolución temporal (Figura 12). En este se observa que los meses de septiembre, octubre y noviembre presentan la mayor cantidad de datos con una resolución temporal mínima de 6 minutos, un promedio de 19.43% de las imágenes en dicho intervalo y de 80.19% en el intervalo de 5 minutos.

Posteriormente, y como parte del proceso de experimentación del estudio, se seleccionó inicialmente el mes de noviembre, ya que presentaba la menor cantidad de datos (0.24%) con resoluciones temporales entre 14 y 16 minutos (como se muestra en el gráfico correspondiente a 2020-11 en la figura 12). De este modo, el conjunto de datos utili-

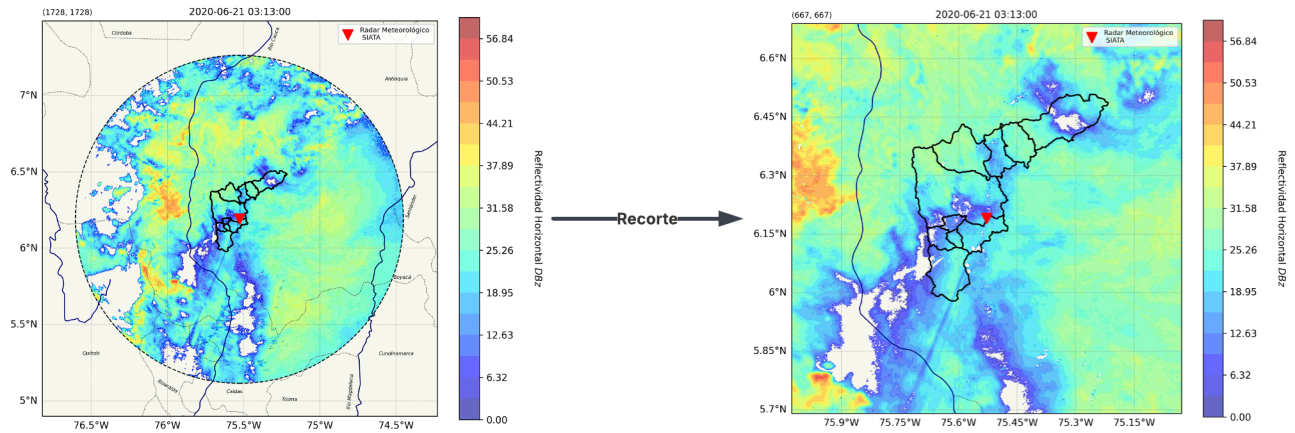


Figura 10: Proceso de recorte aplicado a una imagen de radar. A la izquierda se observa la imagen original del radar, con una resolución de 1728×1728 píxeles, correspondiente a un rango aproximado de 120 km. A la derecha se muestra la región recortada, centrada en el la ubicación del radar, con un área de 0.5° de ancho tanto en latitud como en longitud,

zado incluyó todas las imágenes comprendidas entre el 1 y el 17 de noviembre de 2020, periodo que mostró el mejor desempeño a lo largo del desarrollo del estudio.

Para finalizar, en esta sección se detalló la obtención automatizada de imágenes de radar del SIATA para los años 2018–2020, la evaluación de su completitud temporal mediante histogramas mensuales, y la selección del año 2020 con un periodo de tres meses, elegido con base en criterios de latencia mínima y dinámica atmosférica. Adicionalmente, se aplicó un recorte espacial para obtener imágenes de 667×667 píxeles centradas en la zona del Valle del Aburrá, y se consolidó el periodo del 1 al 17 de noviembre de 2020 como el conjunto de datos utilizado para el entrenamiento, validación y prueba del modelo.

4.2 Preprocesamiento

El preprocesamiento de los datos constituye una fase fundamental para el entrenamiento del modelo de aprendizaje profundo, ya que en esta etapa las imágenes de radar se transformaron no solo para adecuarlas al formato requerido, sino también para aplicar procedimientos computacionales y estadísticos que permitieran asegurar la coherencia de los datos y minimizar la pérdida de información antes de ser suministrada al modelo. El procedimiento en esta fase parte del uso de las imágenes de radar previamente recolectadas y seleccionadas, sobre las cuales se aplican las transformaciones necesarias, descritas en esta sección, para finalmente generar un archivo de tipo *.pth*, que contiene las secuencias de datos destinadas al entrenamiento, validación y prueba del modelo (Figura 13).

4.2.1 Concatenación de imágenes de radar y limpieza preliminar de datos

En esta sección se procesaron los archivos almacenados y seleccionados previamente en la recolección y selección de datos 4.1, los cuales se encontraban en formato *NetCDF*, con un archivo por cada paso de tiempo (01-11-2020 - 17-11-2020). La primera tarea consistió en concatenar todas las imágenes correspondientes al periodo definido, con el

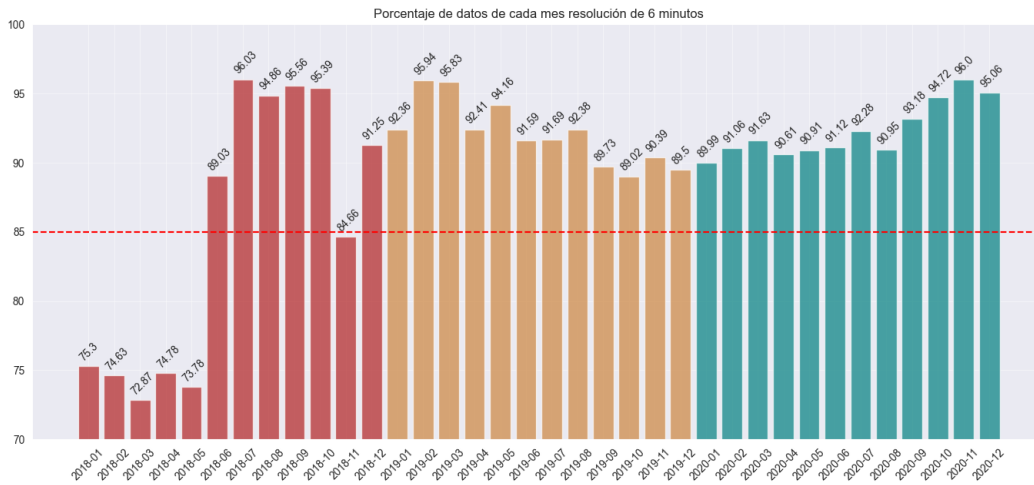


Figura 11: Porcentaje de cantidad de imágenes de resolución temporal de 5 minutos para cada Mes

objetivo de conformar un único arreglo en memoria que representara el conjunto completo de datos. Posteriormente, estos datos fueron convertidos a un tensor de PyTorch, permitiendo así el uso de GPU para los cálculos y acercándose al formato requerido para el entrenamiento del modelo.

Una vez conformado el tensor con dimensiones correspondientes a la cantidad de imágenes, latitud y longitud, se reemplazaron los valores menores a cero por cero. Esta operación tuvo como objetivo eliminar los valores negativos del factor de reflectividad, en particular el valor -32 , que indica ausencia de detección de reflectividad en ese punto de la imagen.

Si bien el factor de reflectividad Z puede tomar valores menores a cero, especialmente cuando el tamaño de las gotas es inferior a 1 mm (Doviak & Zrníc, 1972; Rauber & Nesbitt, 2018), se decidió filtrar todos los valores negativos debido a que generaban ruido en los datos, principalmente en zonas cercanas al radar, teniendo en cuenta que no se afectarían sistemas de precipitación, los cuales son el foco de estudio. Esta decisión se tomó con el fin de mejorar la calidad del conjunto de datos y evitar entradas distorsionadas que pudieran afectar el entrenamiento del modelo.

4.2.2 Reducción dimensional espacial y remuestreo

Una vez conformado el tensor con todas las imágenes de radar seleccionadas, se procedió a reducir su dimensionalidad espacial y a realizar un remuestreo temporal. La reducción espacial se llevó a cabo con el fin de optimizar la eficiencia computacional y minimizar la pérdida de información, mientras que el remuestreo aseguró una resolución temporal uniforme entre las imágenes, indispensable para la construcción de las secuencias empleadas en el entrenamiento del algoritmo.

Por tanto, para las transformaciones espaciales orientadas a la reducción de dimensionalidad, se llevó a cabo un análisis comparativo de diversos métodos computacionales y matemáticos, con el objetivo de seleccionar aquel que minimizara la pérdida de información de la imagen original. El análisis se hizo escogiendo cuatro tipos de imágenes de radar, en las que se pudiese ver distintos eventos de precipitación, los cuales se caracterizaban por la intensidad en el factor de reflectividad y por la distribución espacial de los eventos. (Figura 14)

Tras la selección de los cuatro eventos de precipitación representativos, se aplicó la reducción dimensional mediante tres métodos: agrupamiento de máximos ("Max Pooling"), agrupamiento de promedios ("Average Pooling") e inter-

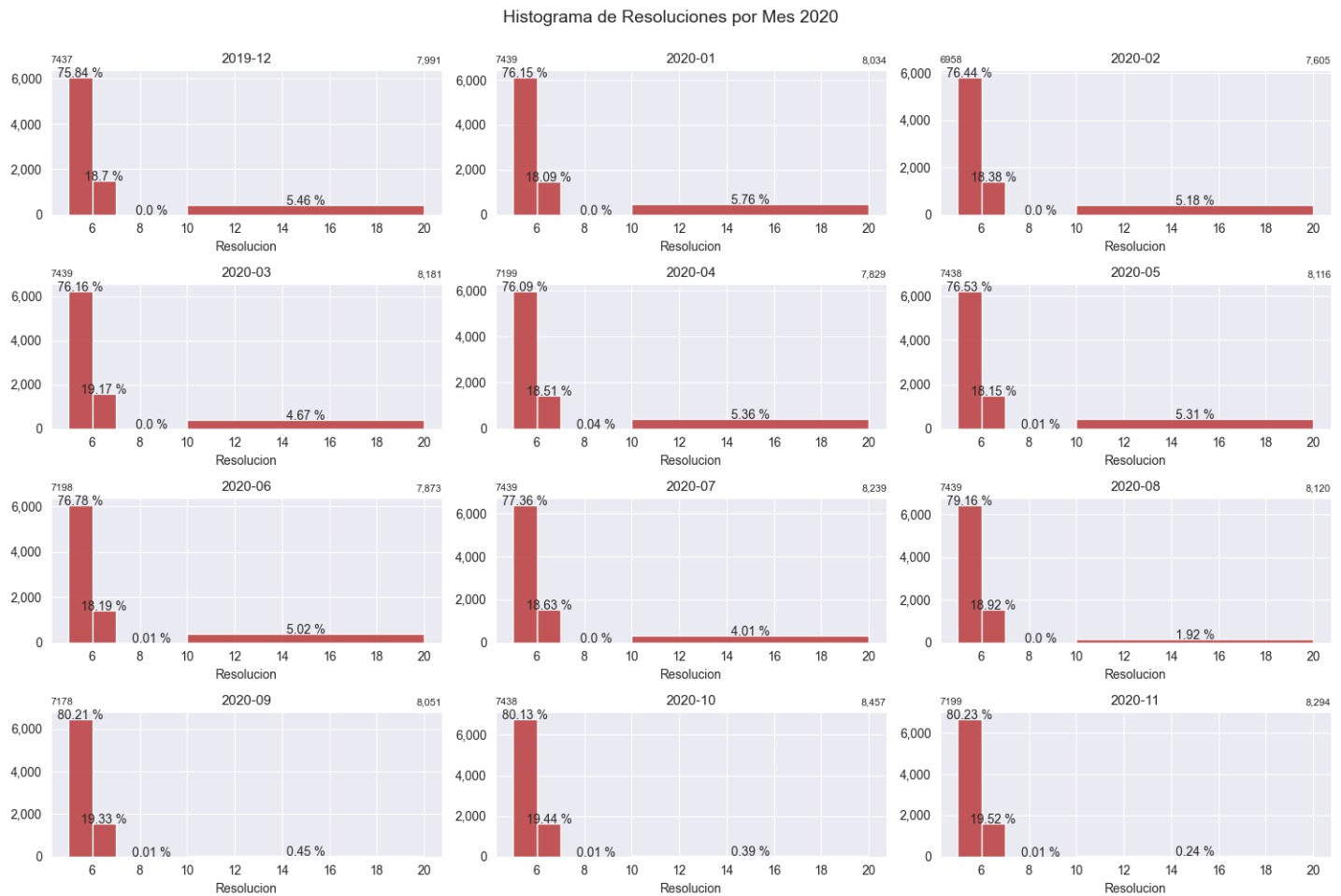


Figura 12: Histograma de la distribución de resoluciones para el conjunto de datos en el año 2020

polación bilineal (‘‘Interpolation’’), con los cuales el tamaño de las imágenes pasó de 667×667 a 110×110 píxeles (Figura 13). Si bien los métodos de agrupamiento ofrecen mayor eficiencia computacional, la interpolación bilineal refleja con mayor precisión la estructura inicial de los sistemas de precipitación y sus valores de reflectividad. Los algoritmos de agrupamiento tienden a engrosar los límites de los sistemas donde no hay medida de reflectividad, mientras que la interpolación preserva su forma original. Además, el agrupamiento por máximos suele sobrestimar la intensidad inicial, y el agrupamiento por promedios disipa los valores, subestimando la intensidad de los eventos.

La interpolación bilineal fue la última transformación espacial aplicada a los datos, reduciendo una imagen cruda de 1728×1728 píxeles a 110×110 (Figura 15), enfocando la zona de estudio y logrando un tamaño significativamente menor sin comprometer la calidad de la información, con el fin de optimizar el rendimiento computacional y facilitar el entrenamiento del modelo.

Tras la reducción espacial de las imágenes, se aplicó un remuestreo temporal mediante interpolación lineal en la dimensión temporal, con el objetivo de asegurar una resolución constante de 6 minutos en todo el conjunto de datos. Esta etapa fue necesaria debido a que los escaneos originales presentaban intervalos irregulares, en algunos

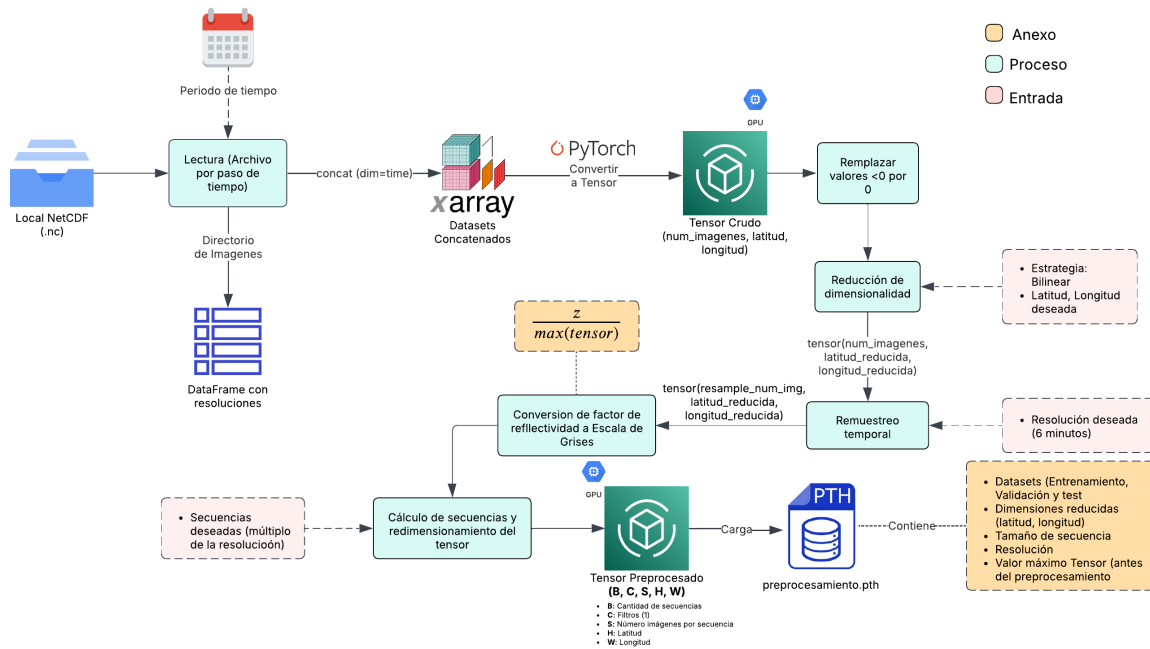


Figura 13: Diagrama detallado del proceso de preprocesamiento

casos superiores a 5 minutos, y la arquitectura *ConvLSTM*, al funcionar como una red neuronal recurrente, procesa secuencias de mapas de reflectividad sin incorporar la resolución temporal como atributo explícito. Por lo tanto, mantener un intervalo fijo garantiza que cada predicción se alinee correctamente con el paso temporal definido por la secuencia de entrada.

Como paso final antes de seleccionar las secuencias del conjunto de datos, se realizó una transformación de todas las imágenes a escala de grises (*Ecuación 2*), la cual consistió en dividir el valor de reflectividad de cada píxel por el valor máximo de reflectividad presente en todo el conjunto de datos. Esta conversión permitió entrenar las redes neuronales convolucionales utilizando un único canal, y además, al escalar los valores al rango $[0, 1]$, se facilitó la convergencia del modelo durante el entrenamiento (Géron, 2019; X. Shi et al., 2015).

$$Z_{gris} = \frac{Z}{\max(tensor)} \quad (2)$$

4.2.3 Cálculo de secuencias y adecuación de formato

El nowcasting es un problema de pronóstico espacio-temporal que se formula como la predicción de una o varias imágenes de radar en pasos de tiempo futuros, a partir de una secuencia de imágenes de entrada (Jianzhu et al., 2024; X. Shi et al., 2015), permitiendo así, que el algoritmo ajuste sus parámetros dada una secuencia de imágenes. De esta forma el conjunto de datos preprocesado anteriormente, se convirtió en secuencias de imágenes, donde cada secuencia contenía 30 imágenes de radar con 6 minutos de resolución, teniendo así, 8 secuencias por día (*240 imágenes de radar diarias*), con un total de 128 secuencias para todo el conjunto de datos utilizado para el ejercicio principal de nowcasting en este trabajo (01-11-2020 - 17-11-2020).

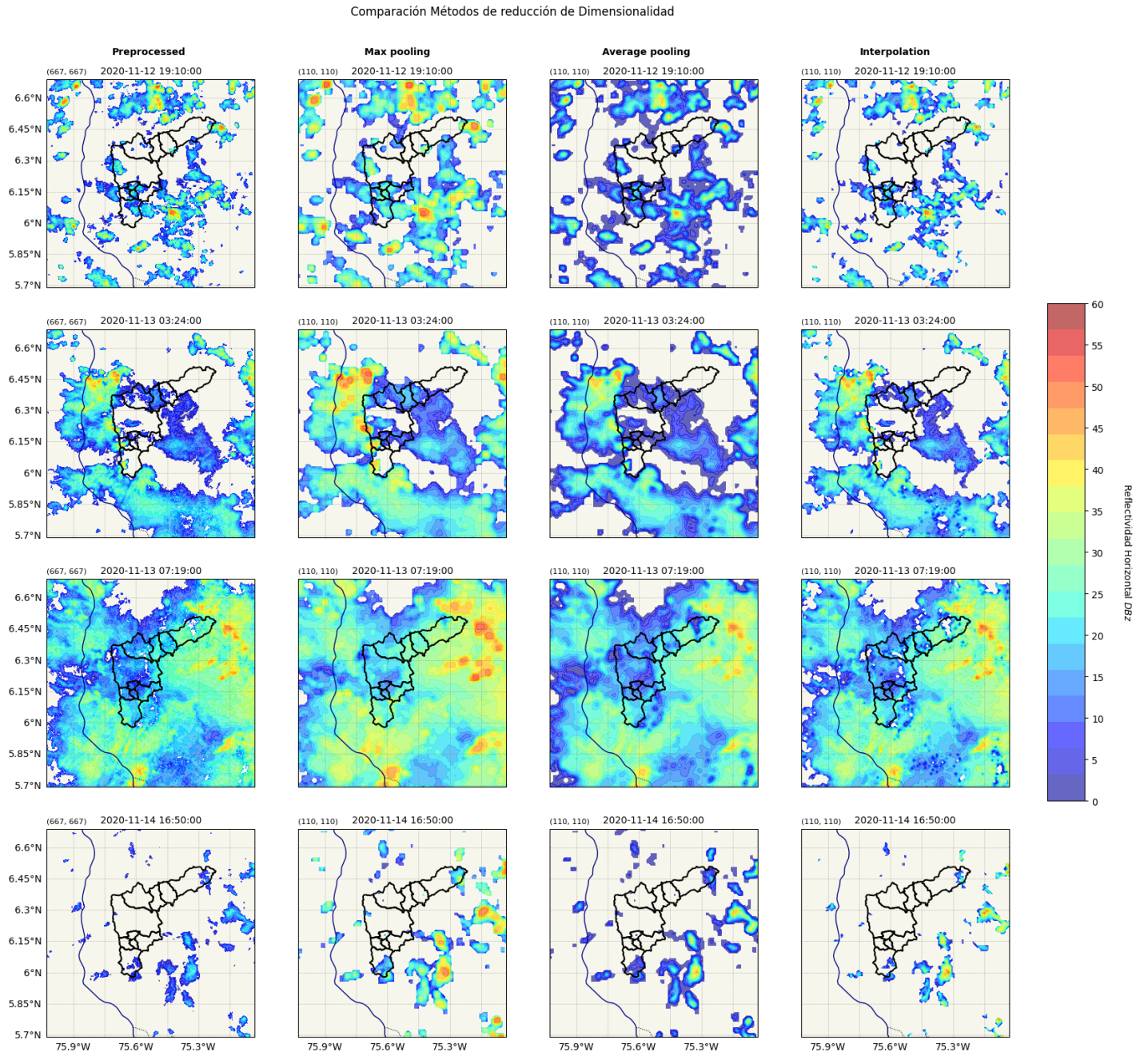


Figura 14: Métodos y eventos de análisis de reducción dimensional, donde cada fila representa la misma imagen de radar y las columnas son las reducciones dimensionales

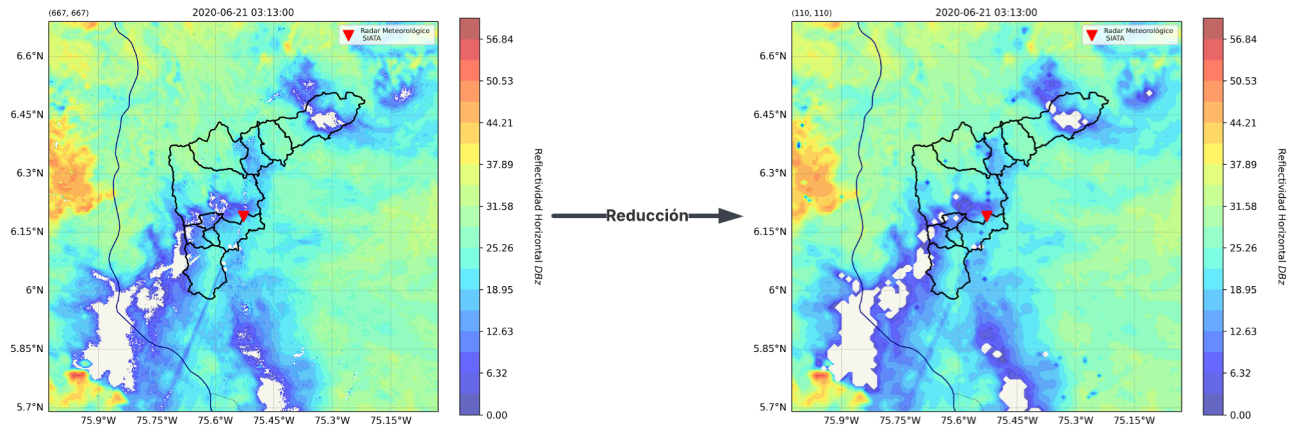


Figura 15: Métodos y eventos de análisis de reducción dimensional, donde cada fila representa la misma imagen de radar y las columnas son las reducciones dimensionales

Posteriormente, las secuencias de todo el conjunto de datos fueron divididas en tres grupos: entrenamiento, validación y prueba. El conjunto de entrenamiento, correspondiente al 70 % (90 secuencias) de los datos, se utilizó para entrenar el modelo y ajustar sus parámetros. El conjunto de validación, que representó el 10 % (12 secuencias), permitió observar el desempeño del modelo a lo largo de las épocas y monitorear la pérdida sobre datos no utilizados durante el entrenamiento. Finalmente, el conjunto de prueba, correspondiente al 20 % (25 secuencias), se empleó para evaluar el rendimiento final del modelo.

Cada secuencia constituye una instancia para el entrenamiento, la validación o las pruebas, en donde cada una contiene 29 imágenes de radar utilizadas como entrada para realizar la predicción, y una imagen adicional correspondiente al paso siguiente en el tiempo, que se emplea como referencia para entrenar, validar o evaluar el modelo, respectivamente. Además, debido a la naturaleza convolucional de la red neuronal, el formato requerido para el entrenamiento es un tensor de cinco dimensiones (B, C, S, H, W); donde B representa la cantidad de instancias (o secuencias), C el número de canales (que en este caso es 1, al tratarse de imágenes en escala de grises), S la cantidad de imágenes por secuencia (30 en total), y H y W el tamaño en píxeles a lo largo de los ejes vertical y horizontal, respectivamente.

4.3 Entrenamiento y predicciones

El proceso de entrenamiento del modelo comienza con la selección de un subconjunto de las secuencias disponibles, correspondiente al 70 % de los datos totales, obtenidos en el paso anterior. Cada una de estas secuencias, compuesta por 30 imágenes consecutivas de radar, se considera una instancia individual dentro del proceso de entrenamiento de la red neuronal. De cada secuencia, se utilizan las primeras 29 imágenes como entrada del modelo, mientras que la imagen número 30 se emplea como etiqueta, con el fin de comparar la salida generada por el modelo con la observación real y calcular el error correspondiente.

La red neuronal se entrena utilizando secuencias de imágenes de radar organizadas en lotes, controlados por un hiperparámetro denominado *tamaño de lote*. En este caso, dicho valor se fijó en 1, lo que significa que en cada paso del entrenamiento se calcula el error y se actualizan los pesos de la red a partir de una única secuencia. De esta

forma, cada lote contiene exactamente una secuencia, y el total de lotes corresponde al conjunto completo de datos de entrenamiento.

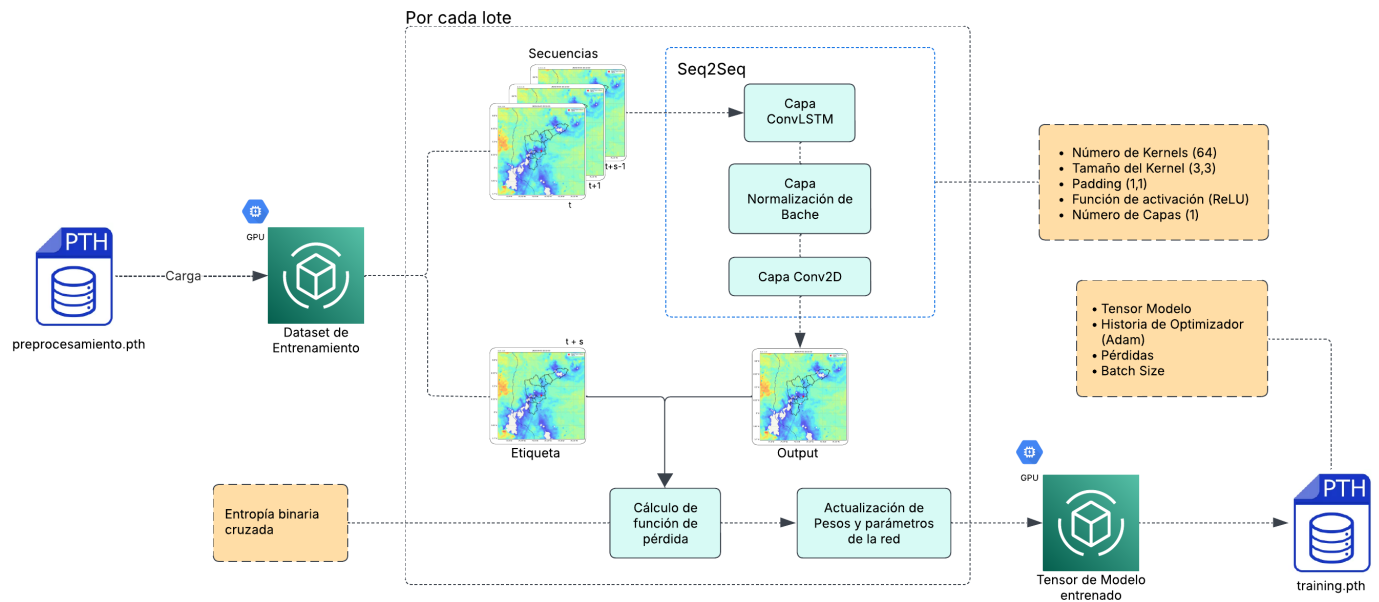


Figura 16: Diagrama detallado del proceso de entrenamiento del modelo, donde se puede diferenciar la forma de entrenar por lote, las secuencias utilizadas y la forma en que se mide el error e inicia el ciclo de propagación y cambio de gradientes.

Una pasada completa del modelo de entrenamiento sobre todos los lotes se conoce como una época. De esta manera, el proceso de entrenamiento se lleva a cabo durante 25 épocas. Al finalizar, se genera un tensor que contiene los parámetros aprendidos por la red neuronal. Este modelo entrenado se utiliza posteriormente para realizar predicciones, tanto sobre el conjunto de evaluación como sobre otras secuencias no vistas por el conjunto de entrenamiento. El modelo recibe como entrada una secuencia de 29 imágenes de radar y tiene como salida generar una predicción de la imagen número 30 (Figura 16).

4.3.1 Arquitectura del modelo

La arquitectura del modelo está compuesta por tres tipos principales de capas (Figura 16, línea punteada azul). En primer lugar, se encuentra la capa *ConvLSTM*, encargada de capturar los patrones espacio-temporales presentes en la secuencia de entrada; esta capa recibe como entrada la serie de imágenes de radar. En segundo lugar, se incorpora una capa de normalización por lote (*Batch Normalization*), que ajusta la escala de las salidas de las capas anteriores, facilitando que el modelo aprenda de manera más eficiente y estable (Géron, 2019). Por último, se utiliza una capa convolucional bidimensional (*Conv2D*), cuya función es transformar la salida del bloque anterior en una única imagen que representa la predicción del siguiente paso temporal. Esta imagen es la que se compara con la observación correspondiente, permitiendo así calcular el error y retroalimentar el modelo durante el entrenamiento. A continuación, se describe con mayor detalle el rol y funcionamiento de las capas que componen el modelo.

ConvLSTM

Esta capa es la más importante del modelo y da nombre a la arquitectura *ConvLSTM*. Está basada en el principio de las redes neuronales recurrentes del tipo *LSTM* (*Long Short-Term Memory*, por sus siglas en inglés), las cuales

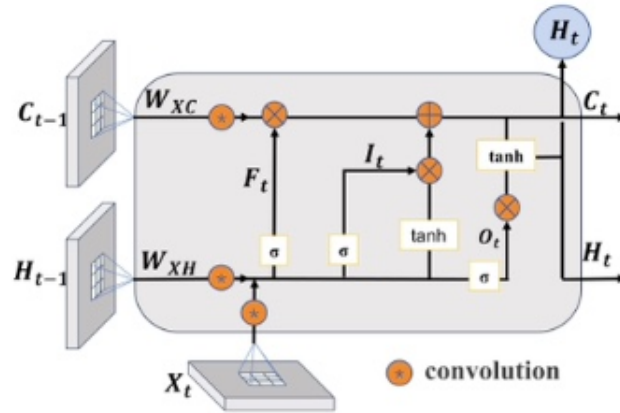


Figura 17: Imagen de una celda de ConvLSTM donde se pueden ver las compuertas de entrada, de olvido, de celda, de salida y el estado oculto i_t, f_t, C_t, o_t, H_t . En esta imagen se puede ver el flujo de los datos dentro de una neurona. (C. Shi et al., 2022)

introducen una estructura interna denominada celda de memoria, en lugar de utilizar neuronas simples. Esta celda permite conservar información a lo largo de múltiples pasos de tiempo, lo cual es fundamental para capturar dinámicas temporales en secuencias de datos. Para lograrlo, las celdas disponen de distintas compuertas, cada una con una función específica relacionada con la retención de la información (Figura 17).

Es importante resaltar que las redes neuronales LSTM, debido a su arquitectura, dependen tanto de las salidas del paso anterior como de las entradas del paso actual. Por esta razón, en las ecuaciones 4–7 se emplean los subíndices t para referirse al paso de tiempo actual y $t - 1$ para representar el paso de tiempo anterior. De esta forma, la letra H representa el estado corriente de la red neuronal, es decir, lo que podría considerarse como el “corto plazo” (incluso es el estado que se utiliza como predicción en la última imagen de la secuencia). Por otro lado, la letra C contiene información que se conserva a lo largo de la red. A este estado se le denomina “largo plazo”, ya que tiene como objetivo capturar información proveniente de pasos de tiempo anteriores. Por otro lado, los términos W y b son los parámetros aprendidos por la red neuronal.

Para comprender el funcionamiento interno de la celda de memoria, es necesario describir el papel que cumple cada una de sus compuertas. Por lo tanto, el estado C y lo que se “recuerda” de éste, se maneja mediante la compuerta de olvido f_t (*forget gate*, Ecuación 3), la cual se encarga de recibir como entrada el paso de tiempo corriente X_t y el estado oculto anterior H_{t-1} . A través de una función sigmoide σ (que entrega valores entre 0 y 1), esta compuerta define qué porcentaje del estado de la celda C se conservará en el paso de tiempo t .

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \otimes C_{t-1} + b_f) \quad (3)$$

Posteriormente, la compuerta i_t (ecuación 4) se encarga, de manera similar, de calcular el porcentaje de memoria potencial que se va a recordar a largo plazo y que se sumará al estado C . Dicha memoria potencial se encuentra en la ecuación 5, dentro de la función *ReLU* (*Rectified Linear Unit*, por sus siglas en inglés), función no lineal encargada de conservar los valores positivos y convertir en cero los valores negativos.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \otimes C_{t-1} + b_i) \quad (4)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \text{ReLU}(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (5)$$

Por lo tanto, el estado a largo plazo de la celda C (ecuación 5) se actualiza teniendo en cuenta las compuertas anteriores, donde la compuerta f_t se multiplica por el estado a largo plazo anterior C_{t-1} , y la compuerta de entrada por la memoria potencial a largo plazo. Así, se obtiene un porcentaje a recordar del estado a largo plazo anterior, al cual se le suma un porcentaje a recordar proveniente del paso de tiempo actual, definiendo así el estado actual a largo plazo C_t .

Por último, se tiene la ecuación 6, que describe el funcionamiento de la compuerta de salida o_t , la cual actúa de manera similar a las compuertas de entrada y de olvido, definiendo un porcentaje de la memoria a corto plazo que debe conservarse. Por consiguiente, la memoria a corto plazo H_t (ecuación 7) se define en función del estado de la memoria a largo plazo, modulada por la compuerta de salida o_t .

$$O_t = \sigma(W_{xo} + W_{ho}h_{t-1} + W_{co} \otimes c_t + b_o) \quad (6)$$

$$h_t = o_t \otimes \text{ReLU}(c_t) \quad (7)$$

Si bien con lo anterior explicamos la forma en la que se maneja la temporalidad del problema de nowcasting, es importante destacar que el modelo también conserva y aprovecha la dimensión espacial de los datos. Esto se logra mediante las operaciones indicadas con el operador $*$ en las ecuaciones, el cual representa convoluciones aplicadas tanto a las entradas como a los estados ocultos anteriores. De esta manera, cada compuerta procesa no solo información temporal, sino también patrones espaciales locales relevantes para el pronóstico de precipitación, permitiendo que el modelo aprenda cómo evolucionan en el espacio las estructuras meteorológicas a lo largo del tiempo.

Normalización de Bache y Conv2D

La capa de normalización de bache se encuentra luego de las capas de *ConvLSTM*, y se encarga de normalizar la entrada de los datos, de manera tal que utiliza la media y la desviación estándar de los mismos para acercarse a una distribución normal y conservar la escala (Géron, 2019). Esta capa es importante para evitar que el gradiente se desvanezca o se explote (Jianzhu et al., 2024). De esta forma, en la arquitectura se dispone una capa *ConvLSTM*, seguida de una de normalización de bache, repitiéndose este patrón tres veces, para un total de tres capas *ConvLSTM* y tres de normalización de bache.

Por otro lado, luego de tener las seis capas apiladas, intercaladas entre ConvLSTM y Normalización de Bache, se encuentra al final de la red neuronal una capa convolucional de dos dimensiones (*Conv2D*), la cual recibe la última imagen producida por la secuencia. A esta imagen se le aplica la función de activación sigmoide σ , con el fin de obtener como salida una imagen de reflectividad con valores entre 0 y 1 (en escala de grises).

En las capas convolucionales utilizadas dentro del modelo, tanto en las *ConvLSTM* como en la *Conv2D*, se emplearon filtros de convolución de tamaño (3x3), los cuales recorren las imágenes y extraen información relevante de estas. Estos filtros corresponden a los parámetros que la red neuronal debe aprender durante el entrenamiento. En todas las capas se utilizaron 64 filtros, permitiendo al modelo extraer un conjunto diverso de características espaciales. Para preservar las dimensiones espaciales de las imágenes a lo largo del proceso de convolución, se aplicó un relleno (1,1), que consiste en añadir un borde de un píxel alrededor de cada imagen antes de aplicar los filtros; esto asegura que la salida conserve el mismo tamaño espacial que la entrada.

4.3.2 Entrenamiento por cada lote

Teniendo en cuenta la arquitectura explicada en la subsección anterior, el proceso de entrenamiento inicia con el tensor que contiene las secuencias, donde cada una representa una instancia que se alimenta a la red. De esta forma, el modelo genera como salida una imagen de radar por cada secuencia, utilizando únicamente las primeras 29 imágenes como entrada. Esta imagen de salida, correspondiente a la imagen número 30, representa la predicción del modelo durante la fase de entrenamiento para cada secuencia.

El proceso de entrenamiento, para cada secuencia, continúa con la comparación entre la imagen número 30 reservada de la secuencia y la predicción generada por el modelo a partir de dicha secuencia. El objetivo de esta comparación es medir el error entre la imagen de radar observada (la imagen 30 de la secuencia original) y la imagen producida por el modelo (Salida de la última capa Conv2D). Este error se calcula mediante la ecuación de entropía binaria cruzada, la cual penaliza con mayor severidad aquellas predicciones que se alejan significativamente del valor real, debido a su naturaleza logarítmica (Géron, 2019), y ha sido empleada en arquitecturas destinadas al problema de nowcasting de precipitación (Agrawal et al., 2019; Jianzhu et al., 2024; X. Shi et al., 2015).

Luego del cálculo del error, se realiza la retropropagación, encargada de recorrer todos los pesos de la red y calcular la contribución de cada parámetro al error total. A partir de esta información, se aplica el algoritmo Adam (*Adaptive Moment Estimation*) para estimar los gradientes y determinar cómo deben ajustarse los pesos de manera efectiva, con el objetivo de reducir el error en la siguiente iteración. Este tipo de algoritmos, conocidos como optimizadores, son los responsables de modificar los parámetros del modelo a lo largo del entrenamiento hasta alcanzar una solución óptima al recorrer todo el conjunto de datos (Géron, 2019).

En síntesis, por cada lote, los datos ingresan a la red neuronal en forma de tensor, con cada instancia y sus respectivas secuencias, reservando la última imagen de cada una. Posteriormente, la red produce una predicción por secuencia, la cual es comparada con la observación reservada; de esta manera, se calcula el error y se ajustan los parámetros de la red en cada capa. Esto significa que la red neuronal fue entrenada para optimizar la predicción de la imagen siguiente $t + 1$, recibiendo como entrada una secuencia de imágenes. Por lo tanto, las predicciones se basan únicamente en las secuencias ingresadas, y la red neuronal aprendió los pesos de cada una de sus capas con el objetivo de generar dicha imagen siguiente.

Cuando el algoritmo recorre todos los lotes del conjunto de datos, este recorrido corresponde a una época, como se explicó anteriormente; por lo tanto, el proceso de entrenamiento culmina tras completar 25 recorridos (valor definido para este hiperparámetro), acumulando gradientes y ajustando los pesos en cada paso por los lotes. Al finalizar este proceso, se obtiene un modelo entrenado que puede ser utilizado para realizar inferencias sobre el conjunto de prueba y generar predicciones.

4.4 Evaluación

Debido a que las predicciones generadas por el algoritmo son imágenes de radar con valores de factor de reflectividad (dBZ), su naturaleza es no probabilística, ya que el modelo produce una única estimación del estado futuro en lugar de una distribución de posibles resultados (Wilks, 2019). Por lo tanto, la evaluación de la habilidad predictiva del modelo se abordó desde dos enfoques: por un lado, considerando las predicciones discretas, evaluadas como un problema de clasificación binaria que determina si la predicción corresponde a la ocurrencia (1) o no ocurrencia (0) de un evento de precipitación; y por otro lado, mediante la comparación directa de los valores continuos del factor de reflectividad, con el objetivo de analizar la precisión de las estimaciones cuantitativas realizadas por el modelo.

Con el fin de comparar los resultados del modelo ConvLSTM frente a un enfoque de referencia, se utilizó el modelo de persistencia, el cual replica la última imagen observada (imagen 29) como predicción para los pasos siguientes.

Para ello, tanto las predicciones del modelo ConvLSTM como las del modelo de persistencia, junto con las imágenes observadas del conjunto de evaluación, fueron transformadas en tensores unidimensionales ($1D$). Esta transformación permite realizar comparaciones directas entre las secuencias de valores predichos y las observaciones, así como entre la secuencia generada por el modelo de persistencia y dichas observaciones.

4.4.1 Evaluación categórica del pronóstico de precipitación

Para esta sección, los valores tanto de las predicciones como del conjunto de evaluación se asignaron a 0 y 1 según un umbral de factor de reflectividad de $2dBZ$, el cual se escogió con el fin de eliminar ruido asociado a predicciones de ocurrencia de reflectividad en píxeles aislados, de esta forma, todos los valores inferiores al umbral se convirtieron en 0 y los superiores en 1. A continuación se construyó una tabla de contingencia (*Tabla 1*) para ilustrar el conteo de píxeles con pronóstico de ocurrencia o no de reflectividad, tanto en predicciones como en observaciones. A partir de ella se identificaron los siguientes casos:

- **Acierto:** predicción y observación positivas (verdadero positivo, TP).
- **Fallo:** observación positiva sin predicción (falso negativo, FN).
- **Falsa alarma o falso positivo:** predicción positiva sin observación (falsa alarma, FP).
- **Verdadero negativo:** predicción y observación negativas (negativo verdadero, TN).

Predicciones	Observaciones	
	Con reflectividad	Sin reflectividad
Con reflectividad	número de aciertos (TP)	número de falsas alarmas (FP)
Sin reflectividad	número de fallos (FN)	número de falsos negativo (TN)

Tabla 1: Matriz de contingencia 2×2 con el fin de ilustrar las evaluaciones categóricas del pronóstico de precipitación

De esta tabla de contingencia se derivan diversas métricas; para este estudio se seleccionaron la probabilidad de detección (POD , por sus siglas en inglés), el índice crítico de éxito (CSI) y la tasa de falsas alarmas (FAR). Estas métricas de evaluación fueron escogidas por su amplio uso en ejercicios de nowcasting de precipitación (Fang et al., 2022; X. Shi et al., 2015) y porque, en conjunto, permiten evaluar el pronóstico desde diferentes perspectivas.

$$POD = \frac{TP}{TP + FN} \quad (8)$$

La métrica de probabilidad de detección (POD) (ecuación 8) mide la proporción de verdaderos positivos respecto a todas las observaciones positivas, es decir, la fracción de eventos de precipitación que el modelo detecta correctamente en comparación con los registrados en las observaciones. Su valor oscila entre 0 y 1, donde 0 indica que el modelo no detectó ningún evento de precipitación y 1 la detección de todos los eventos.

Aunque la probabilidad de detección permite evaluar la capacidad del modelo para capturar la precipitación observada, no informa sobre su tendencia a sobrestimar la variable. Por ejemplo, si el modelo predice precipitación en toda la región, la POD sería igual a 1, pues abarcaría todos los eventos registrados, aun cuando la predicción resulte excesiva.

$$FAR = \frac{FP}{FP + TP} \quad (9)$$

Por otro lado, la tasa de falsas alarmas (*FAR*) (*Ecuación 9*) cuantifica la proporción de falsos positivos respecto al total de predicciones positivas del modelo, es decir, la división entre las falsas alarmas y la suma de aciertos y falsas alarmas. En contraposición a la *POD*, valores cercanos a 0 indican pocas falsas alarmas en comparación con los aciertos, mientras que valores próximos a 1 señalan que la mayoría de las predicciones positivas corresponden a falsas alarmas.

$$CSI = \frac{TP}{TP + FN + FP} \quad (10)$$

Si bien las métricas anteriores proporcionan información sobre la detección de eventos de precipitación y su sobrestimación, el índice crítico de éxito (*CSI*) (*Ecuación 10*) evalúa la relación entre los aciertos, los fallos de detección y las falsas alarmas. Esta métrica es adecuada en escenarios donde la categoría positiva (eventos de precipitación) ocurre con mayor frecuencia que la negativa (no precipitación) (Wilks, 2019).

El índice crítico de éxito (*CSI*) oscila entre 0 y 1, donde un valor de 0 indica que todas las predicciones del modelo son fallos de detección o falsos positivos, y un valor de 1 que todas son aciertos. Esta métrica es especialmente rigurosa porque considera ambas fuentes de error (no detecciones y falsas alarmas). No obstante, proporciona una visión más general de la capacidad predictiva del modelo.

4.4.2 Evaluación cuantitativa continua del pronóstico de precipitación

En esta sección se evaluó el modelo mediante la métrica de la raíz cuadrada del error cuadrático medio (*RMSE*, por sus siglas en inglés), expresada en la ecuación 11. A diferencia de las métricas categóricas, en este caso no se discretizan los valores entre 0 y 1, sino que se utilizan directamente los valores continuos de reflectividad tanto de las predicciones como de las observaciones. Esto permite evaluar la capacidad del modelo para aproximarse cuantitativamente a los valores reales, conservando así la naturaleza continua de los predicandos.

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (Y_k - O_k)^2}{n}} \quad (11)$$

Esta medida de desempeño no está acotada dentro de un rango fijo, como ocurre con las métricas categóricas presentadas en la sección anterior. Sin embargo, el *RMSE* constituye una métrica de precisión que cuantifica el error entre los valores predichos Y_i y los valores observados O_i para cada ocurrencia k . Para cada paso, los valores de la observación son abstraídos del valor pronosticado y elevado al cuadrado, lo cual hace que esta métrica sea más sensible a valores anómalos (Wilks, 2019), posteriormente son divididos por todas los eventos n . Al valor final, se le aplica la raíz cuadrada con el fin de interpretarlo con las mismas unidades de la variable medida *dBZ*, y tener así una métrica estándar la cual es 0 si es un pronóstico perfecto (Wilks, 2019).

5 Análisis de Resultados

Teniendo en cuenta los pasos y procedimientos explicados en la sección pasada, en esta sección se ilustraron los resultados del proceso de entrenamiento y evaluación, donde se analizaron las métricas presentadas en la subsección de Evaluación (4.4) y se examinaron las métricas del proceso de entrenamiento en cuanto al cambio de los valores de la función de pérdida a lo largo de las épocas. Se incluyen visualizaciones y tablas que permitieron comparar los resultados entre el modelo ConvLSTM, la Persistencia y las observaciones reales para las secuencias en el conjunto de evaluación.

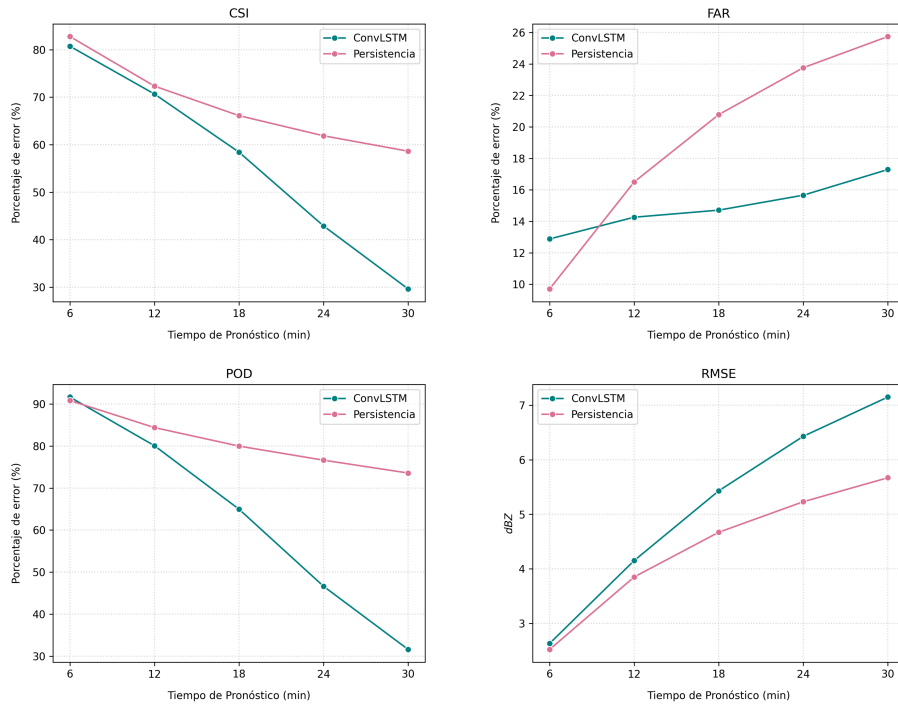


Figura 18: Diagrama de la evolución temporal de las métricas de rendimiento en todos los pasos del tiempo de pronóstico. Se puede observar en ésta el decaimiento del desempeño de ambos modelos con los pasos de tiempo

Por lo tanto, para el periodo de tiempo estudiado (01-11-2020 al 17-11-2020), en el cual las secuencias de evaluación corresponden a los días comprendidos entre el 13 y el 17 de noviembre, el pronóstico se definió con un horizonte de 30 minutos, con pasos de 6 minutos, lo que permitió analizar la evolución de las predicciones en distintos momentos futuros. Esta sección de resultados se organiza en 4 subsecciones: la primera evalúa la evolución del desempeño del modelo a lo largo de todo el conjunto de datos de prueba; la segunda presenta una tabla con las métricas correspondientes al pronóstico a 6 minutos para cada secuencia, destacando las secuencias con mejor y peor rendimiento del modelo ConvLSTM; la tercera pretende analizar el progreso de la función de pérdida en la fase de entrenamiento. Por último se expone otros experimentos en los que se varió la cantidad de datos utilizados para el entrenamiento del modelo, así como ciertos parámetros.

5.1 Evolución de pronóstico en el tiempo

La tabla 2 muestra las métricas de evaluación para los modelos ConvLSTM y Persistencia a lo largo de los cinco pasos de tiempo. En el caso del índice crítico de éxito (*CSI*), se observa que el modelo de Persistencia supera consistentemente al modelo ConvLSTM en todos los pasos, comenzando con una diferencia de aproximadamente un 2% en el primer paso (6 minutos) y alcanzando una diferencia cercana al 30% en el minuto 30. Esta brecha comienza a ampliarse de manera más marcada a partir del minuto 18 de pronóstico. De forma similar, la raíz cuadrada del error cuadrático medio (*RMSE*) muestra un deterioro progresivo en el desempeño, con un aumento de la diferencia entre ambos modelos a partir del segundo paso de tiempo.

Métrica	6min		12min		18min		24min		30min	
	ConvLSTM	Persistencia	ConvLSTM	Persistencia	ConvLSTM	Persistencia	ConvLSTM	Persistencia	ConvLSTM	Persistencia
CSI	80.72	82.76	70.66	72.33	58.42	66.11	42.89	61.87	29.63	58.63
POD	91.67	90.84	80.08	84.40	64.97	79.99	46.61	76.64	31.58	73.58
FAR	12.88	9.70	14.26	16.50	14.71	20.78	15.66	23.76	17.29	25.74
RMSE	2.63	2.52	4.15	3.85	5.43	4.67	6.43	5.23	7.15	5.67

Tabla 2: Comparación de métricas entre *ConvLSTM* y *Persistencia* con respecto a las observaciones en los pasos de tiempo pronosticados. Los mejores resultados para cada métrica por cada paso de tiempo están resaltados en negritas, para el modelo que obtuvo mejores resultados).

Por otro lado, la tasa de falsas alarmas (*FAR*) presenta un comportamiento en el que, en el primer paso de tiempo, el modelo de Persistencia muestra un desempeño superior por aproximadamente un 3%. No obstante, a medida que avanza el horizonte de predicción, el modelo ConvLSTM supera al de Persistencia en esta métrica para todos los pasos subsiguientes hasta el minuto 30. En cuanto a la probabilidad de detección (*POD*), el comportamiento es similar al observado en el *CSI*. Sin embargo, destaca que esta es la única métrica en la que el modelo ConvLSTM supera al de Persistencia en el primer paso de tiempo, con una ventaja de aproximadamente 1%.

El decaimiento del desempeño de ambos modelos con el tiempo (*Figura 18*) evidencia que, para las métricas de *POD* y *CSI*, el modelo ConvLSTM presenta una disminución más pronunciada a medida que avanzan los pasos de tiempo, en comparación con el modelo de Persistencia. Este modelo puede llegar a perder hasta un aproximado de 15% por paso, lo cual se refleja en la pendiente de sus curvas, donde el deterioro en el rendimiento del ConvLSTM es considerablemente más notorio que en el caso de la Persistencia. Este comportamiento también se observa en la métrica *RMSE*, donde el error del ConvLSTM aumenta de forma más pronunciada a medida que avanza el tiempo. Es notable que, en las tres métricas, aunque las curvas parten de valores muy similares en el primer paso de tiempo (minuto 6), estas divergen progresivamente, siendo el desempeño del modelo ConvLSTM consistentemente inferior en los pasos siguientes.

Para el caso de la *FAR*, cuyo comportamiento difiere del de las otras métricas, ambas curvas inician desde valores más distantes. A lo largo del horizonte de predicción, la persistencia incrementa sus valores de forma más pronunciada que el modelo ConvLSTM: mientras que este último experimenta un aumento aproximado del 5% entre el minuto 6 y el 30, la persistencia alcanza un incremento cercano al 10%. Este comportamiento indica que, en términos de falsas alarmas, el modelo ConvLSTM logra mantener un crecimiento del error más moderado a medida que avanza el tiempo.

Como se vio anteriormente, el *CSI* y la *POD* tienen un comportamiento similar. Sin embargo, la *POD* muestra diferencias más marcadas entre el modelo de Persistencia y el ConvLSTM, lo que indica que este último presenta un desempeño deficiente al detectar las ocurrencias de precipitación en las observaciones. En contraste, las diferencias en el *CSI* no son tan notorias, debido a que la *FAR* exhibe un descenso en el desempeño más estable para el modelo ConvLSTM. Dado que el *CSI* se ve afectado tanto por los falsos positivos (relacionados con la *FAR*) como por las omisiones en la detección (relacionadas con la *POD*), se puede deducir que el principal factor que contribuye al

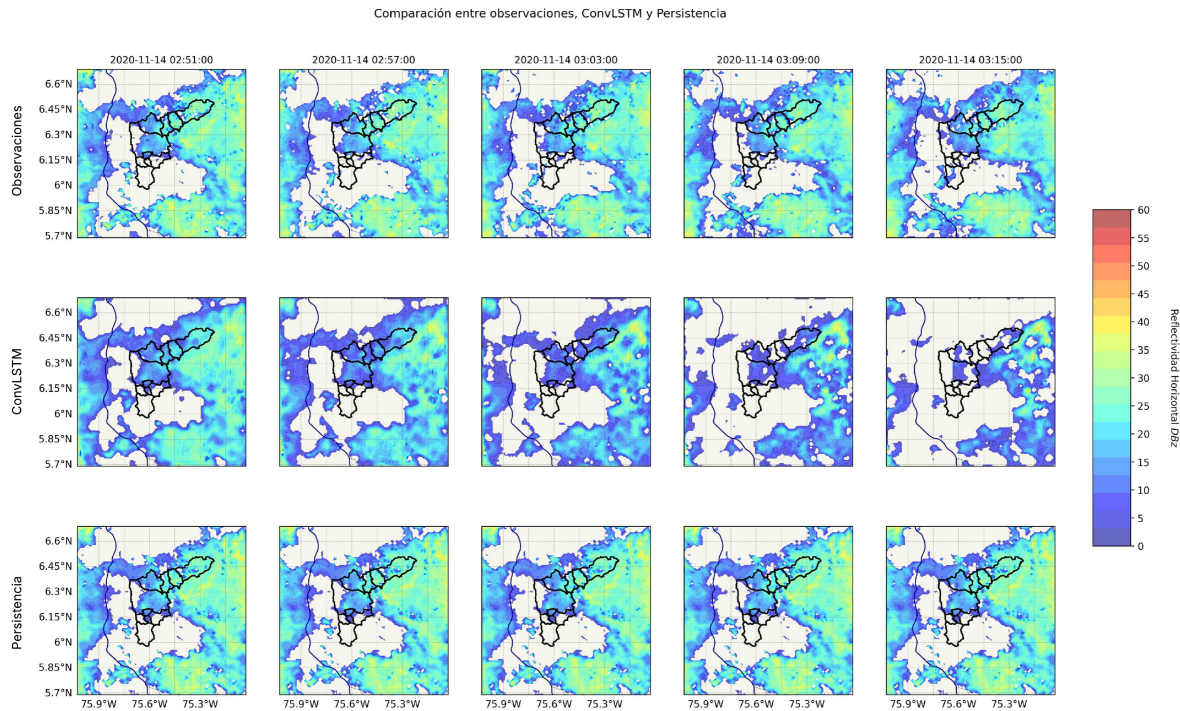


Figura 19: Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la mejor predicción en términos de *CSI* y *POD*, realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3

bajo desempeño general del modelo en términos de *CSI* es la *POD*, es decir, su capacidad limitada para identificar correctamente las ocurrencias de precipitación reales.

Si bien las métricas de evaluación ofrecieron hallazgos relevantes sobre la capacidad predictiva del modelo ConvLSTM en comparación con la Persistencia, estas cifras encuentran su justificación en la forma en que el pronóstico del ConvLSTM se debilita progresivamente con el avance de los pasos de tiempo. Esta tendencia se evidencia en las figuras 19–23, que muestran la evolución del pronóstico del modelo, junto con las observaciones reales y la predicción del modelo de persistencia, a lo largo del horizonte de 30 minutos. Esta pérdida de capacidad predictiva explica, en parte, que la tasa de falsas alarmas se mantenga relativamente constante, mientras que la probabilidad de detección presenta un descenso mucho más marcado.

La causa principal del deterioro progresivo y eventual desaparición de las predicciones del modelo ConvLSTM se relaciona directamente con la forma en que fue entrenado. El modelo fue entrenado únicamente para predecir el siguiente paso temporal ($t + 1$) a partir de una secuencia de entrada fija, lo que implica que no aprendió a corregir o adaptarse a los errores acumulados en predicciones sucesivas. Al extender su uso para predecir varios pasos hacia adelante (hasta 30 minutos), el error de cada predicción se propaga a las siguientes (Géron, 2019), generando un decaimiento pronunciado en la calidad de las imágenes generadas.

Una forma más adecuada de abordar este problema sería entrenar el modelo con secuencias completas de salida, correspondientes a todos los pasos de tiempo utilizados en la inferencia, permitiendo así calcular el error sobre toda la secuencia futura y mejorar su capacidad de generalización en horizontes de tiempo más largos (Géron, 2019). Sin embargo, esta aproximación implicaría que el modelo predijera tantos pasos de tiempo como entradas, por lo que una alternativa viable sería entrenar una red neuronal que codifique los pasos de tiempo anteriores en un vector de alta

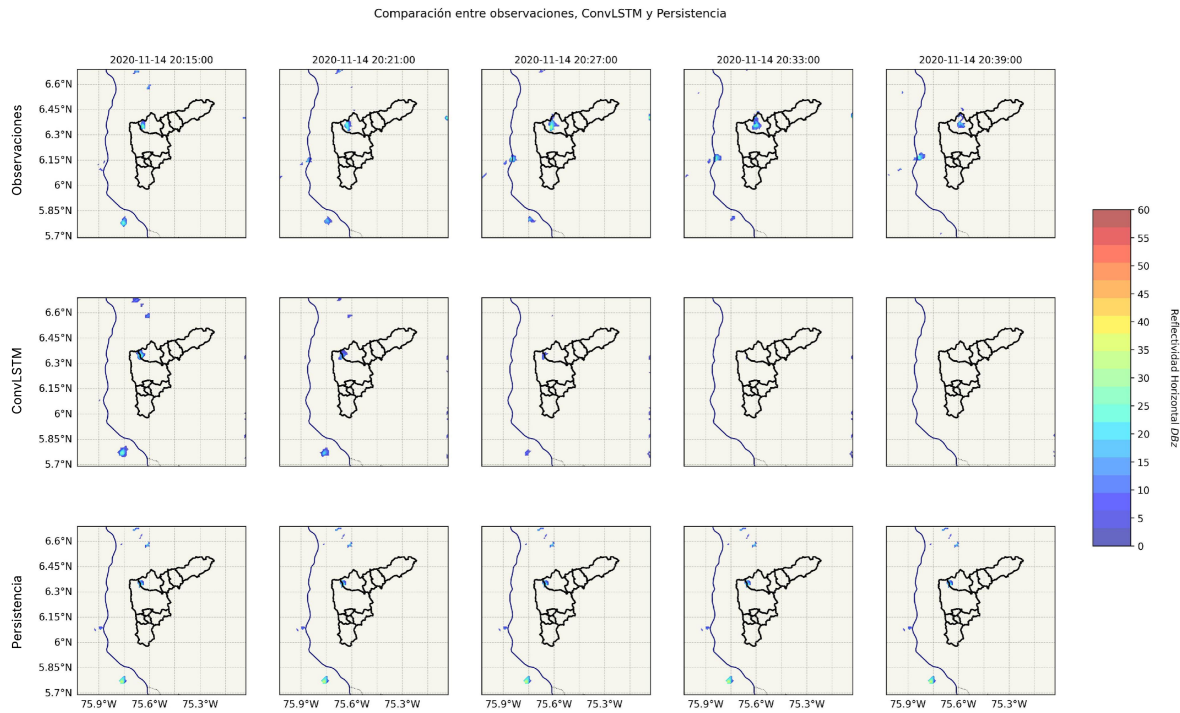


Figura 20: Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la peor predicción en términos de POD , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3

dimensionalidad y otra red neuronal que decodifique dicho vector en los pasos de tiempo requeridos (X. Shi et al., 2015; Sutskever et al., 2014), ($t + 5$) en este caso.

5.2 Primer paso temporal por secuencia: métricas y casos destacados

Este análisis se basó en los resultados presentados en la tabla 3, la cual muestra las métricas obtenidas para cada una de las 25 secuencias del conjunto de prueba en el primer paso de tiempo ($t + 1$). El enfoque en este paso se debe a que representa el horizonte temporal con mejor desempeño para el modelo ConvLSTM, además de ser el único paso directamente optimizado durante el entrenamiento. Esto permitió analizar distintos eventos de precipitación y evaluar el rendimiento de ambos modelos, representando gráficamente la evolución temporal de sus predicciones.

Los casos particulares seleccionados se basaron en los valores de las métricas de desempeño. Se eligió el evento de precipitación del **14 de noviembre a las 02:51**, que presentó el mejor desempeño tanto en la métrica CSI como en POD (Figura 19). También se consideró el evento del **14 de noviembre a las 05:45**, que mostró el peor valor de $RMSE$ pero el mejor valor de FAR (Figura 21). Además, se seleccionó el evento del **14 de noviembre a las 20:15**, que exhibió el peor desempeño en POD (Figura 20), y el del **15 de noviembre a las 02:03**, que presentó los peores valores en CSI y FAR , pero el mejor en $RMSE$ (Figura 22). Por otro lado, se realizó un análisis adicional de la predicción ubicada en la mediana (**14 de noviembre a las 23:09**) (Figura 23) de los valores de CSI (64.27%), ya que este caso representó un evento de precipitación distinto a los analizados anteriormente y representa el valor de la mitad en el amplio rango de valores de CSI (21.43% - 91.51%), siendo mayor por aproximadamente 10% de la media. Cabe resaltar que las secuencias que no contaban con eventos de precipitación, fueron excluidas de este

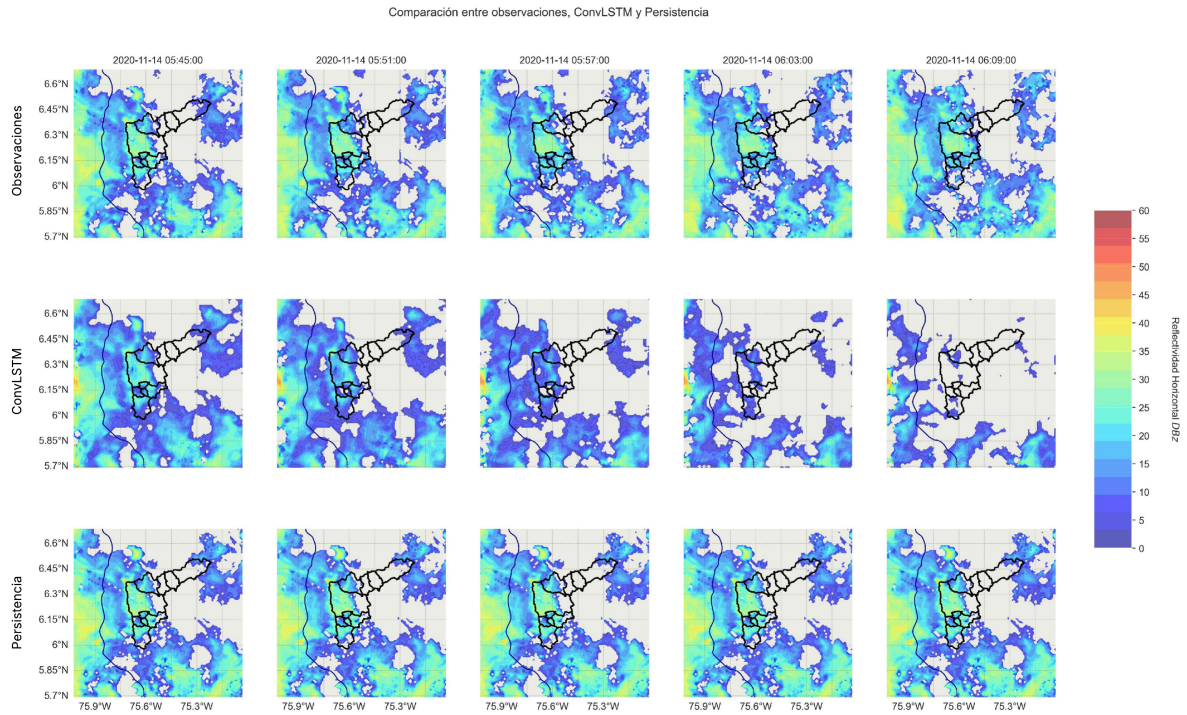


Figura 21: Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la peor predicción en términos de $RMSE$ y la mejor en términos de FAR , realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3

análisis, teniendo así un total de 20 secuencias en el conjunto de prueba (*Tabla 3*).

En el primer paso de tiempo, y en contraste con la evolución temporal observada a partir del minuto 18 en la tabla 2, se evidencia que la tasa de falsas alarmas es menor para el modelo de Persistencia en la mayoría de las secuencias (17 de 20). Esto se debe a que, como se muestra en las figuras 19–23, el modelo ConvLSTM tiende a engrosar los bordes de los sistemas de precipitación, estimando precipitación en zonas donde no se registró precipitación en las observaciones.

En contraposición, el modelo ConvLSTM supera al modelo de Persistencia en la métrica POD en más de la mitad de las secuencias (13 de 20). Este mejor desempeño está relacionado con el comportamiento observado en la FAR : al abarcar una mayor región en sus predicciones de precipitación, el modelo ConvLSTM incrementa la probabilidad de coincidir con los eventos presentes en las observaciones. Esto refleja la compensación inherente entre estas métricas categóricas, donde la mejora en la detección de eventos (POD) puede implicar un incremento en las falsas alarmas (FAR), y viceversa.

En cuanto al CSI , el modelo de Persistencia mostró un mejor desempeño en 17 de las 20 secuencias, coincidiendo con la cantidad de secuencias en las que también obtuvo mejores resultados en la FAR . Aunque no se trata de una correspondencia directa entre cada secuencia, este patrón sugiere que el bajo desempeño del modelo ConvLSTM en el índice crítico de éxito está fuertemente influenciado por su alta tasa de falsas alarmas, lo que penaliza su valor de CSI a pesar de una POD relativamente competitiva.

El $RMSE$ para el primer paso de tiempo es muy similar entre el modelo ConvLSTM y la Persistencia (*tabla 2*), con un valor ligeramente mayor para la Persistencia por $0.19dBZ$. Si bien el $RMSE$ penaliza con mayor severidad

Fecha (24hr)	RMSE		CSI		POD		FAR	
	ConvLSTM	Persistencia	ConvLSTM	Persistencia	ConvLSTM	Persistencia	ConvLSTM	Persistencia
2020-11-13 21:03	3.438	3.397	68.03	70.15	85.42	82.58	23.04	17.67
2020-11-13 23:57	4.313	4.422	81.97	82.74	91.22	88.23	11.01	6.99
2020-11-14 02:51	4.622	4.106	91.51	91.61	97.40	95.88	6.19	4.63
2020-11-14 05:45	4.664	3.516	90.03	91.89	95.10	95.26	5.58	3.71
2020-11-14 08:39	4.450	3.702	85.45	84.66	95.10	94.50	10.61	10.95
2020-11-14 11:33	2.846	2.918	79.72	81.54	89.13	91.11	11.69	11.41
2020-11-14 14:27	2.122	2.592	59.85	57.42	79.63	69.29	29.33	22.98
2020-11-14 17:21	1.707	1.990	65.46	64.31	78.83	78.64	20.58	22.08
2020-11-14 20:15	0.918	0.997	27.64	38.86	53.97	53.97	63.83	41.88
2020-11-14 23:09	1.664	1.731	64.27	74.88	90.19	86.10	30.90	14.82
2020-11-15 02:03	0.383	0.405	21.43	52.83	75.00	77.78	76.92	37.78
2020-11-15 07:51	0.796	0.772	63.77	80.10	90.26	84.62	31.52	6.25
2020-11-15 16:33	2.231	2.310	57.44	67.42	75.03	75.72	28.98	13.99
2020-11-15 19:27	3.179	3.291	70.07	69.92	88.32	83.71	22.78	19.07
2020-11-15 22:21	3.612	3.225	82.56	85.56	90.05	90.93	9.15	6.47
2020-11-16 01:15	4.184	3.899	88.38	89.47	97.44	96.65	9.51	7.67
2020-11-16 04:09	2.033	1.706	82.06	83.77	90.07	92.66	9.78	10.27
2020-11-16 07:03	1.242	1.535	71.24	71.44	86.41	82.36	19.77	15.65
2020-11-16 09:57	0.585	0.661	50.16	60.00	82.56	81.54	43.90	30.57
2020-11-16 15:45	0.999	1.219	44.11	49.42	75.21	72.65	48.39	39.29
2020-11-16 18:39	2.642	3.375	65.18	70.50	79.39	80.94	21.55	15.47

Tabla 3: Tabla de Comparación del desempeño de predicción de 6 minutos ($t + 1$) para todas las secuencias en el conjunto de prueba, donde el modelo que mejor desempeño tuvo se encuentra en **negrilla**, y para cada métrica se encuentra el valor en **verde** para el mejor valor por métrica de ConvLSTM y en **rojo** para el peor

los errores en valores atípicos (Wilks, 2019), en el caso del modelo ConvLSTM no se evidencia una sobrestimación significativa de la precipitación. Una excepción puntual se observa en la figura 21, específicamente en la región de 6.15°N y 75.9°W , donde se presenta un aumento localizado en la reflectividad. No obstante, a pesar de que esta secuencia corresponde al peor *RMSE*, es también la que presenta la mejor tasa de falsas alarmas, lo que indica que su bajo desempeño en esta métrica no se debe a una cobertura excesiva de precipitación, sino más bien a los errores en los valores de intensidad. De hecho, se observa que el modelo tiende a subestimar los valores cercanos a los bordes de los sistemas de precipitación, haciendo que estas zonas de baja reflectividad se representen de forma más extensa (Figuras 19, 21, 18).

Se puede evidenciar que las dos secuencias de imágenes de radar con las métricas categóricas de rendimiento más bajo (Figuras 20, 22) corresponden a eventos de precipitación de menor tamaño en comparación con otras secuencias, con valores de *CSI* inferiores al 30% en ambos casos. En contraste, las imágenes que presentan sistemas de precipitación más amplios y compuestos, como las de las figuras 19 y 21, alcanzan valores de *CSI* superiores al 90%. Por su parte, la secuencia correspondiente a la mediana del *CSI* (Figura 23) muestra sistemas de precipitación algo mayores que los de menor desempeño, pero considerablemente más pequeños y menos estructurados que los de mejor rendimiento.

Lo anterior sugiere una relación entre el desempeño del modelo ConvLSTM y el tamaño de los sistemas de precipitación, la cual puede explicarse por la cantidad de puntos evaluados en cada caso. Dado que las métricas categóricas consideran únicamente aciertos, fallos y falsos positivos, un sistema de mayor tamaño implica más puntos con precipitación en las observaciones, lo que ofrece mayor margen para aciertos. Esto indica que el modelo es sensible al tamaño de las estructuras de precipitación y que su capacidad para localizar correctamente los eventos disminuye cuando se trata de sistemas más pequeños. No obstante, la secuencia con la *POD* más baja muestra que el modelo sí logró generar estructuras de precipitación similares a las observadas, tanto en forma como cercanas en ubicación (Figura 20).



Figura 22: Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la mejor predicción en términos de *RMSE* y peor en términos de *FAR*, realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3

5.3 Evolución de la función de pérdida en el entrenamiento

Como se mencionó anteriormente, el proceso de entrenamiento del modelo ConvLSTM tiene como objetivo disminuir la función de pérdida a lo largo de las épocas. Por esta razón, se recopilieron los valores de la función de pérdida tanto para el conjunto de entrenamiento como para el conjunto de validación. El conjunto de validación se utiliza únicamente para evaluar el desempeño del modelo durante el entrenamiento, pero no interviene en el ajuste de los pesos del modelo. Por lo tanto, su propósito es observar cómo se comporta el modelo frente a datos que no ha visto previamente en el proceso de aprendizaje (Géron, 2019).

Durante el proceso de entrenamiento, se observa que los valores de la función de pérdida para el conjunto de entrenamiento comienzan alrededor de 3300, descienden rápidamente hasta ubicarse apenas por encima de 1000 después de la quinta época, y se estabilizan en torno a ese valor hasta el final del entrenamiento en la época 25. Para el conjunto de validación, se evidencia una tendencia similar, aunque con valores más elevados: la pérdida inicia cerca de 5200 y disminuye hasta aproximadamente 2300, lo que representa una reducción proporcionalmente mayor que en el conjunto de entrenamiento. No obstante, a pesar de esta mejora, el conjunto de validación presentó cierta inestabilidad entre las épocas 16 y 22, reflejada en fluctuaciones en los valores de pérdida.

El comportamiento de las curvas de pérdida para ambos conjuntos a lo largo de las épocas de entrenamiento indica que el algoritmo logró converger hacia un mínimo de la función de pérdida. Además, el hecho de que la curva de pérdida del conjunto de validación siga una tendencia similar refuerza la idea de que el modelo fue capaz de generalizar adecuadamente y mejorar su capacidad predictiva con el transcurso de las épocas. En este sentido, el proceso de entrenamiento para la predicción de la primera imagen, utilizando una secuencia como entrada, puede

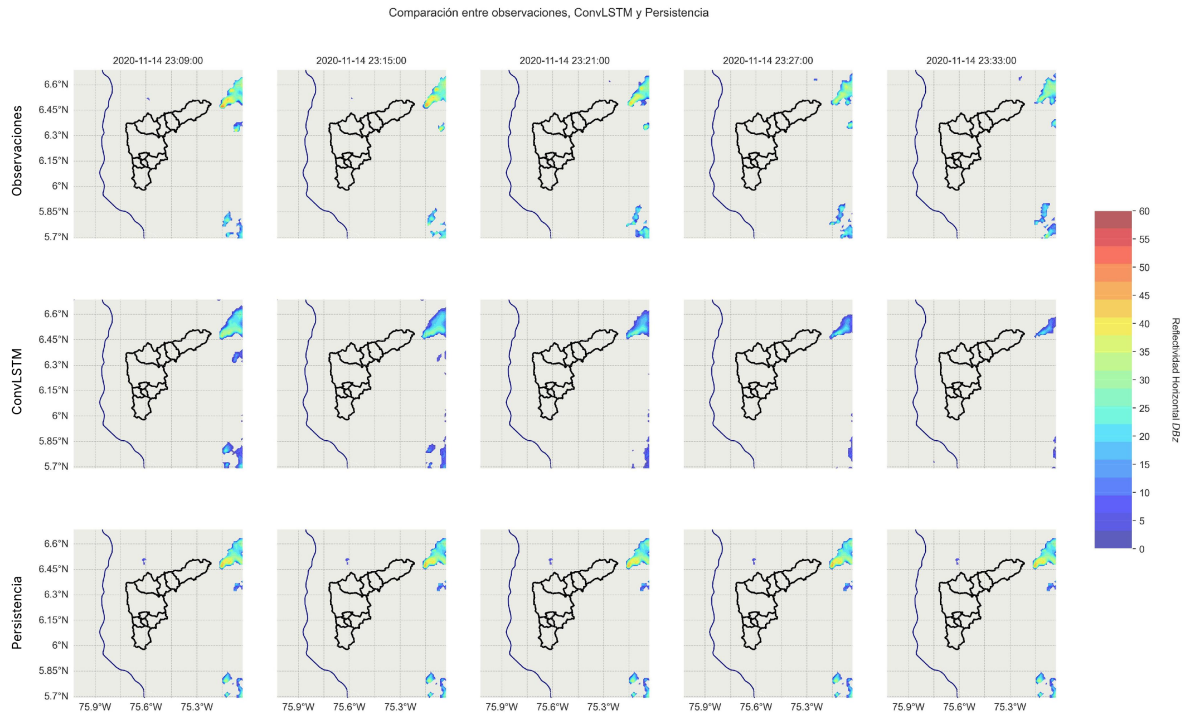


Figura 23: Mapas de reflectividad para los pasos de tiempo ($t + 5$) de la mediana predicción en términos de *CSI*, realizada por el modelo de ConvLSTM, se pueden verificar sus valores en la tabla 3

considerarse satisfactorio. Sin embargo, se enfatiza en la necesidad de realizar el proceso de entrenamiento para los tiempos posteriores a $t + 1$, debido a que el algoritmo estaría optimizado para realizar las predicciones en estos pasos de tiempo.

5.4 Otros Experimentos

Se realizaron diversos experimentos con el objetivo de encontrar un modelo que se desempeñara adecuadamente para el periodo de tiempo utilizado en su entrenamiento. En estos experimentos se variaron distintos hiperparámetros del proceso de entrenamiento, como la tasa de aprendizaje, la cantidad de épocas, la inclusión o exclusión del conjunto de validación, así como el periodo considerado para el entrenamiento y la cantidad de imágenes empleadas. Además, se evaluó el impacto de los recursos computacionales disponibles, llevando a cabo entrenamientos tanto con *CPU* como con *GPU*.

Para los demás experimentos presentados en esta sección, la mayoría fueron realizados con una tasa de aprendizaje de 0.001, un orden de magnitud mayor que la utilizada en los análisis anteriores (0.0001). Este ajuste se implementó con el objetivo de que el algoritmo realizara actualizaciones más amplias en los pesos durante el entrenamiento, permitiéndole así explorar mejor el espacio de soluciones y facilitar la convergencia hacia un mínimo global de la función de pérdida (Géron, 2019); es decir, el valor más bajo posible que esta función puede alcanzar para ese conjunto de datos.

Para el mismo conjunto de datos (01-11-2020 - 17-11-2020) utilizado en los análisis anteriores (Sección 5.2), se realizó

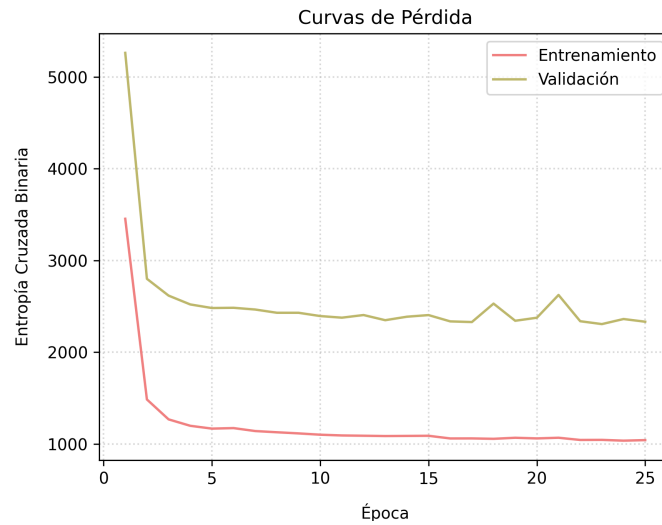


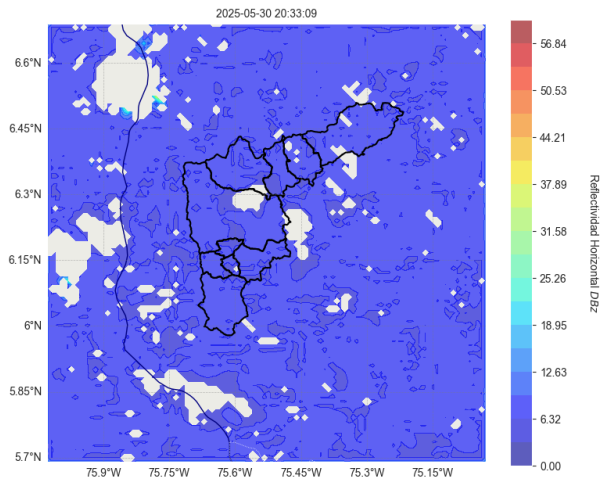
Figura 24: Evolución de la función de pérdida a lo largo del entrenamiento y del conjunto de validación a lo largo de las épocas

un nuevo entrenamiento del modelo, esta vez con una duración de 50 épocas, con el objetivo de brindarle el doble de iteraciones para ajustar sus pesos. En este experimento no se empleó un conjunto de validación, por lo que la única curva disponible corresponde a la evolución de la función de pérdida en el conjunto de entrenamiento. Esta curva muestra una tendencia a estabilizarse alrededor de un valor de 2000; sin embargo, a partir de la época 43, la pérdida se incrementa abruptamente hasta alcanzar un valor cercano a 12000, descendiendo posteriormente, pero sin retornar a niveles cercanos a los observados previamente (*Figura 25b*). Este comportamiento refleja un fenómeno observado en distintos entrenamientos, donde el modelo pierde su capacidad de generalización y produce predicciones carentes de sentido, indicando que no logró aprender un patrón significativo (*Figura 25a*).

Se realizó un experimento con datos comprendidos entre el 01 de junio y el 30 de agosto de 2020, con el objetivo de explorar el desempeño del algoritmo en un conjunto correspondiente a tres meses, los cuales representan una de las temporadas de menos lluvias del Valle de Aburrá. Particularmente, el periodo JJA (junio, julio y agosto) se caracteriza por la presencia de nubes convectivas dentro de las estructuras atmosféricas (Bedoya Soto et al., 2019). En este experimento, a pesar de que el modelo fue entrenado con una cantidad de datos aproximadamente seis veces mayor que la utilizada en los análisis presentados en la sección anterior, el pronóstico resultante no mostraba detección alguna de precipitación, indicando una falla total en la capacidad predictiva del modelo bajo estas condiciones como se muestra en la figura 26a; obtenida, al igual que la figura 25, mediante el registro del valor de la función de pérdida para cada época, y la predicción de una secuencia de tiempo.

Por otro lado, la curva de pérdida para la validación y el entrenamiento (*Figura 26b*) presenta un comportamiento distinto al observado en el experimento presentado en la sección de resultados de este trabajo. En este caso, la curva de validación se mantiene por debajo de la de entrenamiento, con valores cercanos a los 1000 durante la mayor parte del entrenamiento. En contraste, la curva de entrenamiento, aunque muestra una disminución inicial antes de la época 5, no logra estabilizarse por debajo de los 1500, alcanzando incluso valores superiores a 3600 al final del proceso. Esta divergencia entre ambas curvas sugiere que el modelo no logró una convergencia adecuada. El aparente buen desempeño en la validación podría explicarse por la presencia mayoritaria de secuencias sin precipitación en ese conjunto, lo cual haría que la pérdida se mantuviera baja y poco sensible a los errores del modelo.

Se realizaron varios entrenamientos utilizando los mismos hiperparámetros, datos y demás condiciones, variando

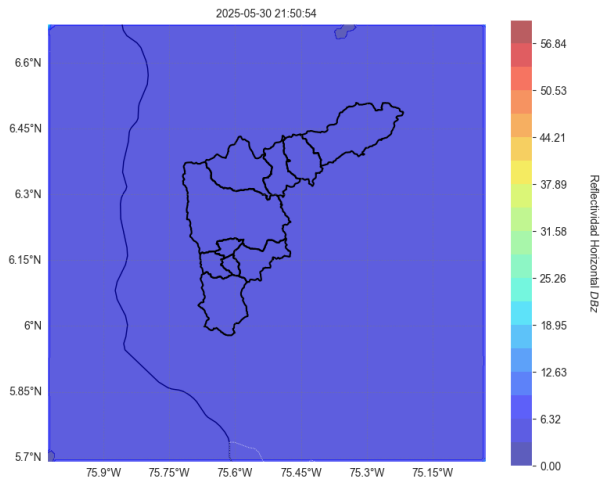


(a) Predicción del algoritmo en una secuencia aleatoria del conjunto de prueba

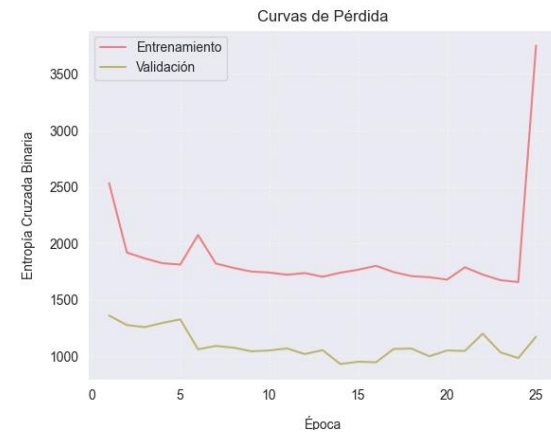


(b) Curva de evolución de la función de pérdida para el conjunto de entrenamiento.

Figura 25: Experimento de entrenamiento con 50 épocas, utilizando *CPU*, en el conjunto de datos en el que los resultados se basaron (01-11-2020 - 17-11-2020) sin utilizar conjunto de validación.



(a) Predicción del algoritmo en una secuencia aleatoria del conjunto de prueba



(b) Curva de evolución de la función de pérdida para el conjunto de entrenamiento.

Figura 26: Experimento de entrenamiento con 25 épocas, utilizando *CPU* para JJA.

únicamente el recurso de cómputo empleado (*CPU* o *GPU*), donde el resultado es similar a la figuras 25 y 26; indicando así resultados significativamente distintos a los planteados en la sección 5.2, y el entrenamiento no logró estabilizarse ni fue posible asegurar su replicabilidad. En general, en la aplicación de técnicas de aprendizaje profundo, lograr la reproducibilidad es un desafío debido, por un lado, a la aleatoriedad inherente a los algoritmos, que suelen inicializar sus parámetros con valores aleatorios; (B. Chen et al., 2022; PyTorch Contributors, 2025) y, por otro, a la

naturaleza no determinista de componentes computacionales como las *GPU* y *CPU*, lo cual se atribuye principalmente a la forma en que se realizan las operaciones con números en punto flotante (B. Chen et al., 2022).

6 Conclusiones

En este trabajo se aplicaron técnicas de aprendizaje profundo al problema del nowcasting de precipitación, utilizando específicamente la arquitectura ConvLSTM para realizar pronósticos con un horizonte temporal de 30 minutos. El conjunto de datos fue sometido a transformaciones significativas tanto en el dominio espacial como temporal, con el objetivo de ajustarlo a los requerimientos del modelo y facilitar su entrenamiento. Además, el desempeño del modelo se evaluó mediante métricas cuantitativas y se comparó frente a un modelo de persistencia, lo que permitió analizar sus capacidades predictivas en distintos escenarios.

Para el primer paso de tiempo en la predicción (6 minutos), el modelo ConvLSTM presenta un mejor desempeño que la persistencia en la métrica de probabilidad de detección. Sin embargo, en los pasos de tiempo posteriores, el modelo de persistencia supera a ConvLSTM en todas las métricas, salvo en la tasa de falsas alarmas. Esto permite concluir que, en el corto plazo, ConvLSTM logra capturar adecuadamente las señales relevantes gracias a que fue entrenado específicamente para ese horizonte. No obstante, a medida que el tiempo avanza, la predicción pierde precisión. Este deterioro evidencia la necesidad de adaptar el entrenamiento para optimizar múltiples pasos de tiempo, permitiendo que el modelo mantenga su capacidad predictiva en horizontes más largos.

En cuanto a la predicción del primer paso de tiempo, los ejemplos analizados en la sección de resultados evidencian que el modelo ConvLSTM presenta un mejor desempeño en las métricas de evaluación binarias cuando se enfrenta a sistemas de precipitación que cubren una porción considerable del área de estudio, en comparación con su rendimiento frente a estructuras más pequeñas y aisladas. No obstante, en estos últimos casos (caracterizados por sistemas de menor extensión) el modelo mostró un comportamiento más favorable en términos del *RMSE*. Lo anterior sugiere que, si bien el modelo resulta más adecuado para temporadas de mayor recurrencia de precipitación en el Valle del Aburrá, donde predominan sistemas más extensos, también ofrece una estimación razonable de los valores de reflectividad en eventos más localizados y aislados.

Debido a la naturaleza de la arquitectura utilizada, así como a las características de los datos y la configuración de los hiperparámetros, se puede inferir que el modelo es sensible a la estacionalidad de los eventos de precipitación. Esto se debe a que, al entrenarse para recibir 29 imágenes de entrada y pronosticar la número 30, el modelo asume cierta coherencia temporal en el comportamiento de los sistemas de precipitación dentro de cada secuencia. No obstante, desde una perspectiva operacional, la implementación del modelo resulta factible, ya que este puede ser reentrenado periódicamente o ajustado dinámicamente en función de su desempeño en tiempo real, sin requerir recursos computacionales excesivos.

Para finalizar, este estudio también evidencia que una mayor cantidad de datos no garantiza un mejor desempeño del modelo, como se observó en el experimento realizado con información del periodo junio–agosto (JJA), donde, a pesar del volumen considerable de datos, el modelo no alcanzó resultados satisfactorios. Además, se identificó una marcada sensibilidad del modelo respecto a la reproducibilidad del entrenamiento: variaciones en los recursos de cómputo utilizados (como el uso de CPU o GPU), así como en la configuración de hiperparámetros y otras condiciones del entorno, impidieron en varios casos la estabilización del proceso de entrenamiento.

Referencias

- Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., & Hickey, J. (2019). Machine Learning for Precipitation Nowcasting from Radar Images.
- Arias, P. A., Garreaud, R., Poveda, G., Espinoza, J. C., Molina-Carpio, J., Masiokas, M., Viale, M., Scaff, L., & van Oevelen, P. J. (2021). Hydroclimate of the Andes Part II: Hydroclimate Variability and Sub-Continental Patterns. *Frontiers in Earth Science, Volume 8 - 2020*. <https://doi.org/10.3389/feart.2020.505467>
- Atencia, A., Rigo, T., Sairouni, A., Moré, J., Bech, J., Vilaclara, E., Cunillera, J., Llasat, M. C., & Garrote, L. (2010). Improving QPF by blending techniques at the Meteorological Service of Catalonia. *Natural Hazards and Earth System Sciences, (7)*, 1443-1455. <https://doi.org/10.5194/nhess-10-1443-2010>
- Ayzel, G., Heistermann, M., Sorokin, A., Nikitin, O., & Lukyanova, O. (2019). All convolutional neural networks for radar-based precipitation nowcasting. *Procedia Computer Science, 150*, 186-192. <https://doi.org/10.1016/j.procs.2019.02.036>
- Bedoya Soto, J. M., Aristizábal, E., Carmona, A. M., & Poveda, G. (2019). Seasonal shift of the diurnal cycle of rainfall over medellin's valley, central Andes of Colombia (1998–2005). *Front. Earth Sci., 7*.
- Canchala, T., Alfonso-Morales, W., Carvajal-Escobar, Y., L., W. C., & Caicedo-Bravo, E. (2020). Monthly Rainfall Anomalies Forecasting for Southwestern Colombia Using Artificial Neural Networks Approaches. *Water, 12(9)*. <https://doi.org/10.3390/w12092628>
- Chen, B., Wen, M., Shi, Y., Lin, D., Rajbahadur, G. K., & Jiang, Z. M. (2022). Towards training reproducible deep learning models.
- Chen, L., Cao, Y., Ma, L., & Zhang, J. (2020). A Deep Learning-Based Methodology for Precipitation Nowcasting With Radar [e2019EA000812 10.1029/2019EA000812]. *Earth and Space Science, 7(2)*, e2019EA000812. <https://doi.org/https://doi.org/10.1029/2019EA000812>
- Coiffier, J. (2011). Half a century of numerical weather prediction. En *Fundamentals of Numerical Weather Prediction* (pp. 1-14). Cambridge University Press.
- CPC NOAA Team, I. (2025). *Cold & Warm Episodes by Season*. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php (accessed: 19.04.2025).
- Demetrakopoulos, P. (2023). Short-term Precipitation Forecasting in The Netherlands: An Application of Convolutional LSTM neural networks to weather radar data. <https://arxiv.org/abs/2312.01197>
- Doviak, R. J., & Zrnic, D. S. (1972, enero). *Doppler Radar and Weather Observations: Second Edition*.
- Düben, P., Modigliani, U., Geer, A., Siemen, S., Pappenberger, F., Bauer, P., Brown, A., Palkovic, M., Raoult, B., Wedi, N., & Baousis, V. (2021). Machine learning at ECMWF: A roadmap for the next 10 years. <https://doi.org/10.21957/ge7ckgm>
- Fang, W., Shen, L., S. Sheng, V., & Xue, Q. (2022). A novel method for precipitation nowcasting based on ST-LSTM. *Comput. Mater. Contin., 72(3)*, 4867-4877.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition* (2nd). O'Reilly Media.
- Houze Jr, R. A. (2014). *Cloud Dynamics, Volume 104, Second Edition* (2.^a ed.). Academic Press.
- IDEAM. (2025). *IMÁGENES DE RADAR*. <http://www.pronosticosyalertas.gov.co/archivos-radar> (accedido: 14.05.2025).
- Jianzhu, L., Shi, Y., Zhang, T., Li, Z., Wang, C., & Liu, J. (2024). Radar precipitation nowcasting based on ConvLSTM model in a small watershed in north China. *Natural Hazards, 120*, 63-85. <https://doi.org/10.1007/s11069-023-06193-6>
- Ladino Rincón, A. (2018). Caracterización de la microfísica de la precipitación mediante información de disdrómetros y radar polarimétrico para la estimación cuantitativa de lluvia en el área metropolitana del Valle de Aburrá. <https://repositorio.unal.edu.co/handle/unal/63867>

- Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., Clare, M. C. A., Lessig, C., Maier-Gerber, M., Magnusson, L., Bouallègue, Z. B., Nemesio, A. P., Dueben, P. D., Brown, A., Pappenberger, F., & Rabier, F. (2024). AIFS – ECMWF’s data-driven forecasting system. <https://arxiv.org/abs/2406.01465>
- Mandapaka, P. V., Germann, U., Panziera, L., & Hering, A. (2012). Can Lagrangian Extrapolation of Radar Fields Be Used for Precipitation Nowcasting over Complex Alpine Orography? *Weather and Forecasting*, 27(1), 28-49. <https://doi.org/10.1175/WAF-D-11-00050.1>
- National Academies of Sciences, E., & Medicine. (2022). *Machine Learning and Artificial Intelligence to Advance Earth System Science: Opportunities and Challenges: Proceedings of a Workshop* (R. Silvern, Ed.). The National Academies Press. <https://doi.org/10.17226/26566>
- Niño Medina, J. S., Barón, M. J. S., & Suarez, J. A. R. (2024). Application of Deep Learning for the Analysis of the Spatiotemporal Prediction of Monthly Total Precipitation in the Boyacá Department, Colombia. *Hydrology*, 11(8). <https://doi.org/10.3390/hydrology11080127>
- Paaß, G. (s.f.). *Deep Learning: How do deep neural networks work?* <https://lamarr-institute.org/blog/deep-neural-networks/> (accedido: 02.05.2025).
- Paul Markowski, Y. R. (2010). Mesoscale Convective Systems. En *Mesoscale Meteorology in Midlatitudes* (pp. 245-272). John Wiley & Sons, Ltd. <https://doi.org/https://doi.org/10.1002/9780470682104.ch9>
- Pérez Velasco, A. (2020, septiembre). *Predicción de lluvia a muy corto plazo mediante el uso de técnicas de aprendizaje profundo* [Tesis de maestría]. <http://hdl.handle.net/10902/20796>
- Prudden, R., Adams, S., Kangin, D., Robinson, N., Ravuri, S., Mohamed, S., & Arribas, A. (2020). A review of radar-based nowcasting of precipitation and applicable machine learning techniques. <https://doi.org/https://doi.org/10.48550/arXiv.2005.04988>
- PyTorch Contributors. (2025). Reproducibility — PyTorch 2.7 documentation [Accedido el 30 de mayo de 2025].
- Rauber, R. M., & Nesbitt, S. W. (2018). *Radar Meteorology (1st ed.) A first course*. Wiley.
- Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., & Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672-677. <https://doi.org/10.1038/s41586-021-03854-z>
- Ritvanen, J., Harnist, B., Aldana, M., Mäkinen, T., & Pulkkinen, S. (2023). Advection-Free Convolutional Neural Network for Convective Rainfall Nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 1654-1667. <https://doi.org/10.1109/JSTARS.2023.3238016>
- Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd). Pearson Education.
- Sepulveda Berrío, J. (2016). Estimación cuantitativa de precipitación a partir de la información de Radar Meteorológico del Área Metropolitana del Valle de Aburrá. <https://repositorio.unal.edu.co/handle/unal/58046>
- Shi, C., Zhang, Z., Zhang, W., Zhang, C., & Xu, Q. (2022). Learning Multiscale Temporal-Spatial-Spectral Features via a Multipath Convolutional LSTM Neural Network for Change Detection With Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2022.3176642>
- Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model.
- Shi, X., Zhou, C., Wang, H., Yung, D.-Y., Wong, W.-k., & Woo, W.-c. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. En C. Cortes, N. Lawrence, D. Lee, M. Sugiyama & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf
- Shroder, J. (2013, marzo). *Treatise on Geomorphology*.
- SIATA. (2025). *Quiénes somos*. https://siata.gov.co/siata_nuevo/ (accedido: 10.04.2025).
- Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Hickey, J., Agrawal, S., & Kalchbrenner, N. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting. *Submission to journal*. <https://arxiv.org/abs/2003.12140>

- Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hoimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., Xu, M., & Pinto, J. (2014). Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. *Bulletin of the American Meteorological Society*, *95*(3), 409-426. <https://doi.org/https://doi.org/10.1175/BAMS-D-11-00263.1>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *arXiv preprint arXiv:1409.3215*.
- Tian, L., Li, X., Ye, Y., Xie, P., & Li, Y. (2020). A Generative Adversarial Gated Recurrent Unit Model for Precipitation Nowcasting. *IEEE Geoscience and Remote Sensing Letters*, *17*(4), 601-605. <https://doi.org/10.1109/LGRS.2019.2926776>
- Vaughan, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Chantry, M., Hosking, J. S., & Turner, R. E. (2024). Aardvark weather: end-to-end data-driven weather forecasting. <https://arxiv.org/abs/2404.00411>
- Velásquez, N. (2022). Assessment of Deep Convective Systems in the Colombian Andean Region. *Hydrology*, *9*(7). <https://doi.org/10.3390/hydrology9070119>
- Wang, Y., Coning, E., Harou, A., Jacobs, W., Joe, P., Nikitina, L., Roberts, R., Wang, J., Wilson, J., Atencia, A., Bica, B., Brown, B., Goodmann, S., Kann, A., Li, P. W., Monterio, I., Schmid, F., Seed, A., & Sun, J. (2017, noviembre). *Guidelines for Nowcasting Techniques*.
- Wilks, D. S. (2019, junio). *Statistical methods in the atmospheric sciences* (4.^a ed.). Elsevier Science Publishing.
- Xiong, T., Wang, W., He, J., Su, R., Wang, H., & Hu, J. (2024). Spatiotemporal Feature Fusion Transformer for Precipitation Nowcasting via Feature Crossing. *Remote Sensing*, *16*(14). <https://doi.org/10.3390/rs16142685>
- Yin, J., Meo, C., Roy, A., Cher, Z. B., Wang, Y., Imhoff, R., Uijlenhoet, R., & Dauwels, J. (2024). Precipitation Nowcasting Using Physics Informed Discriminator Generative Models. *32nd European Signal Processing Conference, EUSIPCO 2024 - Proceedings*, 1967-1971. <https://doi.org/10.23919/EUSIPCO63174.2024.10715141>