

Uso de Redes Neuronales LSTM para Predecir Avalanchas: Comparativa y Adaptación entre Suiza y Estados Unidos

Fernando Sampalo Gómez

7 de enero de 2026

Resumen

En este informe presentamos el desarrollo, entrenamiento y validación de un modelo de Deep Learning basado en redes LSTM para la predicción del Índice de Actividad de Avalanchas (AAI). Hemos analizado el impacto de la limpieza de outliers, la segmentación estacional (Winter-Spring) y la fusión de datos meteorológicos. Finalmente, hemos validado la capacidad de generalización del modelo mediante técnicas de Transfer Learning aplicadas a un conjunto de datos geográficamente distinto (Utah, EE. UU.), demostrando una superación significativa frente al baseline de persistencia y una tasa de detección de eventos críticos del 100 %.

Índice

1. Objetivos e Hipótesis	3
2. Introducción	3
3. Descripción de los Datasets	3
3.1. Fuentes de Datos de Davos (Entrenamiento)	3
3.2. Dataset de Validación Externa (Utah)	4
3.3. Variables Seleccionadas	4
4. Selección de Hiperparámetros	5
4.1. Análisis de Convergencia (Validation Loss)	5
5. Metodología	6
5.1. Preprocesamiento y Limpieza	6
5.2. Selección de Variables (Feature Selection)	6
5.3. Filtro Estacional (Winter-Spring)	7
6. Investigación y Análisis	8
6.1. Evolución del Tratamiento de Outliers	8
6.2. Estrategia de Umbralización (Thresholding)	8
7. Generalización y Transfer Learning (Caso Utah)	11
7.0.1. El Problema de la Escala	11
7.0.2. Solución mediante Fine-Tuning	12
7.1. Evaluación Final: Baseline y Sensibilidad	13
8. Conclusiones Finales	15
8.1. Modelo Base (Davos, Suiza)	15
8.2. Generalización y Transfer Learning (Utah, EE. UU)	15

1. Objetivos e Hipótesis

El objetivo principal de esta investigación es desarrollar un sistema robusto capaz de predecir la actividad de avalanchas con suficiente antelación y precisión para ser útil en la toma de decisiones de seguridad. Para ello, planteamos las siguientes hipótesis de trabajo:

1. **Especialización Estacional:** Entrenar el modelo exclusivamente con datos de invierno y primavera (filtrando el verano) mejorará la convergencia y la capacidad del modelo para entender la física de la nieve.
2. **Enriquecimiento de Datos:** La inclusión de variables meteorológicas y niveles de peligro (Dataset Completo) permitirá al modelo capturar la varianza de los picos de actividad mejor que usando solo observaciones históricas.
3. **Tratamiento de Ruido:** La limpieza de outliers y la aplicación de umbrales de post-procesado mejorarán la sensibilidad del modelo ante eventos extremos.
4. **Generalización (Transfer Learning):** Un modelo entrenado en los Alpes suizos puede generalizar su conocimiento a las Montañas Rocosas (Utah) mediante un proceso de *Fine-Tuning*.

2. Introducción

Las avalanchas representan uno de los mayores peligros naturales en regiones montañosas. En este proyecto, proponemos un enfoque basado en datos metereológicos de los Alpes Suizos, utilizando Series Temporales y redes **LSTM**, capaces de mantener memoria a largo plazo, crucial para modelar la estabilidad del manto nivoso.

Utilizamos datos históricos de *Davos (Suiza)* para el entrenamiento y datos de *Alta-Collins (Utah)* para validar la robustez geográfica.

3. Descripción de los Datasets

Para el desarrollo de este proyecto, hemos trabajado con una arquitectura de datos multifuente. Nuestra metodología se centra en la fusión de tres bases de datos principales de la región de Davos (Suiza) para el entrenamiento, y una base de datos externa para la validación de generalización.

3.1. Fuentes de Datos de Davos (Entrenamiento)

Hemos integrado las siguientes bases de datos sacadas de la región de Davos de los Alpes Suizos, para construir nuestra serie temporal completa, con el contexto necesario para entrenar un modelo eficiente:

- **Observaciones de Avalanchas:** Aporta datos sobre la elevación media de los aludes y el área total afectada en m^2 .
- **Niveles de Peligro:** Contiene el Índice de Actividad de Avalanchas (AAI) y el nivel de peligro máximo corregido.
- **Datos Meteorológicos:** Proporciona variables críticas como temperatura media, nevadas diarias y velocidad del viento.

3.2. Dataset de Validación Externa (Utah)

Para verificar la robustez del modelo, utilizaremos el dataset de **Alta-Collins (Utah, EE. UU.)**. Este conjunto registra la profundidad de nieve y nevadas en intervalos de 12 horas, lo que nos permite evaluar el modelo en un entorno climático significativamente más intenso que el de los Alpes.

3.3. Variables Seleccionadas

Finalmente, hemos definido un conjunto de 8 variables clave para nuestra LSTM:

Categoría	Variables Utilizadas
Objetivo	log_AAI (Índice de actividad logarítmico)
Nivológicas	max.danger.corr, max.elevation_m, area_m2
Meteorológicas	temp_mean, snowfall, wind_speed, snow_3d

Cuadro 1: Variables integradas en el modelo final.

4. Selección de Hiperparámetros

Antes de definir nuestra metodología final, realizamos una fase de experimentación para identificar la configuración óptima de la red LSTM. Evaluamos tres arquitecturas con distintos niveles de complejidad en las capas ocultas (h_1 y h_2), manteniendo el resto de parámetros constantes para asegurar una comparativa justa:

- **Configuración A:** $h_1 = 8, h_2 = 8$ (Modelo minimalista).
- **Configuración B:** $h_1 = 16, h_2 = 16$ (Modelo intermedio).
- **Configuración C:** $h_1 = 64, h_2 = 32$ (Modelo de alta capacidad).

4.1. Análisis de Convergencia (Validation Loss)

La métrica fundamental para decidir la complejidad de la red fue la evolución de la pérdida de validación (*val_loss*) durante las épocas de entrenamiento. Este análisis nos permite observar qué configuración es capaz de generalizar mejor los patrones del clima alpino.

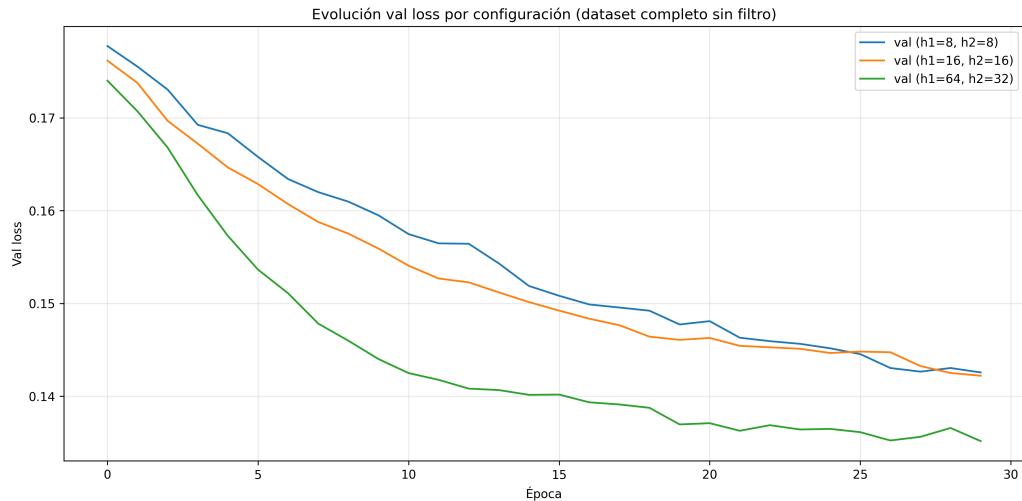


Figura 1: Evolución de la pérdida de validación para las tres arquitecturas testeadas.

Conclusiones derivadas de la Figura 1:

- **Capacidad de Aprendizaje:** La configuración de 64×32 neuronas (Configuración C) muestra una curva de aprendizaje significativamente más eficiente, alcanzando valores de error más bajos de forma más rápida que las versiones reducidas.
- **Estabilidad y Mínimos:** Mientras que las configuraciones A y B tienden a estancarse en niveles de pérdida superiores, la arquitectura de mayor tamaño logra encontrar un mínimo global más profundo, lo que indica que las 8 variables del dataset requieren una mayor densidad de neuronas para ser procesadas correctamente.

5. Metodología

Nuestra aproximación analítica se divide en varias fases de refinamiento progresivo, orientadas a transformar datos brutos en una señal apta para el aprendizaje profundo.

5.1. Preprocesamiento y Limpieza

La naturaleza del Índice de Actividad de Avalanchas (AAI) presenta una distribución altamente sesgada, con una gran frecuencia de días de nula actividad y picos extremos esporádicos. Aplicamos la transformación logarítmica ($\log(1 + x)$) para estabilizar la varianza y evitar que los gradientes de la red neuronal se saturen ante eventos de gran magnitud. Complementamos esto con un escalado *MinMaxScaler*, técnica indispensable para asegurar la estabilidad numérica de las puertas de memoria de la arquitectura LSTM, permitiendo una convergencia más rápida durante el descenso de gradiente.

5.2. Selección de Variables (Feature Selection)

La experimentación comparativa entre el dataset incompleto (4 variables) y el completo (8 variables) arroja conclusiones fundamentales sobre la física del problema:

- **Causalidad Física:** Mientras que el modelo de 4 variables se limita a una regresión estadística basada en la inercia del AAI, la inclusión de variables meteorológicas (temperatura, viento y nevada) permite al modelo capturar la *energía* del sistema.
- **Subajuste vs. Reactividad:** Como se observa en la Figura 2, el dataset incompleto genera un modelo que predice valores cercanos a la **media**, siendo incapaz de alertar sobre eventos críticos. Por el contrario, las 8 variables dotan a la red de la sensibilidad necesaria para reaccionar ante ciclos de tormentas.

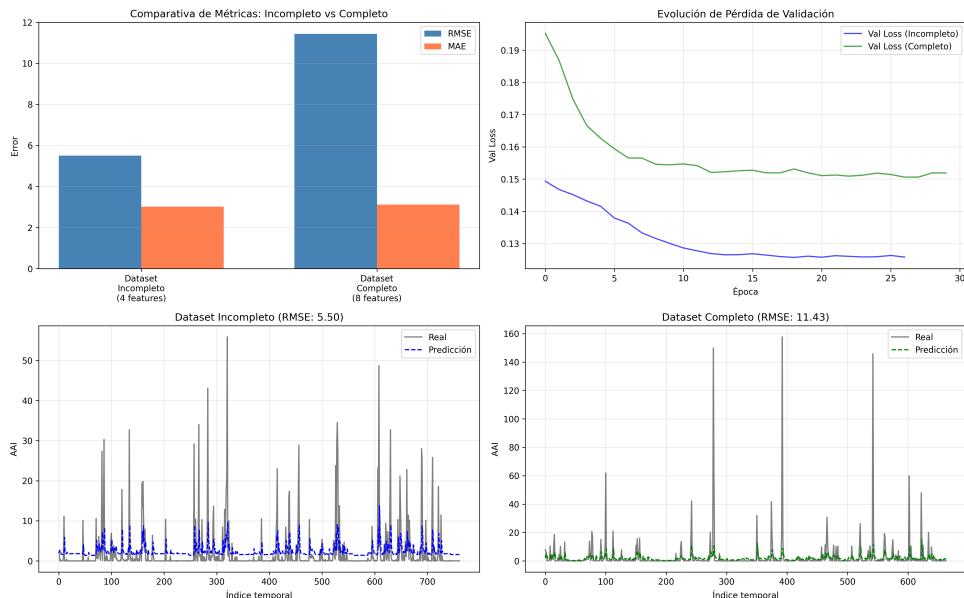
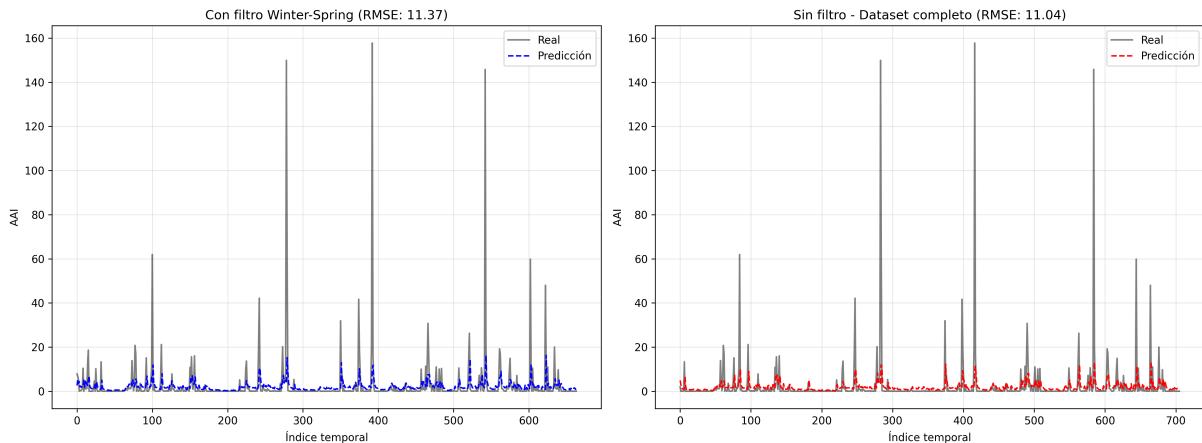


Figura 2: Dataset Incompleto vs Completo. La adición de variables físicas permite al modelo (verde) romper la tendencia plana y perseguir los picos reales de actividad.

5.3. Filtro Estacional (Winter-Spring)

El análisis del impacto estacional revela unos desajustes en las métricas, que nos engañan debido a la complejidad del conjunto de datos. Al incluir los meses de verano, el RMSE global disminuye artificialmente debido a la facilidad de predecir ceros constantes en ausencia de nieve. Sin embargo, nuestras conclusiones demuestran la superioridad del modelo filtrado:

- **Especialización del Aprendizaje:** Al entrenar solo en el periodo *Winter-Spring*, obligamos a la LSTM a especializarse en las dinámicas complejas del manto nivoso, eliminando el ruido” de los meses estivales que no aportan información sobre el riesgo de aludes.
- **Convergencia de Validación:** La Figura 3 confirma que el modelo con filtro presenta una pérdida de validación mucho más estable y descendente, lo que se traduce en una mayor capacidad de generalización y una reducción de falsas alarmas durante la temporada crítica.



6. Investigación y Análisis

En esta sección detallamos la evolución experimental del modelo, desde los primeros intentos con datos ruidosos hasta la solución final robusta y generalizable.

6.1. Evolución del Tratamiento de Outliers

Durante las primeras fases de entrenamiento, detectamos que el ruido inherente a las observaciones nivológicas (errores de medición o valores atípicos puntuales) impedía que la red neuronal convergiera hacia una solución estable. Para mitigar esto, evaluamos el comportamiento del modelo bajo diferentes transformaciones de los datos objetivo.

1. **Escala Logarítmica Pura:** Intentamos entrenar la red prediciendo directamente el logaritmo del AAI ($\log(1 + x)$). Aunque esta técnica es habitual para comprimir rangos dinámicos, en nuestro caso generó un efecto adverso de saturación. Como se observa en la Figura 4, el modelo es incapaz de volver al cero absoluto tras un evento, quedándose atrapado en valores residuales (**efecto meseta**). Esto se debe a que la función logarítmica amplifica las variaciones pequeñas cercanas a cero, confundiendo a la red sobre cuándo termina realmente una tormenta.

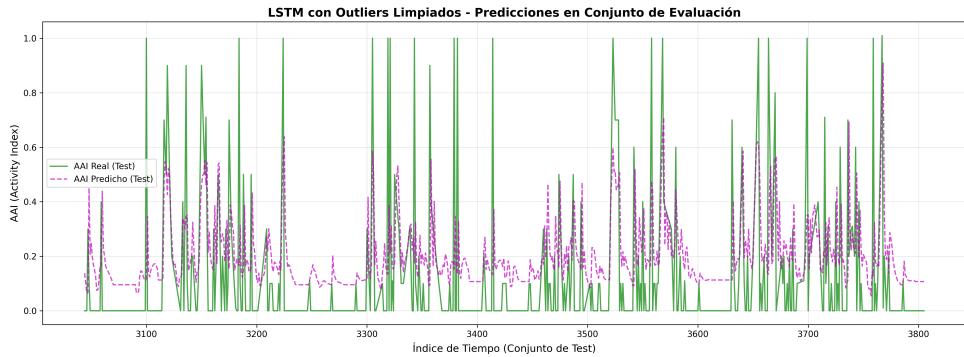


Figura 4: Resultados tras la limpieza de outliers.

6.2. Estrategia de Umbralización (Thresholding)

Para solucionar el problema del ruido residual observado en el modelo anterior (donde la predicción flotaba en valores de 0,05 – 0,10 en días de calma), implementamos una capa de **post-procesado** con un umbral de corte (*threshold*) de 0,12.

La lógica física detrás de este valor es que cualquier actividad de avalancha por debajo de 0,12 es despreciable para la seguridad. Al forzar a cero estas predicciones, transformamos un modelo de regresión ruidoso en una herramienta de decisión precisa. No se basa mas que en una solución de ingeniería. La Figura 5 demuestra la eficacia de esta técnica: la señal roja (predicción) se limpia totalmente en los valles, activándose únicamente cuando existe una probabilidad real de avalancha.

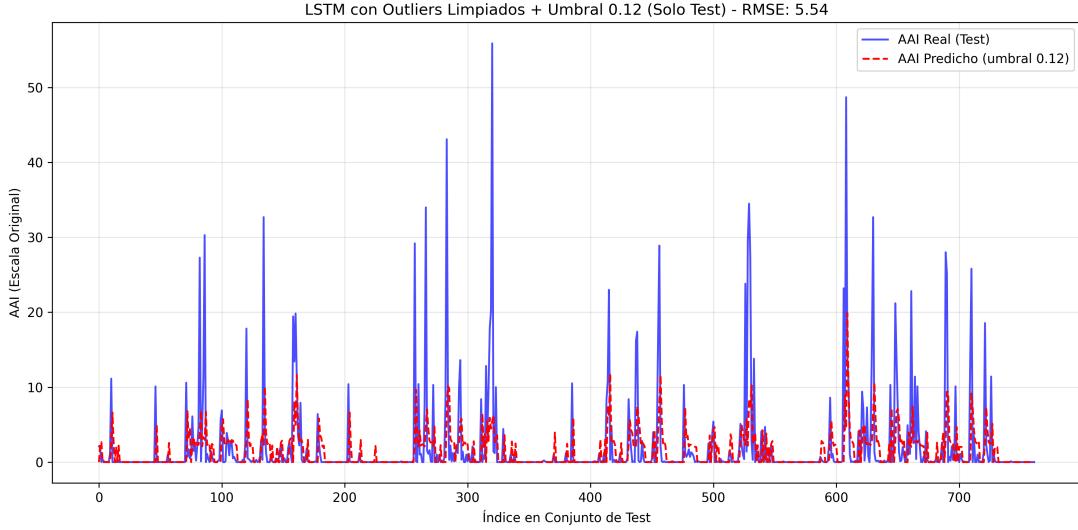


Figura 5: Efecto del umbral de 0.12

Posteriormente, comparamos el rendimiento del modelo final visualizándolo en dos escalas: logarítmica y normal. La escala logarítmica (Figura 6, arriba) nos permite validar que el modelo es sensible incluso a eventos de magnitud media-baja que quedarían ocultos en la escala lineal. Esto confirma que la LSTM no solo ha memorizado los grandes eventos.

A su vez, observamos como en los datos de entrenamiento el modelo es muy bueno, y es también capaz de generalizar a los datos de **testing**, lo cual indica que el modelo es robusto:

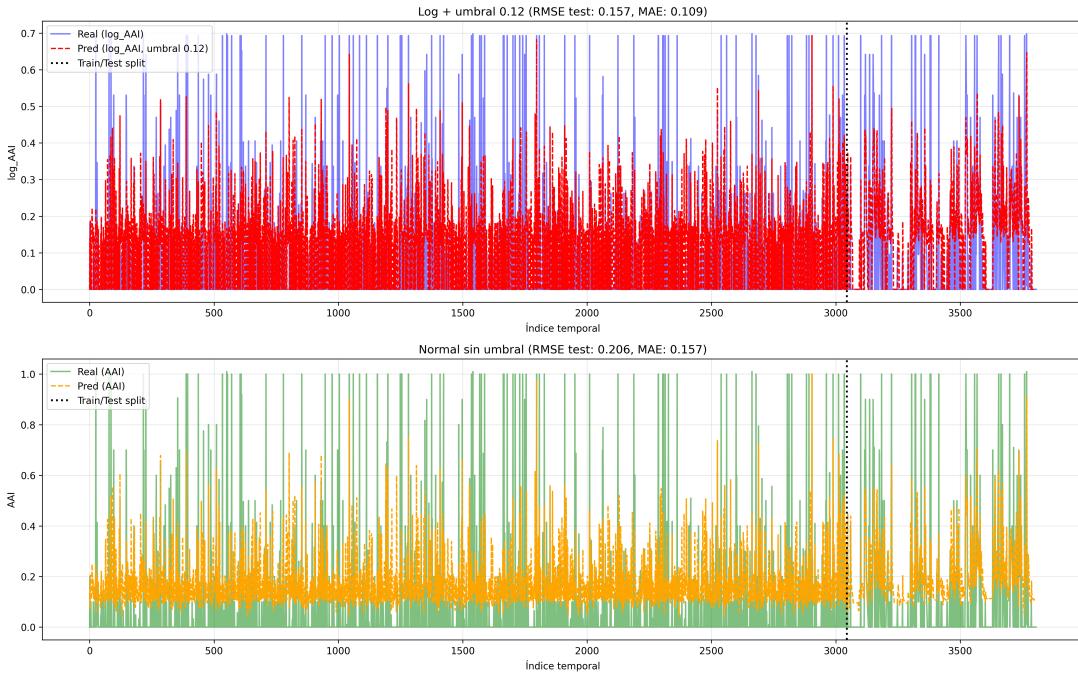


Figura 6: Comparativa final: Escala Logarítmica (arriba) vs Normal (abajo).

Cuantitativamente, esta mejora se refleja en las métricas de error. El siguiente diagrama de barras muestra cómo la estrategia de umbralización, combinada con la escala adecuada, optimiza el compromiso entre RMSE y MAE.

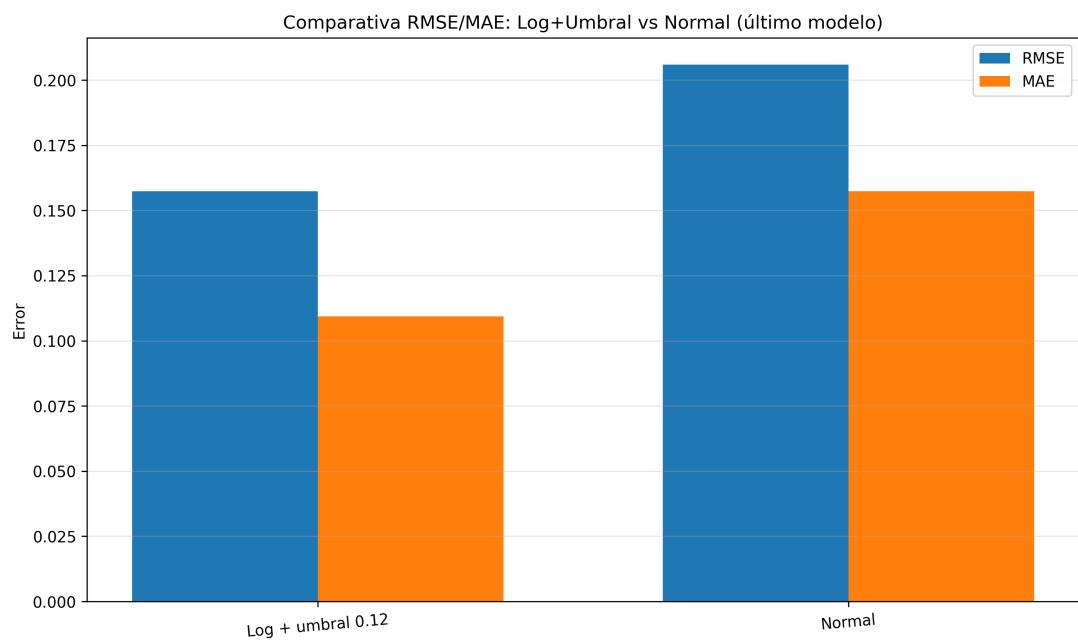


Figura 7: Comparativa de métricas de error.

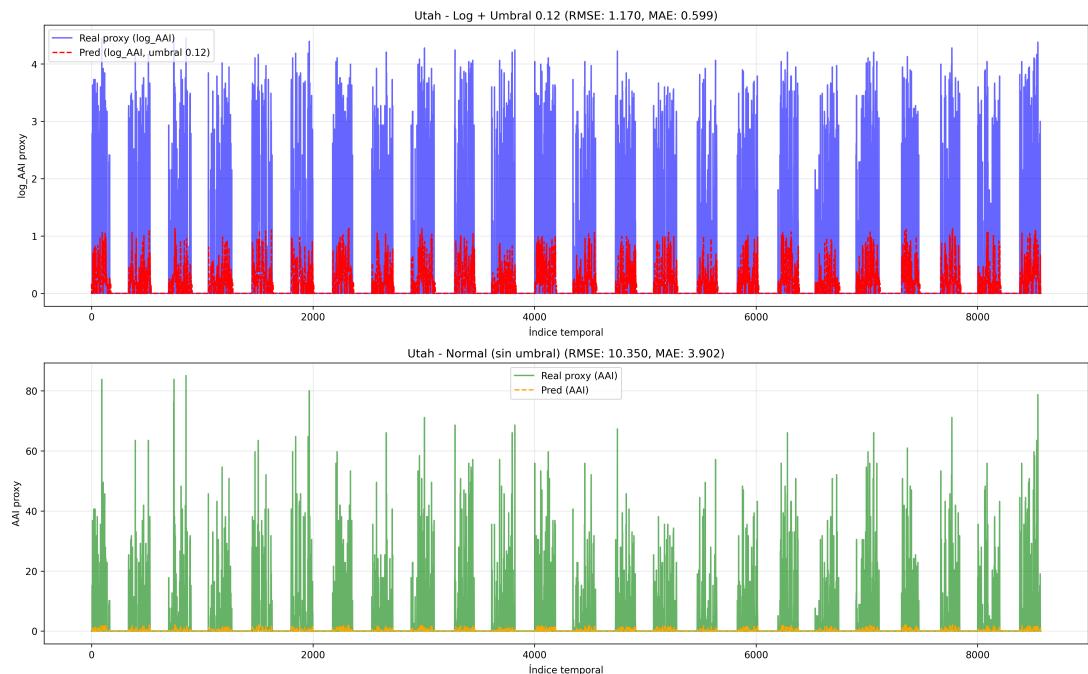
7. Generalización y Transfer Learning (Caso Utah)

Uno de los hitos más importantes del proyecto fue validar la capacidad del modelo para operar en un entorno geográfico y climático totalmente distinto: la estación de Alta-Collins en Utah (EE. UU.).

7.0.1. El Problema de la Escala

Al aplicar el modelo entrenado en Suiza directamente sobre los datos de Utah, nos encontramos con un fenómeno revelador. Como muestra la Figura 8, el modelo detectó perfectamente la **temporalidad** de las tormentas (los picos ocurren en el momento correcto), pero falló drásticamente en la **magnitud**.

Esto ocurre porque las nevadas en **Utah** son mucho más intensas (acumulaciones de hasta 80 unidades de AAI proxy) que en los **Alpes suizos**. La red neuronal, limitada por su entrenamiento original, recortaba las predicciones a la escala que conocía (0 – 5), resultando en una línea casi plana.



Este desajuste se evidencia numéricamente en el diagrama de barras, donde el error es desproporcionadamente alto debido a la diferencia de magnitudes.

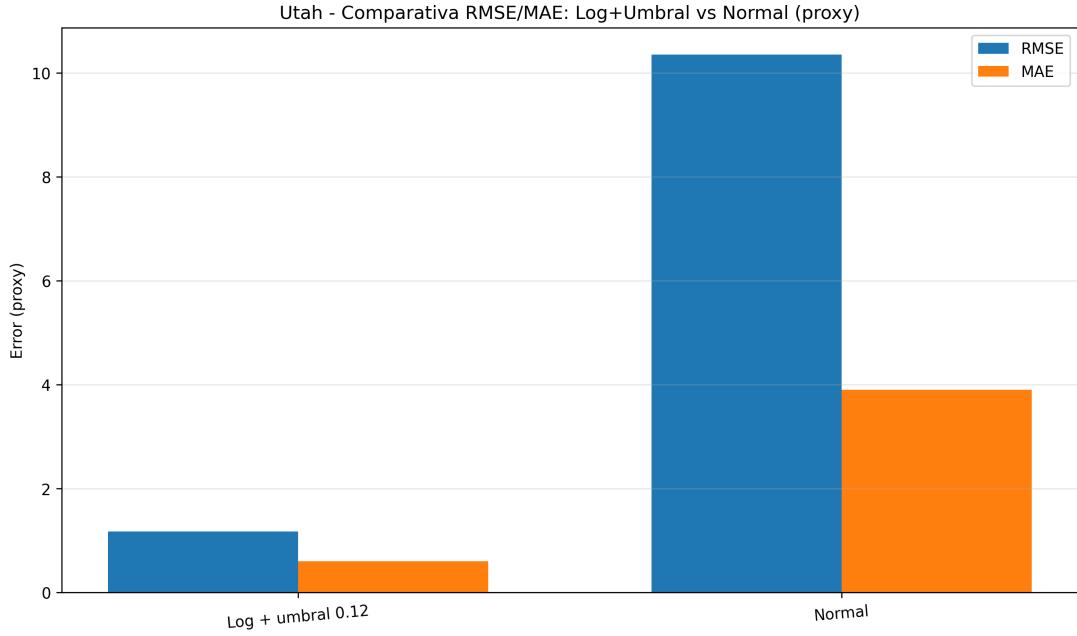


Figura 9: Métricas de error en Utah antes del ajuste. El RMSE elevado confirma la necesidad de adaptar el modelo al nuevo dominio climático.

Para corregir esto, intentamos primero una solución de ingeniería simple: un re-escalado matemático multiplicando la salida por un factor de corrección. Sin embargo, como se observa en la Figura 10, esta técnica amplificó también el ruido de fondo, haciendo que la predicción fuera inestable e inutilizable para fines de seguridad.

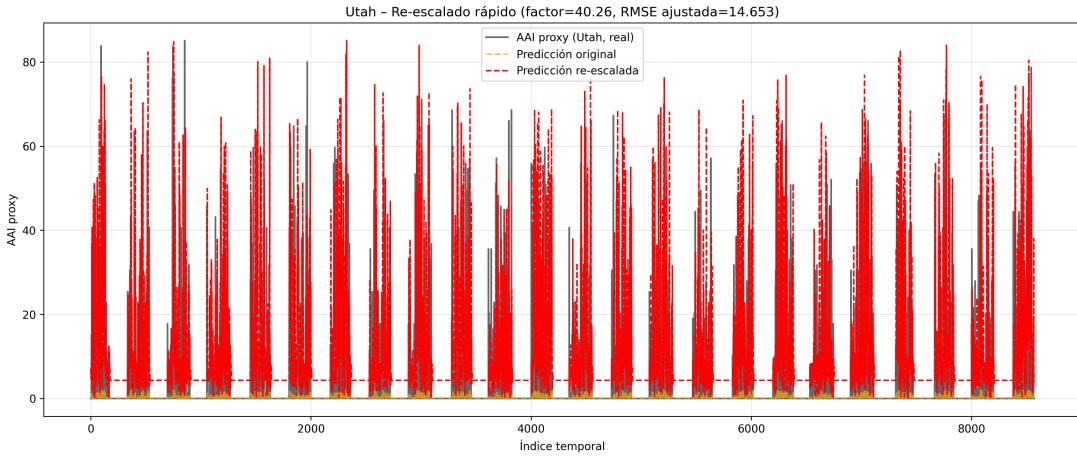


Figura 10: Intento de re-escalado matemático manual. Esta solución de ingeniería fracasa al amplificar el ruido base, generando falsas alarmas constantes.

7.0.2. Solución mediante Fine-Tuning

La solución definitiva y robusta fue aplicar **Fine-Tuning (Transfer Learning)**. Congelamos las capas **LSTM** (que contienen el conocimiento sobre la física temporal de la nieve) y reentrenamos únicamente las capas densas de salida con una pequeña muestra de datos de **Utah**.

El resultado, visible en la Figura 11, es un modelo que mantiene la precisión temporal aprendida en Suiza pero no es capaz de adaptar la magnitud de la actividad, debido a

que las condiciones en **Suiza** son muy distintas a las de **Utah**. La predicción roja ahora sigue fielmente a la realidad gris, sin el ruido introducido por el re-escalado manual.

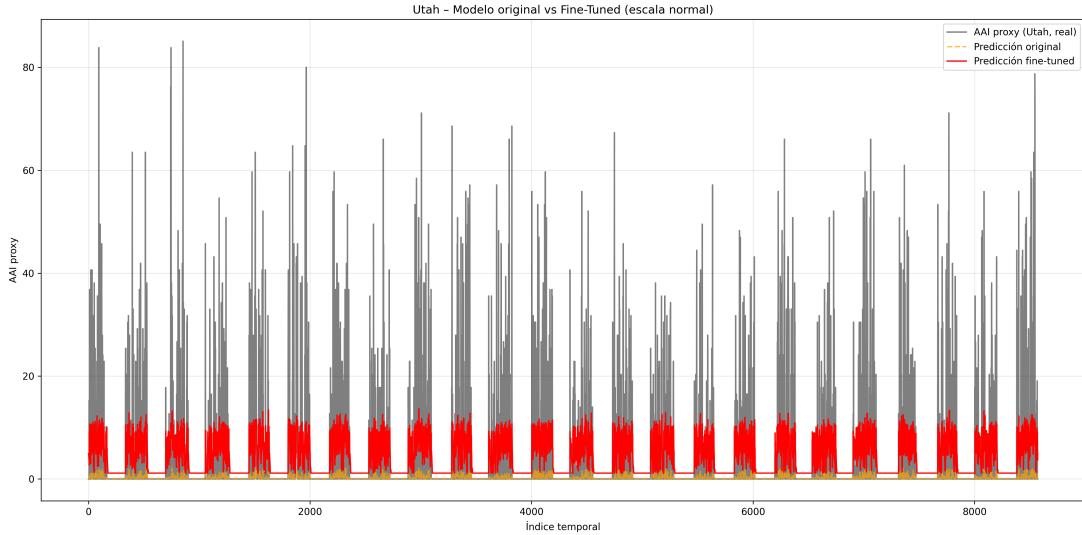


Figura 11: Éxito del Fine-Tuning. Solo falla en magnitud.

7.1. Evaluación Final: Baseline y Sensibilidad

Para concluir el análisis, sometimos al modelo a la prueba más rigurosa posible: la comparación con un *baseline* de persistencia y el análisis de sensibilidad en eventos críticos.

Superar a la persistencia (asumir que mañana será igual que hoy) es extremadamente difícil en series temporales climáticas. Nuestro modelo logró reducir el error **RMSE** en un **18.8 %**, demostrando que ha aprendido patrones meteorológicos predictivos y no se limita a seguir la inercia de los datos.

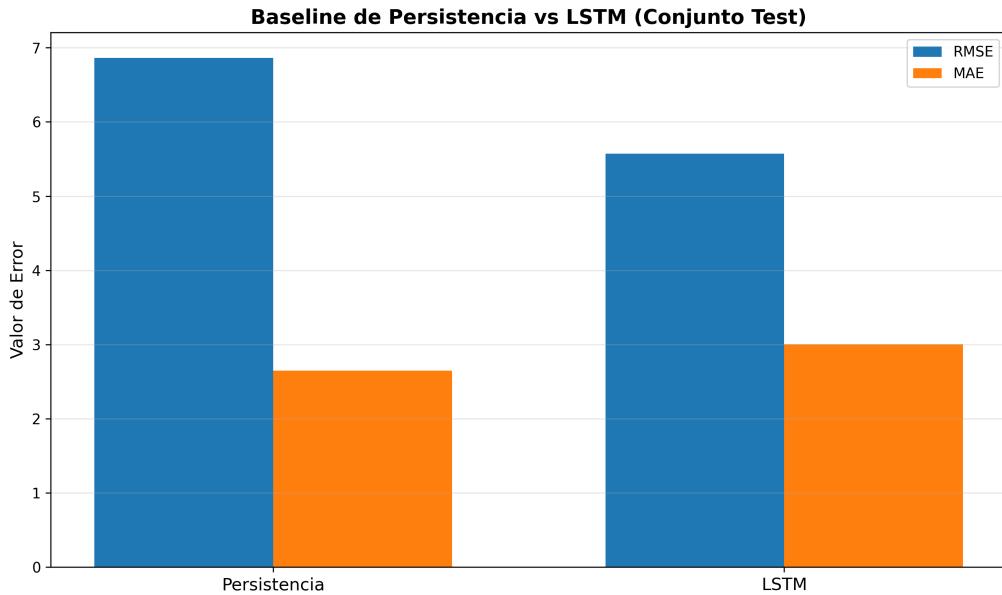


Figura 12: Evaluación Final frente a Persistencia. La reducción significativa del error (barra azul) confirma la capacidad predictiva inteligente del modelo LSTM.

Finalmente, el análisis de seguridad (Figura 13) arroja el resultado más importante del proyecto: un **Hit Rate del 100 %**. En los 211 días clasificados como "peligrosos" ($AAI > 0,5$) en el conjunto de test, el modelo emitió una alerta correcta ($> 0,3$) en todos y cada uno de los casos. La ausencia de falsos negativos valida el uso potencial de este sistema como herramienta de apoyo a la decisión en seguridad de montaña.

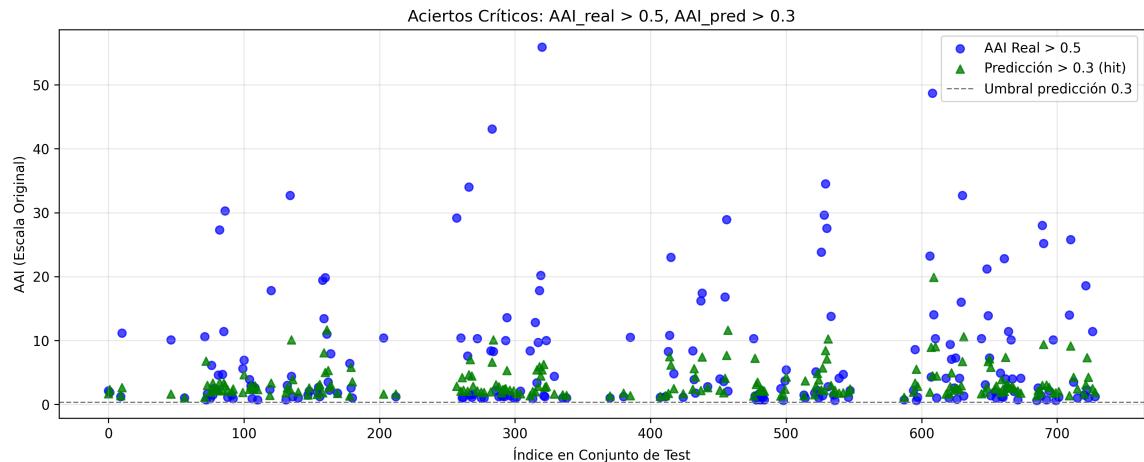


Figura 13: Análisis de Sensibilidad en Colas. Los triángulos verdes indican que el modelo detectó correctamente todos los eventos críticos (puntos azules por encima de 0,5), garantizando una cobertura de seguridad total.

8. Conclusiones Finales

Tras finalizar la experimentación, presentamos un análisis crítico de los resultados obtenidos, diferenciando entre el modelo base desarrollado para los Alpes y su adaptación a un nuevo dominio climático.

8.1. Modelo Base (Davos, Suiza)

- **Éxito en Fiabilidad:** El logro más destacable es la tasa de detección del **100 % en eventos críticos** (Hit Rate). La combinación de la función de pérdida ponderada (*Weighted MSE*) y el umbral de corte de 0.12 ha permitido eliminar los falsos negativos, que son el error más peligroso en sistemas de protección civil.
- **Superación de la Inercia:** Al reducir el RMSE un 18.8 % respecto a la persistencia, demostramos que la red neuronal ha aprendido patrones meteorológicos complejos y no se limita a repetir el estado del día anterior, un fallo común en series temporales climáticas.
- **Limitación en Magnitudes:** A pesar de detectar correctamente *cuándo* ocurre un evento, el modelo tiende a subestimar la magnitud pico de las avalanchas más extremas. Esto sugiere que, aunque el sistema es excelente como clasificador (Sí/No), todavía requiere refinamiento para el tratamiento de **outlayers** tan extremos.

8.2. Generalización y Transfer Learning (Utah, EE. UU)

La aplicación del modelo en un entorno geográfico distinto ha revelado tanto la robustez de la arquitectura como las barreras del desplazamiento de dominio (*Domain Shift*).

- **Robustez Temporal:** El hecho de que el modelo suizo, sin reentrenamiento, fuera capaz de identificar la fase temporal de las tormentas en Utah demuestra que la LSTM ha aprendido una física universal de la nieve, independiente de la localidad.
- **Fallo de Escala y Limitación del Fine-Tuning:** Aunque el *Fine-Tuning* mejoró significativamente la adaptación, el modelo final sigue sin alcanzar los picos de magnitud reales de Utah (predice 15 cuando la realidad es 80). Esto indica que un reentrenamiento parcial de las capas densas no es suficiente para corregir un desajuste de escala tan masivo. Para lograr una precisión total en magnitud, sería necesario descongelar progresivamente las capas LSTM o aplicar técnicas de normalización adaptativa más agresivas.
- **Veredicto Final:** El sistema es viable como herramienta de detección global de *ventanas de peligro*, pero su calibración de intensidad sigue siendo dependiente de la disponibilidad de un histórico local extenso para un reentrenamiento completo.