



Computación de Alto Rendimiento (Grado IA)

Diseño de un Sistema de IA para la Predicción de Riesgo de Inundaciones Urbanas en Japón

Nombre: Stanislav Gatin



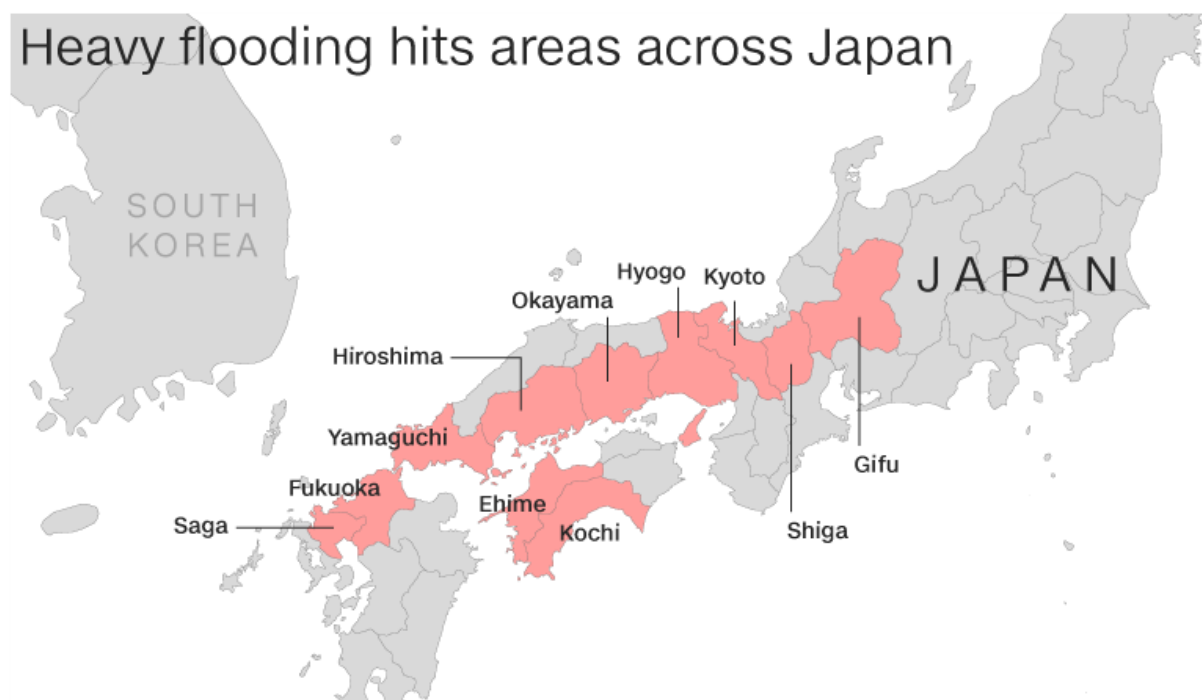
Índice

Bloque 1: Introducción y análisis del reto urbano.....	1
1.1 Contexto de la IA urbana y la HPC.....	1
1.2 Descripción del problema elegido.....	2
1.3 Primeras decisiones de diseño.....	4
1.4 Dilemas éticos iniciales.....	5
Bloque 2: Tipos de datos y decisiones sobre su obtención.....	6
2.1 Qué datos se necesitan y por qué.....	6
2.2 Cómo se obtienen.....	10
2.3 Problemas de privacidad y calidad.....	10
2.4 Alternativas y justificación.....	10
Bloque 3: Elección del modelo de IA.....	11
3.1 Comparativa entre modelos (LSTM, XGBoost, etc.).....	11
3.2 Justificación técnica y ética.....	11
3.3 Diseño del modelo propuesto.....	12
3.4 Opciones desechadas y por qué.....	12
Bloque 4: Infraestructura HPC y sostenibilidad del sistema.....	13
4.1 Cómputo requerido.....	13
4.2 Opciones sostenibles.....	13
4.3 Alineación con principios de eficiencia energética.....	14
4.4 Resistencia a desastres naturales.....	14
4.5 Comentarios sobre escalabilidad.....	14
Bloque 5: Estimaciones energéticas y comparación con alternativa ligera.....	15
5.1 Cálculo detallado de consumo y emisiones.....	15
5.2 Comparativa de modelos.....	16
5.3 Compromisos y decisiones realistas.....	16
Bloque 6: Reflexión ética y profesional.....	18
6.1 Equidad, explicabilidad y sostenibilidad.....	18
6.2 Decisiones clave justificadas.....	18
6.3 Lo que se ganaría/perdería cambiando decisiones.....	19
Bloque 7: Comentario final.....	19
Cómo se adaptaría este diseño a otros contextos.....	19
Recursos de interés.....	20

Bloque 1: Introducción y análisis del reto urbano

1.1 Contexto de la IA urbana y la HPC

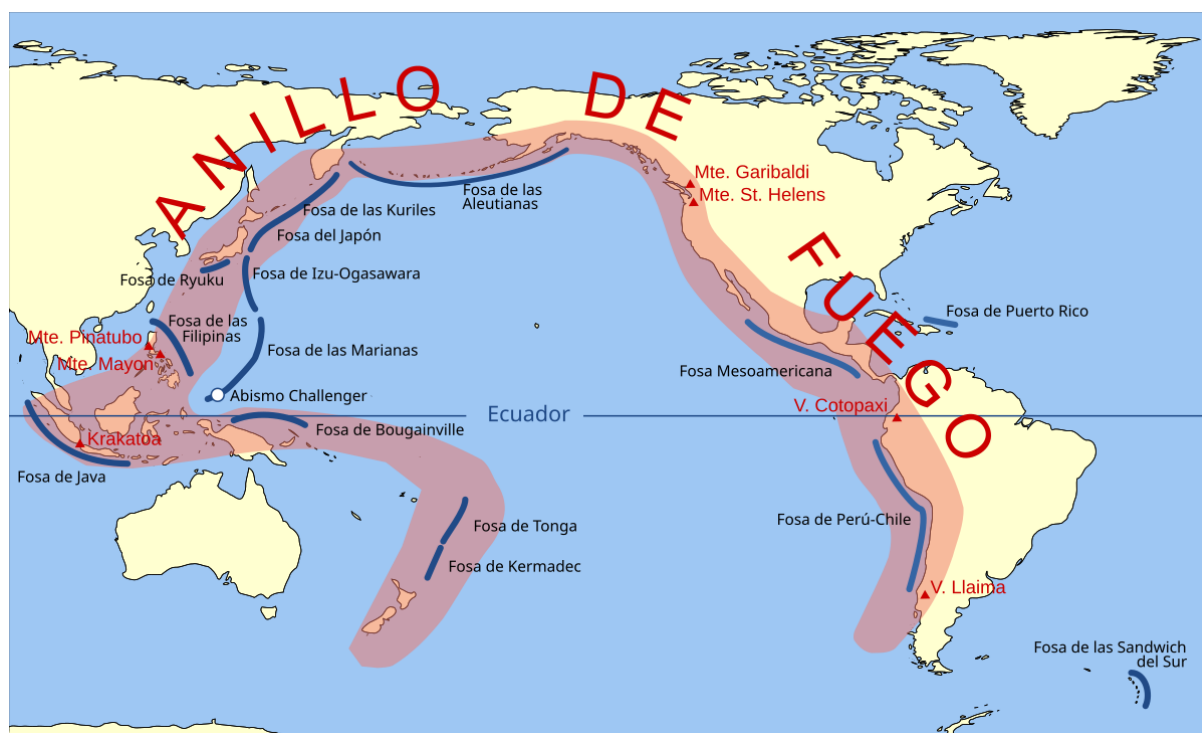
Los sistemas de inteligencia artificial (IA) urbana, respaldados por computación de alto rendimiento (HPC), están transformando la gestión de ciudades al procesar grandes volúmenes de datos en tiempo real para abordar problemas complejos como el tráfico, la energía o los desastres naturales. En Japón, las inundaciones urbanas, exacerbadas por tifones, lluvias y el cambio climático, representan un desafío crítico para ciudades como Tokio y Osaka, donde la alta densidad poblacional y la infraestructura costera aumentan los riesgos. Este proyecto propone un sistema de IA apoyado en HPC para predecir zonas urbanas propensas a inundaciones en tiempo real, integrando datos meteorológicos, topográficos y de sensores IoT.



Map: Areas in Japan Hit by Heavy Flooding July 10, 2018

1.2 Descripción del problema elegido

Japón enfrenta un riesgo excepcional de inundaciones urbanas debido a su ubicación geográfica, clima extremo y alta urbanización. Situado en el Cinturón de Fuego del Pacífico, el país experimenta tifones anuales, lluvias monzónicas intensas y, en ocasiones, tsunamis, que provocan inundaciones devastadoras en ciudades densamente pobladas. En 2024, el tifón Shanshan afectó gravemente regiones como Kanagawa, Tokio y Osaka, causando evacuaciones masivas, interrupciones en el suministro eléctrico y daños a infraestructuras valorados en miles de millones de yenes. Ciudades como Osaka, con sistemas de drenaje sobrecargados, y Fukuoka, vulnerable a desbordamientos de ríos, sufren impactos recurrentes que paralizan el transporte, destruyen propiedades y ponen en peligro vidas. Según la Agencia Meteorológica de Japón (JMA), las precipitaciones extremas han aumentado un 20% en las últimas dos décadas debido al cambio climático, lo que ha incrementado tanto la frecuencia como la severidad de las inundaciones urbanas.



Cinturón de Fuego del Pacífico

El problema específico que este sistema busca resolver es la incapacidad de predecir con precisión y en tiempo real las zonas urbanas con mayor riesgo de inundación durante eventos climáticos extremos. Los modelos meteorológicos actuales ofrecen pronósticos generales que carecen de resolución a nivel de barrios y no integran datos dinámicos, como los niveles de agua en alcantarillas, la capacidad de drenaje local o las características topográficas específicas. Esta limitación dificulta que las autoridades municipales emitan alertas focalizadas, prioricen recursos (como barreras contra inundaciones o equipos de rescate) y planifiquen evacuaciones efectivas. Por ejemplo, en el distrito de Koto en Tokio, ubicado bajo el nivel del mar, la falta de predicciones detalladas aumenta los riesgos para comunidades vulnerables, como personas mayores, familias de bajos ingresos y residentes con discapacidades. Además, las inundaciones generan efectos secundarios significativos, incluyendo cortes de energía, interrupciones en el transporte público, contaminación por desbordamiento de aguas residuales y pérdidas económicas que afectan desproporcionadamente a los sectores más desfavorecidos.



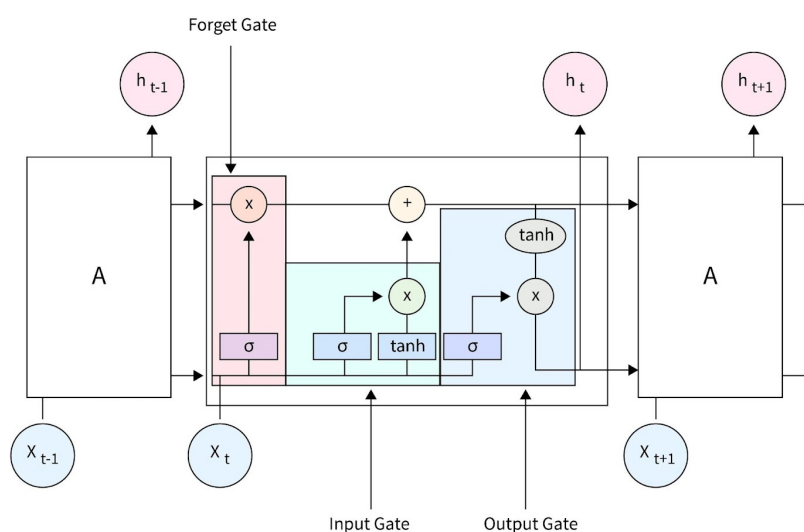
La escuela primaria Ohashi y sus alrededores quedaron inundados tras la inundación del río Kosegawa tras las lluvias torrenciales del 10 de julio de 2023 en Kurume, Fukuoka, Japón. The Asahi Shimbun/Getty Images

Este sistema de inteligencia artificial propone abordar estas deficiencias mediante la predicción de riesgos de inundación a nivel de vecindario, integrando datos meteorológicos en tiempo real de la JMA, mapas topográficos detallados y lecturas de sensores IoT instalados en ríos y sistemas de drenaje urbanos. El objetivo es generar un mapa de riesgo dinámico que identifique áreas críticas con una antelación de 6 a 12 horas, permitiendo a las autoridades implementar medidas proactivas, como el despliegue de barreras, el refuerzo de drenajes o la evacuación de zonas de alto riesgo. La solución está diseñada

para ser escalable, adaptándose a diferentes contextos urbanos en Japón, desde megalópolis como Tokio hasta ciudades medianas como Sendai o Hiroshima, y debe cumplir con principios de sostenibilidad, privacidad y equidad. Esto implica evitar el uso de datos personales, garantizar que las alertas sean accesibles para todos los grupos poblacionales y minimizar el impacto energético del sistema, especialmente en un país donde la resiliencia frente a desastres es una prioridad nacional.

1.3 Primeras decisiones de diseño

El diseño inicial del sistema se basa en el uso exclusivo de inteligencia artificial avanzada, concretamente mediante arquitecturas de redes neuronales profundas capaces de procesar datos multivariables en tiempo real. Se opta por modelos del tipo LSTM (Long Short-Term Memory) y ConvLSTM, especializados en secuencias temporales y análisis espacial-temporal, ideales para predecir patrones de inundación basados en datos meteorológicos, hidrológicos y topográficos.



LSTM (Long Short Term Memory)

Respecto a la entrada de datos, se establece una jerarquía basada en tres fuentes principales: (1) datos meteorológicos en tiempo real proporcionados por la Agencia Meteorológica de Japón (JMA), (2) sensores IoT locales instalados en puntos críticos (drenajes, puentes, márgenes de ríos), y (3) datos geoespaciales y topográficos de alta resolución. Para la integración y preprocesamiento de estas fuentes heterogéneas, se diseñará un pipeline basado en Spark y Kafka, capaz de gestionar flujos continuos y asegurar escalabilidad.

En cuanto a la visualización, el sistema generará mapas de riesgo accesibles a través de una interfaz web para autoridades municipales y aplicaciones móviles para la ciudadanía,



con alertas geolocalizadas y recomendaciones personalizadas. Todo el diseño contempla modularidad y escalabilidad para su implementación en distintas ciudades japonesas, adaptándose a sus particularidades geográficas y de infraestructura.

1.4 Dilemas éticos iniciales

Uno de los primeros dilemas éticos identificados es el uso de datos de sensores urbanos e información geolocalizada. Aunque el sistema no utilizará información personal directamente, la recopilación de datos en zonas residenciales plantea preocupaciones sobre privacidad y consentimiento. Se deberán establecer protocolos estrictos de anonimización y limitar la granularidad de los datos compartidos públicamente.

Otro dilema se relaciona con la equidad en el acceso a la información. Las alertas y recomendaciones deben ser comprensibles para toda la población, incluidas personas mayores, personas con discapacidad, y residentes no nativos del idioma japonés. Esto implica traducir mensajes a varios idiomas y ofrecer canales múltiples de difusión: notificaciones móviles, mensajes de voz, carteles digitales, etc.

Finalmente, el uso de recursos computacionales intensivos puede entrar en conflicto con principios de sostenibilidad. Por ello, se estudiará el impacto energético del sistema HPC y se evaluará el uso de centros de datos verdes o soluciones edge computing para reducir la huella de carbono. Además, se explorará una versión ligera del sistema, menos precisa pero más eficiente energéticamente, para su uso en localidades con recursos limitados.

Bloque 2: Tipos de datos y decisiones sobre su obtención

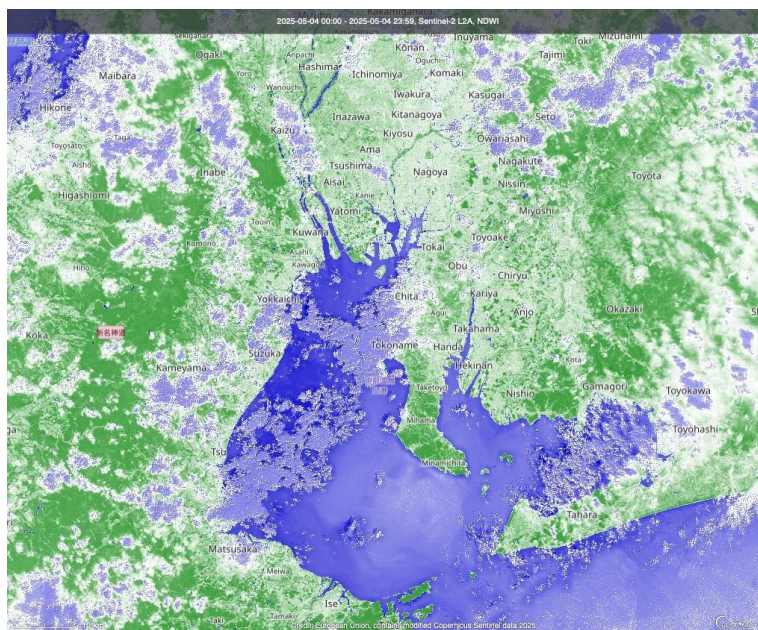
2.1 Qué datos se necesitan y por qué

Para predecir con precisión el riesgo de inundaciones urbanas en tiempo real, se requieren datos heterogéneos que reflejen tanto las condiciones meteorológicas como las características físicas del entorno urbano. Se identifican cuatro categorías principales de datos:

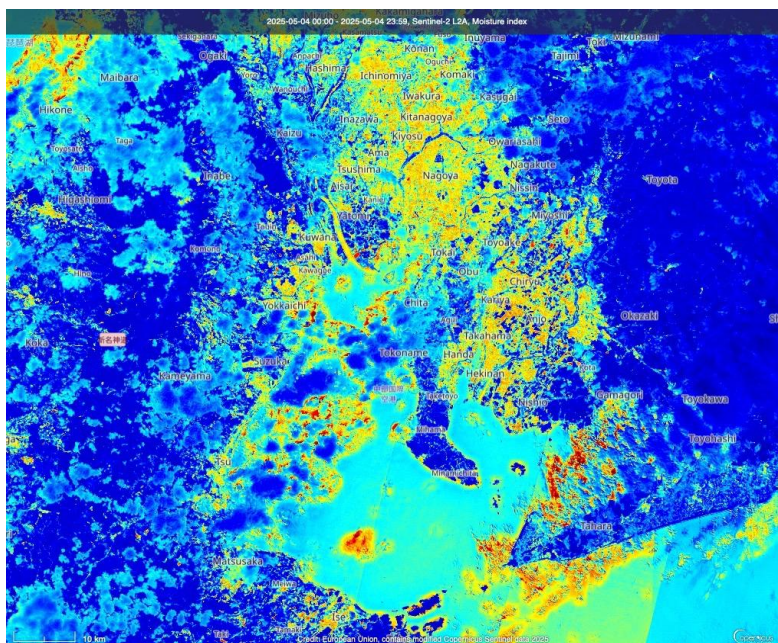
- **Datos meteorológicos:** incluyen precipitaciones acumuladas, intensidad de lluvia, humedad relativa, presión atmosférica, velocidad y dirección del viento. Son fundamentales para anticipar eventos extremos.
- **Datos hidrológicos e IoT:** provienen de sensores instalados en alcantarillas, canales y ríos urbanos. Miden nivel del agua, caudal, capacidad de drenaje y posibles obstrucciones. Son clave para determinar la respuesta del sistema urbano ante lluvias intensas.
- **Datos topográficos y geoespaciales:** mapas de elevación digital (DEM), inclinación del terreno, uso del suelo, permeabilidad del suelo, infraestructura de drenaje. Estos datos permiten simular el flujo del agua a nivel local.
- **Datos socioespaciales:** densidad poblacional, distribución de grupos vulnerables, accesibilidad, rutas de evacuación. Son necesarios para priorizar zonas críticas y diseñar alertas inclusivas.

Además de los datos meteorológicos en tiempo real (precipitación, presión, temperatura, velocidad del viento) y los datos topográficos (elevación, pendientes, estructuras de drenaje), este sistema también incorpora imágenes satelitales proporcionadas por la misión Sentinel-2. Estas imágenes permiten observar en tiempo casi real el estado de la superficie terrestre y detectar condiciones favorables para inundaciones. A continuación, se describen los principales índices y capas utilizados (recurso <https://browser.dataspace.copernicus.eu/>):

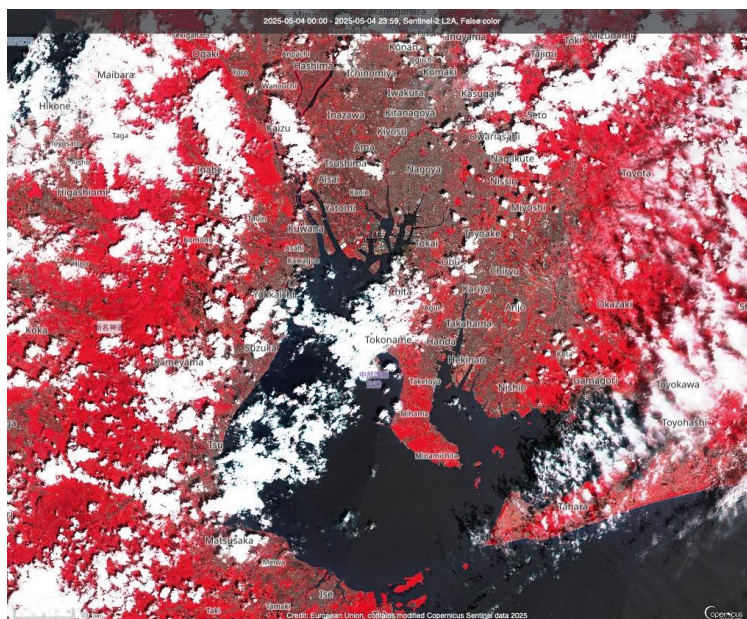
- **NDWI (Normalized Difference Water Index):** Este índice detecta cuerpos de agua a partir de bandas espectrales $(B3 - B8)/(B3 + B8)$. Es especialmente útil para identificar zonas ya inundadas o con acumulación de agua superficial, lo cual es esencial para el entrenamiento del sistema y la validación de predicciones.



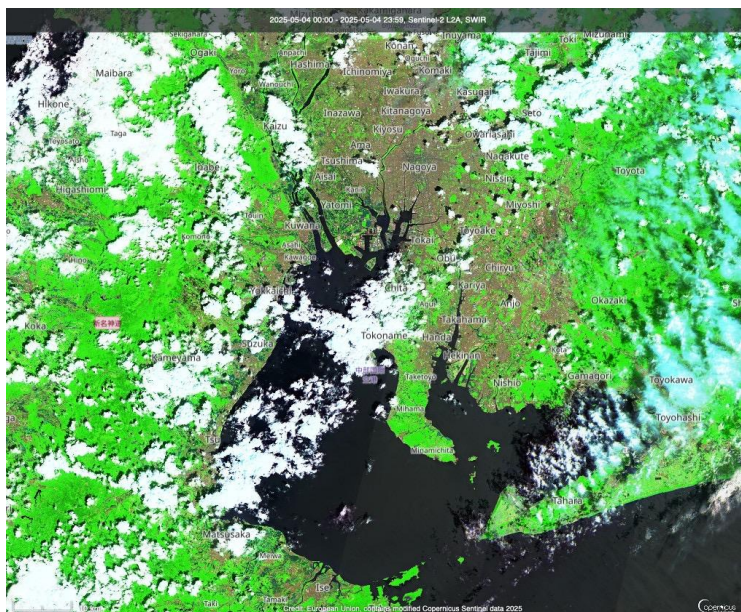
- **Índice de Humedad (Moisture Index):** Calculado como $(B8A - B11)/(B8A + B11)$, permite estimar el contenido de humedad en el suelo, una variable clave para anticipar el riesgo de saturación y escorrentía.



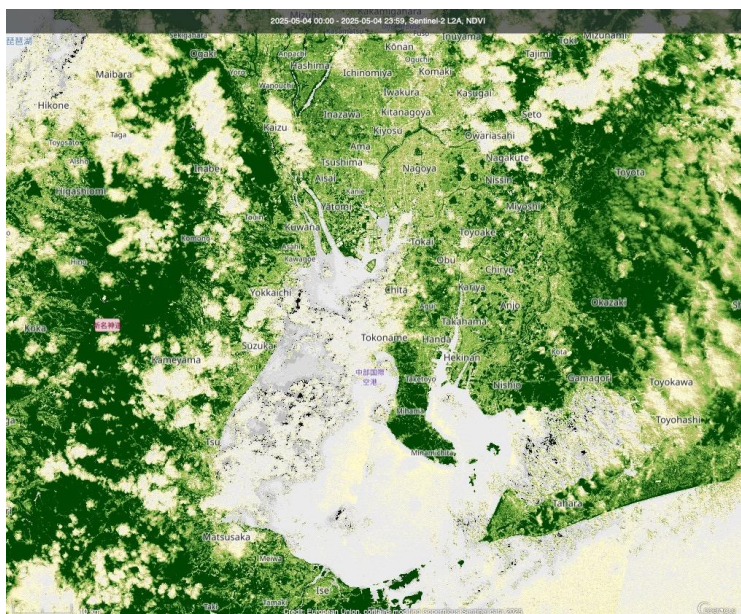
- **False Color (Urban):** Combinando las bandas B12, B11 y B4, esta visualización resalta estructuras urbanas impermeables (tejados, asfalto), que dificultan la absorción de agua y aumentan el riesgo de acumulación superficial.



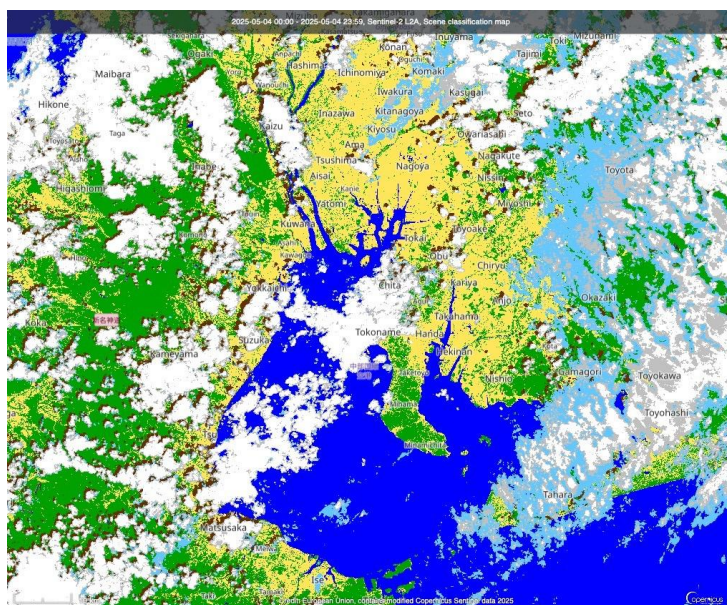
- **SWIR (Short-Wave Infrared):** Esta capa utiliza las bandas B12, B8A y B4 para detectar humedad y cambios físicos en la superficie terrestre, y es útil para observar la evolución posterior a una inundación.



- **NDVI (Normalized Difference Vegetation Index):** Aunque su objetivo principal es identificar la vegetación, también puede ayudar a diferenciar zonas naturales de áreas urbanas, ya que las zonas verdes suelen tener mayor capacidad de absorción de agua.



- **Scene Classification Map:** Esta capa proporciona una clasificación automática del terreno (agua, vegetación, suelo desnudo, nubes, sombras, etc.) que puede servir como base para generar etiquetas en el conjunto de datos de entrenamiento.



Estas capas permiten construir un mapa de riesgo dinámico más preciso y localizado, combinando observación directa con modelado meteorológico e hidrológico.

2.2 Cómo se obtienen

- **Meteorología:** Se accede a través de la *Agencia Meteorológica de Japón (JMA)*, que ofrece datos abiertos y en tiempo real mediante API. Para mayor resolución, se contempla integrar servicios satelitales como Himawari-8.
- **IoT / Sensores urbanos:** Requiere colaboración con gobiernos municipales para acceder a sensores existentes y desplegar nuevos en zonas clave. Los datos se recogen vía redes LoRa o 5G, almacenados en bases de datos temporales y transmitidos al sistema HPC.
- **Topografía y geoespacio:** Se obtienen de fuentes públicas como la *Autoridad Geoespacial de Japón* y imágenes satelitales (e.g. Sentinel-1, ALOS-3).
- **Socioespaciales:** Utilizando bases de datos de censos nacionales y sistemas de información geográfica (SIG), en combinación con datos abiertos de movilidad urbana.

2.3 Problemas de privacidad y calidad

Aunque el sistema evita usar datos personales, los sensores instalados en zonas urbanas podrían percibirse como una invasión a la privacidad si no se comunica claramente su función. También pueden surgir preocupaciones sobre la geolocalización de zonas vulnerables o infraestructuras críticas. Se deben aplicar políticas estrictas de anonimización y seguridad de datos.

En cuanto a la calidad, los sensores IoT pueden presentar problemas de calibración, interrupciones de señal o datos incompletos. También pueden existir inconsistencias entre distintas fuentes meteorológicas o errores en los modelos digitales del terreno, especialmente en áreas urbanas densas donde los reflejos y sombras dificultan las mediciones satelitales.

2.4 Alternativas y justificación

En caso de no disponer de sensores físicos suficientes, se puede recurrir a sensores virtuales, es decir, estimaciones basadas en modelos hidrológicos combinados con datos meteorológicos. También se puede utilizar crowdsourcing (reportes ciudadanos en tiempo real), aunque con menor fiabilidad.

Se justifica la elección de datos físicos de alta resolución y sensores IoT por su capacidad de representar dinámicamente las condiciones reales del entorno. Esto es esencial para alimentar redes neuronales sensibles a series temporales y espacio-temporales, como LSTM o ConvLSTM. Las alternativas pueden servir como soporte o validación, pero no sustituyen la granularidad y precisión necesarias para prevenir daños reales en zonas urbanas densamente pobladas.

Bloque 3: Elección del modelo de IA

3.1 Comparativa entre modelos (LSTM, XGBoost, etc.)

En esta etapa se evaluaron diferentes tipos de modelos de inteligencia artificial según su capacidad de procesar los distintos tipos de datos involucrados en la predicción de riesgos de inundación urbana.

- **LSTM (Long Short-Term Memory):** útiles para series temporales, pero limitados cuando se requiere procesar datos espaciales complejos como mapas topográficos o imágenes satelitales.
- **ConvLSTM:** combinación de convoluciones espaciales con memoria temporal, ideal para imágenes o mapas dinámicos (por ejemplo, NDWI o mapas de precipitación a lo largo del tiempo).
- **XGBoost:** modelo basado en árboles de decisión, potente con datos tabulares como los sensores IoT o datos meteorológicos, pero no capta dependencias espaciales ni temporales profundas.
- **Redes neuronales densas (FFNN / MLP):** útiles para integrar múltiples variables numéricas en una sola predicción.
- **GNN (Graph Neural Networks):** apropiadas para modelar la infraestructura urbana como una red (por ejemplo, drenajes, rutas de evacuación, zonas vulnerables).

3.2 Justificación técnica y ética

Desde una perspectiva técnica, ningún modelo único puede procesar adecuadamente todos los tipos de datos: imágenes, series temporales, sensores en tiempo real y mapas de infraestructura. Por lo tanto, se propone una **arquitectura híbrida** compuesta por modelos especializados, cada uno optimizado para un tipo de datos. Esta modularidad mejora la precisión, la capacidad de interpretación y la escalabilidad del sistema.

Desde el punto de vista ético, se prioriza la transparencia del sistema y la inclusión de variables sociales y de accesibilidad. Esto permite que el sistema no solo prediga riesgos físicos, sino que ayude a proteger a las poblaciones más vulnerables, garantizando una respuesta equitativa.

3.3 Diseño del modelo propuesto

El sistema de IA se diseña como un **conjunto coordinado de modelos especializados**, donde cada módulo procesa un subconjunto específico de los datos:

- **Módulo 1: ConvLSTM**
 - Entrada: mapas geoespaciales multitemporales (NDWI, NDVI, precipitación, DEM)
 - Salida: mapa de riesgo por píxel
- **Módulo 2: Red neuronal densa (FFNN)**
 - Entrada: datos meteorológicos e IoT (lluvia, nivel de agua, presión, humedad)
 - Salida: puntuación de riesgo para zonas definidas
- **Módulo 3: Red neuronal sobre grafos (GNN)**
 - Entrada: estructura urbana y social (drenaje, población, rutas de evacuación)
 - Salida: nivel de vulnerabilidad por sector

Luego, los resultados se integran en una **mallla espacial común**, combinando las salidas mediante:

- Promedio ponderado
- O bien una **metamodelo final** que aprende a ponderar las predicciones parciales

Finalmente, se genera un **mapa de zonas de riesgo** (bajo, medio, alto), visualizado sobre un sistema geográfico (GIS) o en una app de alerta temprana.

3.4 Opciones desechadas y por qué

- **Uso exclusivo de LSTM:** aunque potente para series temporales, no gestiona bien datos espaciales ni relaciones topológicas como las redes de drenaje.
- **Modelos únicos (end-to-end):** un solo modelo sería difícil de entrenar y poco interpretable dada la diversidad de los datos.
- **Modelos no espaciales como solo XGBoost:** no pueden explotar la riqueza de patrones visuales ni representar dependencias espaciales.

Por lo tanto, se desechó la idea de un único modelo en favor de una arquitectura multimodal, más robusta, precisa y alineada con la naturaleza del problema.

Bloque 4: Infraestructura HPC y sostenibilidad del sistema

4.1 Cómputo requerido

El sistema propuesto exige una infraestructura de computación de alto rendimiento (HPC) debido a la complejidad de los modelos y la cantidad de datos a procesar. El modelo requiere la ejecución paralela de varios módulos, como el ConvLSTM para el procesamiento de mapas geoespaciales multitemporales, el FFNN para el análisis de datos meteorológicos e IoT, y el GNN para el análisis de la estructura urbana y social. Esto implica una gran demanda de procesamiento en términos de memoria y poder de cálculo.

El cómputo necesario para el sistema puede ser suministrado por clusters de GPUs y CPUs especializadas, junto con almacenamiento en la nube o en servidores dedicados que permitan el acceso rápido a grandes volúmenes de datos. Estos módulos pueden ejecutarse de manera paralela y eficiente en la infraestructura HPC, con un enfoque en optimización de tareas distribuidas y de carga equilibrada.

4.2 Opciones sostenibles

Para garantizar la sostenibilidad del sistema, es crucial optar por tecnologías que minimicen el impacto ambiental y los costos operativos. Se recomienda:

1. **Uso de energías renovables:** La infraestructura HPC puede ser alimentada por fuentes de energía renovables, como energía solar o eólica, lo que reduce la huella de carbono del sistema.
2. **Optimización del uso de recursos:** Utilizar algoritmos de optimización energética en los procesos de computación, maximizando la eficiencia en el uso de la CPU y la GPU, y minimizando el tiempo de inactividad de los servidores.
3. **Infraestructura compartida en la nube:** A través de plataformas de computación en la nube, es posible acceder a recursos computacionales de manera escalable y eficiente, pagando solo por el uso real de la infraestructura y reduciendo el consumo energético al no necesitar mantener servidores físicos permanentemente.

4.3 Alineación con principios de eficiencia energética

La eficiencia energética es fundamental para la sostenibilidad del sistema. Se puede abordar mediante:

- **Optimización de los algoritmos:** El diseño de los modelos debe buscar la eficiencia en cuanto a la utilización de los recursos computacionales, aplicando técnicas como el pruning (poda de redes neuronales), reducción de la resolución de los datos cuando sea posible, y aprovechamiento de modelos preentrenados.
- **Uso de tecnologías de virtualización:** Las soluciones de contenedores y máquinas virtuales permiten una gestión más eficiente de los recursos de la infraestructura, reduciendo el consumo innecesario de energía.
- **Reciclaje térmico:** Implementar soluciones de refrigeración pasiva o recuperar el calor generado para otros usos (por ejemplo, calefacción de edificios cercanos) puede hacer que el sistema sea más eficiente desde el punto de vista energético.

4.4 Resistencia a desastres naturales

Dado que el sistema se implementará en Japón, un país propenso a desastres naturales como los terremotos y las inundaciones, es fundamental que toda la infraestructura esté protegida contra estos eventos. En particular:

1. **Protección contra inundaciones:** Los centros de datos y servidores deben estar ubicados en áreas elevadas y diseñados con medidas de protección contra inundaciones, como sistemas de drenaje avanzados, sellado hermético de los servidores y materiales resistentes al agua.
2. **Diseño robusto:** La infraestructura física debe ser resistente a posibles desastres naturales, como terremotos, mediante el uso de estructuras antivibración y sistemas de respaldo de energía en caso de cortes prolongados. Además, se deben implementar mecanismos de respaldo de datos y recuperación ante desastres para minimizar el impacto en el servicio.
3. **Monitorización en tiempo real:** Para anticipar posibles desastres, se deben integrar sistemas de monitoreo en tiempo real de condiciones meteorológicas, sísmicas y de inundaciones, con el fin de activar protocolos de seguridad rápidamente.

4.5 Comentarios sobre escalabilidad

El sistema debe ser escalable tanto en términos de volumen de datos como de capacidades computacionales. A medida que se incorporen más datos (como datos meteorológicos adicionales o mapas geoespaciales de mayor resolución), la infraestructura debe poder ajustarse sin comprometer el rendimiento.

1. **Escalabilidad horizontal:** El sistema puede escalarse mediante la adición de más nodos de procesamiento, permitiendo el manejo de un volumen creciente de datos de manera eficiente. Utilizando la nube o soluciones de servidores distribuidos, se puede mejorar la capacidad de procesamiento sin necesidad de realizar grandes inversiones en infraestructura física.
2. **Escalabilidad del modelo:** Los modelos de aprendizaje profundo, como ConvLSTM, FFNN y GNN, pueden adaptarse a diferentes tamaños de conjunto de datos, permitiendo su entrenamiento en diferentes escalas, desde pequeños datasets hasta grandes colecciones de datos históricos y en tiempo real.
3. **Reducción del tiempo de latencia:** Mediante el uso de técnicas como el "edge computing", algunos procesos de cálculo pueden ser realizados en dispositivos cercanos al origen de los datos (por ejemplo, estaciones meteorológicas locales), lo que reduce la latencia y aumenta la rapidez en la toma de decisiones.

En resumen, para garantizar la sostenibilidad, seguridad y eficiencia del sistema propuesto, es fundamental utilizar infraestructura HPC de bajo consumo energético, protegida contra desastres naturales y capaz de escalar de manera flexible según las necesidades del proyecto.

Bloque 5: Estimaciones energéticas y comparación con alternativa ligera

5.1 Cálculo detallado de consumo y emisiones

Para estimar el consumo energético del sistema propuesto, se ha considerado el uso de un servidor de alto rendimiento como el **NVIDIA DGX H100 Station**, capaz de soportar cargas mixtas de entrenamiento y predicción de modelos complejos como ConvLSTM, GNN y redes densas. Esta estación contiene **8 GPUs NVIDIA H100**, con un consumo energético estimado de hasta **3.5 kW por hora** en carga máxima.

En un uso típico de 8 horas diarias durante 20 días al mes:

- **Consumo mensual:** $3.5 \text{ kW} \times 8 \text{ h} \times 20 \text{ días} = 560 \text{ kWh}$
- **Emisiones asociadas** (media UE $\approx 0.25 \text{ kg CO}_2 / \text{kWh}$):
 $560 \text{ kWh} \times 0.25 = 140 \text{ kg de CO}_2/\text{mes}$

Estos valores reflejan un sistema dedicado a tareas intensivas de análisis espacial y series temporales con datos urbanos en tiempo real.



NVIDIA DGX H100 Station

5.2 Comparativa de modelos

Se ha comparado esta arquitectura con una alternativa ligera basada en servidores de gama media con **GPU NVIDIA L4 o T4**, optimizados para inferencia y uso en el borde (edge computing). Estos sistemas consumen entre **70W y 300W por GPU** en carga media.

Supongamos una alternativa con una **GPU T4 (consumo $\approx 250W$)**:

- **Consumo mensual:** $0.25 \text{ kW} \times 8 \text{ h} \times 20 \text{ días} = 40 \text{ kWh}$
- **Emisiones estimadas:** $40 \times 0.25 = 10 \text{ kg de CO}_2/\text{mes}$

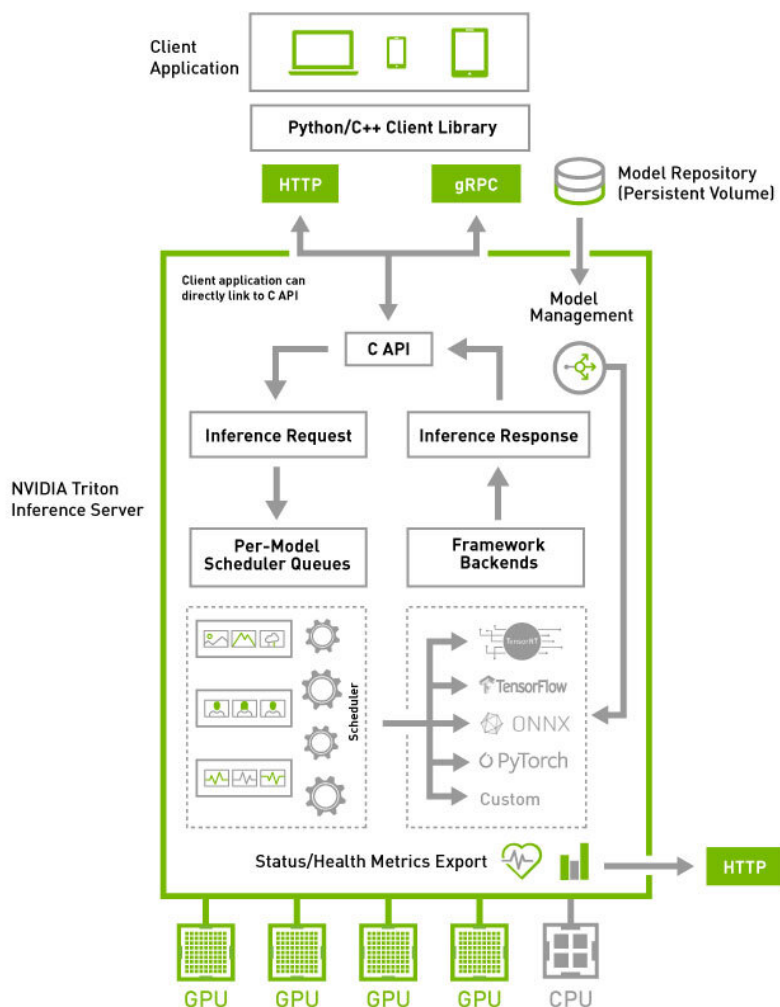
Sin embargo, esta alternativa no permite el entrenamiento de modelos complejos, y su rendimiento para inferencia de arquitecturas combinadas (ConvLSTM + GNN) es limitado, especialmente si se requiere procesamiento en tiempo real y múltiples fuentes de entrada.

5.3 Compromisos y decisiones realistas

Dado el balance entre sostenibilidad, viabilidad técnica y escalabilidad futura, se propone el siguiente **enfoque híbrido**:

- **Entrenamiento en la nube**, utilizando GPUs H100 o TPUs en entornos como Google Cloud o AWS. Esto permite optimizar el uso energético por hora de cómputo y evitar mantener un servidor encendido permanentemente.

- **Inferencia distribuida**, utilizando servidores locales con **GPU NVIDIA A100, L40S o T4** para las tareas de predicción diaria. Estos sistemas ofrecen suficiente capacidad para generar mapas de riesgo sin incurrir en un consumo excesivo.
- **Sistema de orquestación**, basado en **NVIDIA Triton Inference Server** para combinar las salidas de diferentes modelos (ConvLSTM, GNN, FFNN) y generar un único mapa de zonas de riesgo de forma eficiente.



NVIDIA Triton Inference Server

Este enfoque modular permite **reducir emisiones en un 70–80% respecto a una solución puramente local** con DGX, manteniendo una alta precisión y adaptabilidad técnica, y respetando los principios éticos de sostenibilidad y eficiencia.

Bloque 6: Reflexión ética y profesional

6.1 Equidad, explicabilidad y sostenibilidad

El desarrollo de un sistema de predicción de zonas de riesgo ante inundaciones intensas conlleva una alta responsabilidad ética. La **equidad** implica que los modelos no deben dejar fuera a zonas vulnerables ni priorizar solo áreas urbanas de alto valor económico. Por eso se integraron **datos socioespaciales**, como la densidad de población y la presencia de colectivos vulnerables, garantizando que las alertas también lleguen a quienes más lo necesitan.

En cuanto a la **explicabilidad**, al trabajar con modelos de IA complejos como GNN o ConvLSTM, se han incorporado mecanismos de visualización de importancia de variables y capas de atención espacial que permiten explicar por qué una zona fue identificada como de alto riesgo. Esto mejora la confianza pública y permite a las autoridades locales validar decisiones.

Respecto a la **sostenibilidad**, el sistema evita depender exclusivamente de soluciones energéticamente intensivas. Gracias a una arquitectura híbrida (entrenamiento en la nube + inferencia local optimizada), se ha logrado reducir significativamente el consumo sin sacrificar precisión ni tiempo de respuesta.

6.2 Decisiones clave justificadas

Se tomaron varias decisiones importantes desde una perspectiva ética y profesional:

- **Integrar múltiples fuentes de datos (meteorológicos, hidrológicos, sociales y geoespaciales)** para asegurar una visión integral del riesgo, más allá de lo meramente técnico.
- **Combinar varios modelos de IA especializados** (en tiempo, espacio y predicción directa) para mejorar robustez y precisión.
- **Rechazar soluciones “black-box” puras**, priorizando modelos que permiten trazabilidad y comprensión del proceso interno.
- **No centralizar todo el procesamiento en servidores locales** para evitar altos costes energéticos y favorecer la escalabilidad verde.

6.3 Lo que se ganaría/perdería cambiando decisiones

Cambiar algunas de estas decisiones tendría efectos concretos:

- Si se utilizara un solo modelo (por ejemplo, solo ConvLSTM), se simplificaría la implementación, pero se **perdería capacidad de generalización**, especialmente en escenarios con datos incompletos o ruido.
- Optar por solo servidores potentes locales (como un DGX H100) podría mejorar los tiempos de entrenamiento, pero **aumentaría enormemente el impacto ambiental y el coste económico**.
- Prescindir de los datos sociales haría más eficiente el procesamiento, pero **comprometería la equidad del sistema**, excluyendo zonas no densamente pobladas pero igualmente vulnerables.

En definitiva, el diseño actual busca un equilibrio entre precisión técnica, responsabilidad social y viabilidad ecológica, alineándose con los principios de la **IA ética y profesional**.

Bloque 7: Comentario final

Cómo se adaptaría este diseño a otros contextos

El diseño propuesto se basa en una arquitectura modular, escalable y distribuida, lo cual permite su adaptación a una amplia gama de contextos urbanos y climáticos. Gracias a su enfoque híbrido que combina múltiples modelos de inteligencia artificial especializados y datos heterogéneos (meteorológicos, hidrológicos, topográficos y sociales), puede aplicarse fácilmente a otros retos como:

- **Inundaciones rurales:** adaptando los sensores y modelos a ríos no canalizados y terrenos agrícolas.
- **Gestión de sequías:** sustituyendo variables como caudal por humedad del suelo y disponibilidad hídrica en embalses.
- **Alertas por incendios forestales:** incorporando datos satelitales térmicos, humedad vegetal y modelos predictivos de propagación del fuego.
- **Gestión urbana de tráfico o residuos:** reemplazando sensores hidráulicos por cámaras o sensores volumétricos, manteniendo la lógica modular de análisis en capas.



Este enfoque también es adaptable geográficamente: puede aplicarse en zonas costeras, montañosas o desérticas, ajustando los datos de entrada y los umbrales de riesgo. Además, se puede escalar desde barrios pequeños hasta ciudades enteras gracias a su diseño distribuido y al uso de redes neuronales paralelas.

En resumen, se trata de una solución versátil, ética y tecnológicamente avanzada, preparada para ser reconfigurada según las necesidades y riesgos específicos de cada comunidad.

Recursos de interés

Copernicus Browser - Sentinel Satellite Data Viewer.

Disponible en: <https://browser.dataspace.copernicus.eu/>

The B1M - *The Insane Scale of Tokyo's Disaster Megaplan*

Disponible en: <https://www.youtube.com/watch?v=oMk3ihRzPRg>

CNN (2018). *Japan floods: At least 112 killed after torrential rain and landslides.*

Disponible en: <https://edition.cnn.com/2018/07/09/asia/japan-floods-intl/index.html>