




Effectiveness of Machine Learning in environmental pollution from remote sensing images. Systematic Literature Review

Christian Dios-Castillo, Dr en Administración de la Educación¹, Carlos Chafloque-Llontop, Estudiante de Ingeniería de Sistemas e Informática¹, and Raúl Sánchez-Rentería, Estudiante de Ingeniería de Sistemas e Informática¹

¹Universidad Tecnológica del Perú, Perú, cdios@utp.edu.pe, U19309934@utp.edu.pe, U19311375@utp.edu.pe

Abstract– Machine learning algorithms for the analysis of remote sensing images make it possible to identify and track sources of environmental pollution effectively. The objective or purpose of this research was to evaluate the effectiveness of these techniques in the identification and monitoring of sources of environmental pollution through remote sensing images. The systematic review of scientific articles on the topic in question was proposed as a methodology, with the support of the PICO questions and the PRISMA procedure for the analysis of these articles. The information was obtained from the Scopus database, obtaining 58 documents of open access and from the last 5 years. The results found show that machine learning algorithms allow a more accurate identification of sources of environmental pollution compared to conventional methods. It is concluded that these techniques are a valuable tool for the identification and monitoring of sources of environmental pollution, which can contribute to decision-making and the implementation of measures to reduce environmental pollution.

Keywords-- machine learning, remote sensing, environmental contamination, deep learning.

Efectividad del Machine Learning en el medio ambiente contaminación de imágenes de teledetección. Sistemático Revisión de la literatura

Christian Dios-Castillo, Dr en Administración de la Educación¹, Carlos Chafloque-Llontop, Estudiante de Ingeniería de Sistemas e Informática¹, and Raúl Sánchez-Rentería, Estudiante de Ingeniería de Sistemas e Informática¹

¹Universidad Tecnológica del Perú, Perú, cdios@utp.edu.pe, U19309934@utp.edu.pe, U19311375@utp.edu.pe

Resumen– Los algoritmos de machine learning para el análisis de imágenes de teledetección permiten identificar y rastrear fuentes de contaminación ambiental de manera efectiva. El objetivo o propósito de esta investigación fue evaluar la eficacia de estas técnicas en la identificación y seguimiento de fuentes de contaminación ambiental a través de imágenes de teledetección. Se planteó como metodología la revisión sistemática de artículos científicos sobre el tema en cuestión, teniendo como apoyo las preguntas PICO y el procedimiento PRISMA para el análisis de estos artículos. La obtención de la información se realizó a partir de la base de datos Scopus obteniéndose 58 documentos de acceso libre y de los últimos 5 años. Los resultados encontrados evidencian que los algoritmos de machine learning permiten una identificación más precisa de fuentes de contaminación ambiental en comparación con los métodos convencionales. Con lo cual se concluye que estas técnicas son una herramienta valiosa para la identificación y seguimiento de fuentes de contaminación ambiental, lo que puede contribuir a la toma de decisiones y la implementación de medidas para reducir la contaminación ambiental.

Palabras clave– machine learning, remote sensing, environmental contamination, deep learning.

I. INTRODUCCION

En la actualidad, se observa un creciente interés en la calidad del entorno ambiental y su influencia en la salud de las personas. Esto ha llevado a que se realicen investigaciones sobre cómo detectar y reducir la contaminación ambiental. Un enfoque prometedor y novedoso en este campo involucra el uso de métodos de aprendizaje automático junto con imágenes de teledetección. Varias investigaciones han examinado este enfoque innovador en una variedad de contextos ambientales. Sin embargo, a pesar del avance rápido en el uso del aprendizaje automático para identificar y rastrear fuentes de contaminación, aún existen dudas sobre su eficacia en comparación con los métodos tradicionales [1]. El núcleo de este estudio se centra en evaluar cuán efectivas son los métodos de aprendizaje automático en el procesamiento de imágenes de teledetección con el propósito de identificar y rastrear fuentes de contaminación ambiental. Aunque estas técnicas han demostrado ser muy efectivas en áreas como la clasificación de objetos y la detección de patrones, su aplicación específica en la detección de contaminantes ambientales y la identificación de sus orígenes aún no se ha

explorado lo suficiente. Este problema se agrava aún más debido a la falta de investigaciones que se centren en esta perspectiva en el contexto de la contaminación ambiental [1]–[3]. La pandemia de COVID-19 ha tenido un impacto dual en la contaminación ambiental. La reducción de la actividad industrial y del transporte ha disminuido las emisiones de contaminantes, mientras que el uso creciente de materiales desechables ha aumentado la generación de residuos. Este fenómeno ha ocasionado cambios en la calidad del aire en diversas regiones, con algunas experimentando mejorías y otros empeoramientos, dependiendo de la gestión de residuos y la actividad económica [4], [5]. Por otra parte, el uso de técnicas de aprendizaje profundo en problemas de visión por computadora ha evidenciado avances notables. Las Redes Neuronales Convolucionales, como una de estas técnicas, han encontrado aplicación en diversos campos, incluyendo la detección de objetos, reconocimiento facial, acciones humanas y estimación de postura [6]. Este avance en el aprendizaje profundo podría ser extrapolado para mejorar las técnicas de detección de contaminación ambiental a través de imágenes de teledetección. La justificación de este estudio se fundamenta en la necesidad de evaluar cómo las técnicas de aprendizaje automático pueden ayudar en la detección y rastreo de fuentes de contaminación en el entorno ambiental. Los resultados de esta investigación influirán de manera importante en la toma de decisiones y en la implementación de medidas para reducir la contaminación ambiental. Además, se busca responder a la pregunta crucial de por qué estas técnicas funcionan en algunos casos y en otros no. Esta comprensión profunda permitirá mejorar continuamente los métodos utilizados y su aplicación en proyectos específicos para controlar y prevenir la contaminación ambiental [2], [3], [7].

II. METODOLOGÍA

El presente estudio se llevó a cabo mediante una revisión sistemática de la literatura haciendo uso del enfoque PICO, reconocido por su aplicación en diversos campos de investigación. El esquema PICO ofrece una técnica universal para estructurar preguntas de investigación, independientemente de la disciplina. Se resalta que este enfoque proporciona claridad y precisión en la formulación de preguntas de investigación [8]. la utilización de la estructura

PICO ayuda a los autores a formular preguntas precisas, lo que facilita la comprensión del objetivo del estudio por parte de los lectores, así como identifica los elementos clave del estudio, como la población, la intervención, el grupo de comparación y los resultados [9]. En consonancia con estas perspectivas, se formula la siguiente pregunta de investigación para este estudio: ¿Cuál es la eficacia de las técnicas de machine learning en la identificación y seguimiento de fuentes de contaminación ambiental a través de imágenes de teledetección?

TABLA I
ESTRUCTURA PICO

PICO	PREGUNTA	PALABRAS CLAVE
P	¿Cuál es la incidencia de la contaminación ambiental en áreas específicas como sitios contaminados, fuentes de contaminación y la presencia de contaminantes en el entorno?	Environmental contamination, Contaminated sites, Pollution sources, Water pollution, Air pollution, Soil pollution, Industrial pollution y Contaminants in the environment
I	¿Cómo se aplican las técnicas de Aprendizaje automático en la identificación y seguimiento de fuentes de contaminación ambiental en imágenes de teledetección?	Machine Learning, Artificial intelligence, Deep learning, Neural networks, Data mining, Predictive modeling, Computer visión y Pattern recognition
C	Pregunta no aplicable (dado que no estamos comparando con otra intervención, se centra en evaluar la efectividad de la intervención de aprendizaje automático).	No aplicable (dado que no estamos comparando con otra intervención, se centra en evaluar la efectividad de la intervención de aprendizaje automático).
O	¿Cuáles son los métodos y resultados más comunes utilizados para la identificación y seguimiento de fuentes de contaminación ambiental en imágenes de teledetección, incluyendo la identificación de fuentes de contaminación, el seguimiento de fuentes contaminantes y la evaluación de la contaminación ambiental?	Pollution source identification, Contaminant source tracking, Remote sensing, Teledetection, Satellite imagery, Image análisis, Detection of pollutants in images y Environmental pollution assessment

Este estudio se fundamenta en una investigación bibliográfica realizada utilizando la base de datos SCOPUS, involucrando una búsqueda automatizada de términos relevantes. La ecuación de búsqueda utilizada para esta investigación fue la siguiente: (TITLE-ABS-KEY ("Environmental contamination" OR "Contaminated sites" OR "Pollution sources" OR "Water pollution" OR "Air pollution" OR "Soil pollution" OR "Industrial pollution" OR "Contaminants in the environment") AND TITLE-ABS-KEY ("Aprendizaje automático" OR "Artificial intelligence" OR "Deep learning" OR "Neural networks" OR "Data mining" OR "Predictive modeling" OR "Computer vision" OR "Pattern recognition") AND TITLE-ABS-KEY ("Pollution source identification" OR "Contaminant source tracking" OR "Remote sensing" OR "Teledetection" OR "Satellite imagery" OR "Image analysis" OR "Detection of pollutants in images" OR "Environmental pollution assessment"))

Criterios de inclusión y exclusión:
Los criterios para la inclusión y la exclusión fueron establecidos en función de los objetivos de la investigación.

TABLA II
CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

CRITERIOS DE INCLUSIÓN	CRITERIOS DE EXCLUSIÓN
CI1: Los estudios incluidos deben abordar el tema de contaminación ambiental.	CE1: Artículos que no estén disponibles en acceso abierto.
CI2: Los estudios deben aplicar técnicas de Aprendizaje automático y Teledetección en relación con la contaminación ambiental.	CE2: Publicaciones en idiomas diferentes al inglés.
CI3: Los estudios relacionados con la contaminación ambiental deben estar disponibles en formato PDF o texto y ser de acceso abierto.	CE3: Excluir otros tipos de publicaciones que no sean artículos científicos.
CI4: Los estudios relacionados con las áreas temáticas de Ciencias del Medio Ambiente y Ciencias de la computación.	CE4: Publicaciones no están relacionados con las áreas temáticas de Ciencias del Medio Ambiente y Ciencias de la Computación
CI5: Los estudios incluidos deben haber sido publicados entre los años 2018 y 2022	CE5: Publicaciones publicadas antes de 2018 o después de 2022

Cadena de búsqueda final:
(TITLE-ABS-KEY ("Environmental contamination" OR "Contaminated sites" OR "Pollution sources" OR "Water pollution" OR "Air pollution" OR "Soil pollution" OR "Industrial pollution" OR "Contaminants in the environment") AND TITLE-ABS-KEY ("Aprendizaje automático" OR "Artificial intelligence" OR "Deep learning" OR "Neural networks" OR "Data mining" OR "Predictive modeling" OR "Computer vision" OR "Pattern recognition") AND TITLE-ABS-KEY ("Pollution source identification" OR

"Contaminant source tracking" OR "Remote sensing" OR "Teledetection" OR "Satellite imagery" OR "Image analysis" OR "Detection of pollutants in images" OR "Environmental pollution assessment")) AND PUBYEAR > 2017 AND PUBYEAR < 2023 AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (SUBJAREA , "ENVI") OR LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))

Proceso de selección de estudios:

Después de la identificación de 675 artículos obtenidos en SCOPUS la fecha 20/09/2023, se consideró apropiado llevar a cabo una revisión sistemática siguiendo las pautas establecidas por la declaración PRISMA. Esta declaración se utiliza para obtener resultados de las publicaciones científicas relacionadas con los objetivos de la investigación. El protocolo PRISMA se fundamenta en un proceso de dos etapas de filtrado. En la primera etapa, se aplica un filtro basado en los criterios de inclusión y exclusión establecidos para la investigación del proyecto, evaluando elementos como el título, el resumen, el año de publicación, entre otros. En la segunda etapa, se realiza un filtrado adicional basado en los mismos criterios, pero esta vez se analiza el texto completo de los documentos.

A partir de la obtención de 675 artículos, se dio comienzo a la primera etapa de filtrado basada en los criterios de inclusión y exclusión mencionados anteriormente. Siguiendo estos criterios, se excluyeron un total de 590 publicaciones por varias razones: 423 publicaciones no eran de acceso abierto, 6 estaban escritas en idiomas distintos al inglés, 44 provenían de diferentes artículos y 72 no estaban relacionadas con las áreas de Ciencias Medioambientales y Ciencias de la Computación. Posteriormente, se eliminaron 45 artículos que no estaban comprendidos en el período de 2018 a 2022, y hubo una publicación que no pudo ser recuperada en su totalidad (PDF o HTML). Esto resultó en un total de 84 registros seleccionados para evaluar su idoneidad. Finalmente, después de una lectura completa de los artículos seleccionados, se excluyeron 12 artículos que no se enfocan en análisis de imágenes y 14 artículos que no tratan el tema de interés. En consecuencia, se incluyeron 64 registros para la revisión sistemática.

Este enfoque, basado en la declaración PRISMA, se apoya en una guía especializada para autores de revisiones sistemáticas, garantizando informes transparentes, completos y precisos. La declaración PRISMA 2020, una actualización de 2009, ofrece pautas actualizadas sobre cómo presentar informes, incluyendo una lista detallada de 27 elementos, recomendaciones específicas y diagramas de flujo, fundamentales para las revisiones sistemáticas [10]. Mejora la calidad de informes en estos campos, facilitando la construcción sobre trabajos anteriores y simplificando la realización y actualización de revisiones, además, garantiza evaluaciones más precisas de los manuscritos de metaanálisis [11].

El uso de Prisma no solo asegura la transparencia y exhaustividad en la selección de estudios, sino que también se ajusta a los estándares actualizados en presentación de informes en revisiones sistemáticas, beneficiando a la comunidad científica en general [10], [11].

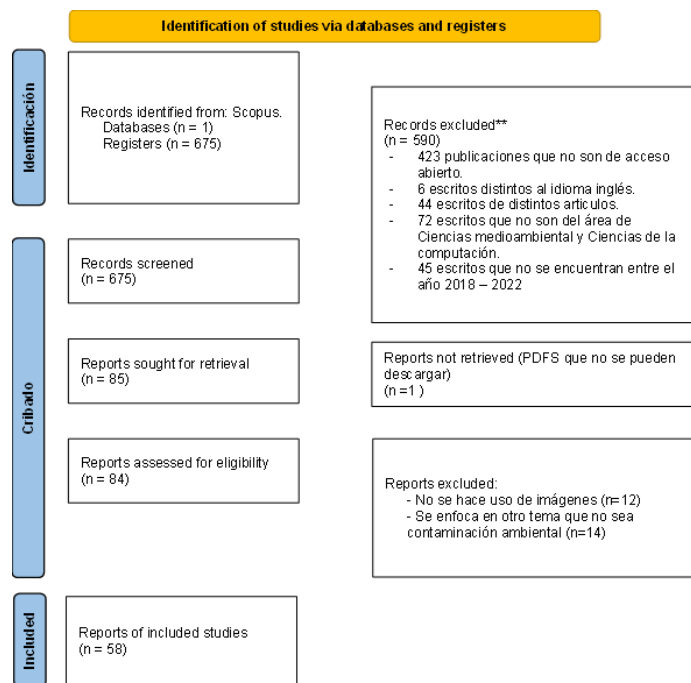


Figura 1. Diagrama PRISMA.

III. RESULTADOS Y DISCUSIONES

El informe de investigación presenta una recopilación de los artículos examinados, los cuales han sido categorizados según su fecha de publicación, lugar de origen y campo temático. Al examinar la figura 2, se aprecia el análisis de los datos abarcando el período comprendido entre 2018 y 2022, lo que permite concluir que en 2022 se experimentó un aumento significativo de 33 publicaciones de artículos. Además, se aprecia un incremento en las publicaciones de artículos durante los años 2021 y 2022. En esta investigación de revisión, se presenta un informe detallado de los artículos examinados, organizados meticulosamente según su fecha de publicación, lugar de origen y temática. Al analizar la Figura 2, se llevó a cabo un estudio minucioso de los datos que abarcan el periodo entre 2018 y 2022. A diferencia de años anteriores, se destaca un aumento significativo de 33 publicaciones de artículos en 2022, marcando un crecimiento notable en la producción científica. También se encontró que, en comparación con el pasado, hay un incremento evidente en las publicaciones de artículos durante los años 2021 y 2022, indicando una tendencia ascendente en la actividad de investigación. En contraste con la posible estabilidad en años anteriores, estos hallazgos subrayan un renovado interés y una contribución

continua al campo temático analizado durante el periodo en estudio. Este análisis proporciona una visión actualizada y detallada de la evolución temporal de la investigación científica en comparación con investigaciones previas.

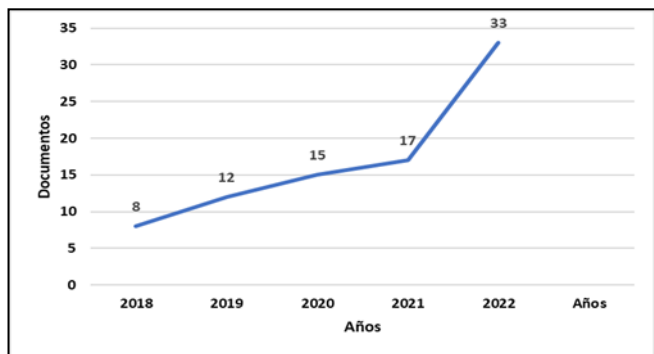


Figura 2. Artículos organizados por año de publicación.

Al analizar el gráfico de la Figura 3 que representa la distribución de artículos según su país de origen, se destaca que China es el país más prominente, contribuyendo con aproximadamente el 31% del total, seguido de Estados Unidos, que representa alrededor del 19%. A medida que se desciende en el gráfico, se aprecia que otros países, como el Reino Unido, Australia, Italia, Francia, Hong Kong y los Países Bajos, tienen una participación más reducida, cada uno contribuyendo en torno al 5%, 3%, 3%, 3%, 3% y 3%, respectivamente, en la producción de documentos científicos. En este estudio de revisión, se examinó detalladamente la distribución de artículos científicos según su país de origen, como se ilustra en el gráfico de la Figura 3. En comparación con otros países, China se destaca como el país más prominente, contribuyendo con aproximadamente el 31% del total. A diferencia de China, Estados Unidos ocupa la segunda posición, representando alrededor del 19% de la producción científica analizada. También se encontró que, a medida que se desciende en el gráfico, otros países como el Reino Unido, Australia, Italia, Francia, Hong Kong y los Países Bajos tienen una participación más reducida. Cada uno contribuye en torno al 5%, 3%, 3%, 3%, 3% y 3%, respectivamente, en la producción de documentos científicos. En contraste con las contribuciones más destacadas de China y Estados Unidos, estos países presentan una presencia más modesta en el panorama de la investigación examinada en este análisis.

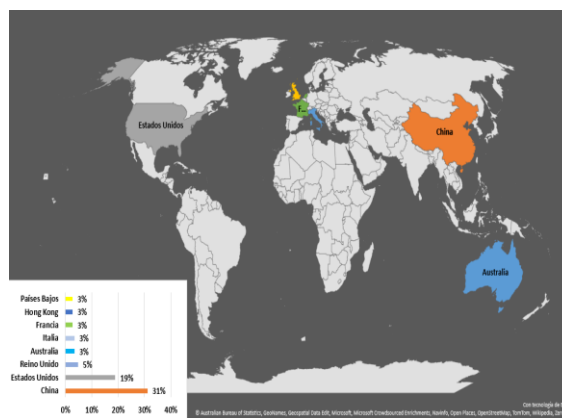


Figura 3. Artículos organizados por país de publicación.

Al examinar detenidamente la información presentada en la figura 4, que representa la categorización de los artículos según sus respectivos campos temáticos, se nota que la gran mayoría de ellos se ubican en el ámbito de las Ciencias Ambientales, alcanzando aproximadamente el 41.9% del total. A continuación, se encuentran representados en proporciones más bajas en los campos de Informática (alrededor del 8.7%), Farmacología, Toxicología y Farmacia (aproximadamente un 7.0%), Ingeniería (cercano al 6.4%), Ciencias de la Tierra y Planetarias (aproximadamente un 5.2%), Medicina (también un 5.2%), Bioquímica, Genética y Biología Molecular (alrededor del 4.7%), Ciencias Agrícolas y Biológicas (aproximadamente un 3.5%), Ciencias Sociales (igualmente alrededor del 3.5%), Química (cercano al 2.9%), Física y Astronomía (alrededor del 2.9%), Ciencias de la Decisión (alrededor del 2.3%), Ingeniería Química (alrededor del 1.7%), Energía (también cerca del 1.7%), Ciencia de los Materiales (cercano al 1.2%) y Matemáticas (igualmente cerca del 1.2%). En este estudio exhaustivo, se llevó a cabo un análisis detenido de la información presentada en la figura 4, la cual ofrece una clasificación de los artículos según sus campos temáticos respectivos. En comparación con la diversidad de temas investigados, se destaca que la mayoría de los artículos se sitúan en el ámbito de las Ciencias Ambientales, representando aproximadamente el 41.9% del total. A diferencia de esta predominancia en Ciencias Ambientales, se observa que otros campos, como Informática (alrededor del 8.7%), Farmacología, Toxicología y Farmacia (aproximadamente un 7.0%), Ingeniería (cercano al 6.4%), y otros, tienen una presencia proporcionalmente menor en la investigación científica abordada en este análisis. También se encontró que, en contraste con la preponderancia de Ciencias Ambientales, hay una distribución más equitativa en áreas como Ciencias Sociales (alrededor del 3.5%), Química (cercano al 2.9%), Física y Astronomía (alrededor del 2.9%), y otras disciplinas, lo que resalta la diversidad temática presente en la producción científica examinada. Este análisis proporciona una visión completa y matizada de la distribución temática en comparación con investigaciones anteriores.

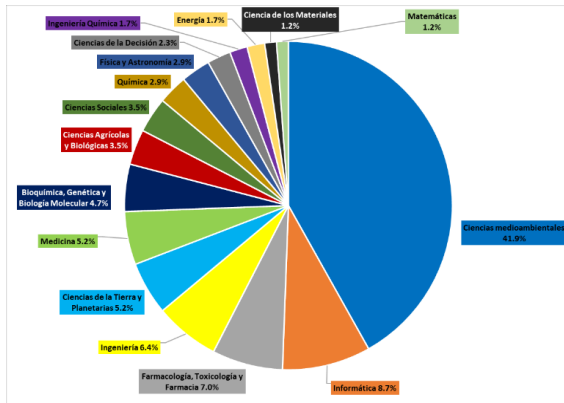


Figura 4. Artículos organizados por área temática.

¿Cuál es la incidencia de la contaminación ambiental en áreas específicas como sitios contaminados, fuentes de contaminación y la presencia de contaminantes en el entorno?

La incidencia de la contaminación ambiental en áreas específicas, como sitios contaminados, fuentes de contaminación y la presencia de contaminantes en el entorno, se manifiesta de manera significativa. Se ha observado que la concentración de partículas PM_{2.5} es elevada en diversas regiones y ciudades, lo que está relacionado con factores como el desarrollo económico rápido y la concentración de industrias, especialmente durante el invierno. La topografía local también influye un impacto en la pureza del aire, intensificando la problemática de la contaminación [12]. Además, se proporciona información detallada sobre la conexión entre la polución atmosférica y la salud de las personas, junto con métodos para evaluar la calidad del aire. Se exploran las consecuencias de la polución atmosférica en el bienestar y salud, considerando factores como los incendios forestales y la prolongada exposición a la polución. Se identifican las fuentes de contaminación, que incluyen emisiones vehiculares e industriales, y se discuten los desafíos relacionados con la caracterización de las exposiciones a la contaminación del aire, tanto a nivel individual como poblacional [13].

En los artículos vemos como destaca la repercusión de eventos extremos como los incendios forestales en la salud regional, señalando cómo estos fenómenos pueden elevar de manera considerable la concentración de contaminantes atmosféricos y afectar la calidad del aire, impactando así la salud de las personas en esas áreas [12]. A parte, la importancia de las herramientas de evaluación de la calidad del aire, empleando métodos avanzados como redes neuronales para estimar la presencia de partículas finas (PM_{2.5}) en áreas específicas. Este análisis detallado resulta esencial para comprender las fuentes de contaminación y evaluar los riesgos para las poblaciones expuestas [13]. Ambos estudios resaltan la necesidad de considerar no solo las fuentes convencionales de contaminación, como las emisiones

vehiculares e industriales, sino también eventos extremos como los incendios forestales, que pueden tener impactos significativos en la calidad del aire y la salud pública. En la tabla III se muestra las referencias que muestran los tipos de incidencias.

TABLA III
ESTUDIOS SOBRE LOS TIPOS DE CONTAMINACIÓN AMBIENTAL

TIPOS DE INCIDENCIAS	ID REFERENCIA.
CONTAMINACIÓN DEL AIRE, CONTAMINACIÓN INDUSTRIAL Y CONTAMINANTES EN EL MEDIO AMBIENTE	[12], [13]

¿Cómo se aplican las técnicas de Aprendizaje automático en la identificación y seguimiento de fuentes de contaminación ambiental en imágenes de teledetección?

Las técnicas de Machine Learning se aplican en diversos contextos para la identificación y seguimiento de fuentes de contaminación en el entorno del medio ambiente en imágenes de teledetección. Estos enfoques contribuyen significativamente a mejorar la comprensión y control de la calidad del aire y del agua, así como a la protección del medio ambiente [14].

En primer lugar, se utilizan modelos de Aprendizaje Automático, como las Redes Neuronales Convolucionales (CNN), para inferir tipos de contaminantes atmosféricos a partir de imágenes de satélite, demostrando la viabilidad de este enfoque [15]. Además, se utiliza un enfoque híbrido que combina información de uso del suelo y datos satelitales para identificar fuentes específicas de contaminantes, utilizando métodos como el análisis de componentes principales y el Random Forest [16].

En otro estudio en China, se emplea el modelo Random Forest para estimar las concentraciones diarias de PM_{2.5} mediante datos de teledetección, como la profundidad óptica de aerosoles y condiciones meteorológicas [17]. Este enfoque demuestra una alta capacidad predictiva y contribuye a una mejor comprensión de la pureza del aire.

Además, en la identificación y seguimiento de fuentes de contaminación en redes de suministro de agua, se utilizan algoritmos de Machine Learning, como las Redes Neuronales Artificiales, para localizar las fuentes de contaminación y analizar variables significativas en eventos de contaminación [2].

En un enfoque espacial-temporal, se desarrolla un modelo interpretable basado en aprendizaje profundo (SIDLM) para estimar concentraciones de PM_{2.5} a partir de datos satelitales, mejorando la precisión predictiva y la capacidad de interpretación [18]. Además, se aplican técnicas de segmentación semántica basada en aprendizaje profundo y clasificación supervisada para evaluar la idoneidad de techos

de edificios para la plantación de vegetación, con el objetivo de reducir la contaminación ambiental [19].

En otros estudios, se utiliza Machine Learning para predecir la concentración de clorofila-a en embalses de agua dulce y estuarios. También se emplean algoritmos como el índice de agua normalizado por diferencia modificada y redes neuronales para evaluar la calidad del agua en diferentes contextos geográficos [20].

Además, se propone un novedoso método de detección automática de posibles vertidos de aguas residuales basado en un modelo de aprendizaje automático, que ha demostrado ser eficaz en la detección y rastrear de fuentes de contaminación ambiental en imágenes de teledetección [21]. Las aplicaciones de Machine Learning en la identificación y seguimiento de fuentes de contaminación en el entorno ambiental en sistemas de IoT, se utilizó modelos de aprendizaje profundo, aprendizaje federado y algoritmos como LSTM y Random Forest [22].

Las técnicas de Machine Learning desempeñan un papel fundamental en la identificación y monitorización de las fuentes de contaminación en el medio ambiente en imágenes de teledetección. Estos enfoques permiten mejorar la comprensión y control de la calidad del aire y del agua, así como contribuir a la protección del medio ambiente [23].

Las redes neuronales convolucionales profundas se aplican para predecir la cantidad de partículas en la atmósfera y analizar patrones complejos en los datos de calidad del aire y las condiciones meteorológicas. Estos modelos aprenden eficientemente de imágenes de satélite y otros datos relacionados, lo que permite estimar la concentración de contaminantes en diferentes ubicaciones [24], [25].

Además, se emplea la clasificación binaria para identificar zonas con alto riesgo de contaminación, utilizando datos espectrales de satélite para detectar la presencia de arsénico y otros contaminantes [26]. El uso de autoencoders permite estimar la concentración de PM_{2.5} en áreas como California, aprovechando datos de teledetección, condiciones meteorológicas y tráfico, este enfoque proporciona una comprensión más precisa de la calidad del aire [27].

Los enfoques modelísticos de Machine Learning, como LASSO, SVR, RF, kNN y XGBoost, se emplean con el propósito de anticipar la cantidad de agentes contaminantes presentes en la atmósfera y se proponen modelos de ensamble basados en algoritmos como RF y GBoost. Esto mejora la precisión de las predicciones [28].

Para abordar la detección de hidrocarburos en suelos contaminados en condiciones tropicales, se recurre al algoritmo Random Forest [29]. Además, el uso de técnicas de ML en drones equipados con cámaras hiperespectrales permite identificar contaminantes en suelo y agua de manera precisa [30].

El aprendizaje automático híbrido se utiliza para predecir la concentración de PM_{2.5} en regiones como Iraq y Kuwait, aprovechando datos de satélite y estaciones de monitoreo

existentes [31]. También, el método de Random Forest se aplica para predecir concentraciones de PM₁₀ en Italia [32].

La aplicación de modelos de Machine Learning, como regresión de vectores de soporte, redes neuronales artificiales y Random Forest, para estimar el contenido de metales pesados en el suelo, proporciona información valiosa sobre la contaminación ambiental [33]. Estas técnicas permiten predecir concentraciones de contaminantes, detectar zonas críticas y respaldar la toma de decisiones en cuestiones de salud pública y protección del medio ambiente, que caracteriza relaciones no lineales y combinar variables de diversas fuentes, mejorando la estimación de la concentración de contaminantes, como en el caso del PM_{2.5} [34], [35].

El Machine Learning se consolida como una técnica emergente en la investigación de la contaminación ambiental, empleando redes neuronales para predecir concentraciones de contaminantes [36]. Además, el aprendizaje profundo se emplea para evaluar los efectos de la contaminación atmosférica en la salud, tal como se ilustra en el contexto del confinamiento por la COVID-19 en India [7].

La detección de hidrocarburos de petróleo en el suelo a través de datos espectrales de teledetección se apoya en algoritmos de Machine Learning [37]. La combinación de teledetección con datos de percepción social y el uso de aprendizaje profundo permite mapear la concentración de PM_{2.5} en áreas urbanas, ilustrando la versatilidad de estas técnicas [38].

Asimismo, en China, se emplean modelos de Machine Learning para estimar la concentración de PM_{2.5} a partir de datos de teledetección satelital, subrayando su utilidad en el control de la pureza del aire [39]. El aprendizaje automático no supervisado se utiliza para identificar emisiones de dióxido de nitrógeno en Chipre, lo que muestra la utilidad de estas técnicas en la identificación de fuentes de contaminación [40].

En el ámbito del agua subterránea, se emplean algoritmos de clasificación y sistemas de visión por computadora con el propósito de reconocer las fuentes de contaminación y mejorar la selección de ubicaciones para pozos de control [41].

Viendo una novedosa metodología donde utiliza algoritmos de Machine Learning y datos de teledetección de alta resolución para predecir los niveles de contaminación del suelo. Esta metodología involucra la extracción de características de los píxeles que representan los puntos de muestreo, que se dividen en tres categorías. Primero ver valores de R, G, B e índices compuestos; segundo, las distancias y gradientes con respecto a características superficiales como vegetación, ríos y fábricas, y por último, las distancias y gradientes en relación a áreas específicas de la fábrica, como zonas de almacenamiento de residuos industriales [42].

Las técnicas de ML desempeñan un papel crucial en la identificación y seguimiento de las fuentes de contaminación ambiental en imágenes de teledetección, se utilizan algoritmos de aprendizaje automático, como redes neuronales profundas,

para la clasificación de datos de teledetección y la detección de derrames de petróleo en entornos marinos [1]. Asimismo, se hacen uso de modelos como LightGBM y XGBoost para predecir las concentraciones de sustancias contaminantes en el aire, aprovechando datos de teledetección, meteorológicos y de uso del suelo, lo que demuestra una notable precisión y eficacia [43].

Un ejemplo interesante es el modelo híbrido de teledetección y Machine Learning llamado RSRF, que utiliza variables de calidad del aire, variables meteorológicas y variables de contaminación del aire para predecir las concentraciones diarias de PM2.5 [44], [45]. Además, se recurre a modelos de aprendizaje profundo para anticipar la polución atmosférica utilizando información meteorológica, de incendios forestales y de temperatura [46]. También, se emplea un modelo de ConvLSTM para pronosticar las cantidades de PM2.5 en la región del condado de Los Ángeles a partir de la fusión de datos obtenidos por satélites de teledetección y observaciones en tierra [47].

Estas técnicas se utilizan igualmente en la anticipación de los niveles de contaminación del aire en varias áreas geográficas, como el empleo del modelo de RF en China [48].

Se emplea un modelo de para anticipar los niveles de polución atmosférica, específicamente las concentraciones de PM2.5 en el Norte de China, basado en datos de teledetección, parámetros meteorológicos, cobertura terrestre y mediciones de PM2.5 en tierra[49]. Además, se aplican al Machine Learning para evaluar las emisiones de vehículos en carretera utilizando sensores remotos[50].

En el ámbito del agua, se construyen modelos de regresión utilizando Machine Learning para predecir la concentración de contaminantes en el agua, como el contenido de arsénico en un sitio de escoria de arsénico [51] o la demanda química de oxígeno (COD) en el agua utilizando una red de cápsulas [52].

También, se utilizan algoritmos como SVM, ANN, y otros para construir modelos de recuperación de la concentración de nitrógeno total en el agua[53]. Estos modelos tienen un valor importante en el monitoreo en tiempo real de la calidad del agua, y, adicionalmente, se emplea el Machine Learning para pronosticar la concentración de PM2.5 en Beijing a partir de imágenes de teledetección [54], [55].

La aplicación exitosa de técnicas como Random Forest para estimar la concentración de PM2.5 y PM10 en distintas áreas geográficas ha demostrado su utilidad al utilizar datos de teledetección y meteorológicos para predecir niveles de polución atmosférica con mayor precisión [21]. Asimismo, se ha empleado el Machine Learning para estimar concentraciones de contaminantes en cuerpos de agua, mejorando la evaluación de la calidad del agua y su monitoreo en tiempo real, especialmente la concentración de clorofila-a en embalses de agua dulce [20].

Por otro lado, se han aprovechado las técnicas de Machine Learning junto con datos de teledetección para

estimar la calidad del aire en áreas urbanas, mejorando la evaluación de zonas con alto riesgo de contaminación [15]. Además, se han identificado emisiones de dióxido de nitrógeno, lo cual representa un paso crucial para el control y la prevención de la contaminación[40].

En el análisis del impacto de la contaminación atmosférica en la salud durante el confinamiento por la COVID-19, se han utilizado técnicas de Machine Learning para evaluar correlaciones entre la contaminación del aire y la salud de la población [24]. También, se han empleado modelos interpretables basados en aprendizaje profundo para estimar concentraciones de PM2.5 y otros agentes contaminantes en la atmósfera [18], [28].

La aplicabilidad de técnicas de Machine Learning para predecir la contaminación atmosférica en distintos contextos geográficos, desde áreas específicas como el condado de Los Ángeles hasta regiones tan vastas como China, se ha demostrado efectiva [39]. Del mismo modo, se han utilizado técnicas de aprendizaje automático para predecir la concentración de metales pesados en el suelo, brindando información crucial para la protección del medio ambiente y la salud pública [33].

Finalmente, la combinación de modelos de ensamble y técnicas de aprendizaje profundo ha contribuido a mejorar la precisión en la estimación de la contaminación del aire. Se han identificado hidrocarburos en suelos contaminados utilizando Random Forest, lo cual destaca la eficacia de las herramientas de Machine Learning en la identificación de focos de contaminación [37], [43]. Estos diversos enfoques y su aplicación integral ilustran el impacto significativo del Machine Learning en la identificación, prevención y gestión de fuentes de contaminación ambiental. La Tabla IV muestra las referencias por tipos de técnicas de Machine Learning.

TABLA IV
ESTUDIOS SOBRE LAS TÉCNICAS DE MACHINE LEARNING

TIPOS DE TÉCNICAS	ID REFERENCIA.
DEEP LEARNING, NEURAL NETWORKS	[17]
MACHINE LEARNING	[2], [7], [16],[18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [50], [51], [52], [53], [54], [55]
DEEP LEARNING, LSTM, RANDOM FOREST	[14], [15],[49]

DEEP LEARNING, LIGHTGBM, XGBOOST	[1]
--	-----

¿Cuáles son los métodos y resultados más comunes utilizados para la identificación y seguimiento de fuentes de contaminación ambiental en imágenes de teledetección, incluyendo la identificación de fuentes de contaminación, el seguimiento de fuentes contaminantes y la evaluación de la contaminación ambiental?

Los métodos y resultados más comunes empleados para identificar y rastrear fuentes de contaminación ambiental en imágenes de teledetección involucran una amplia gama de enfoques. Estos métodos se valen de la capacidad de la teledetección para recopilar datos acerca de la pureza del aire, del agua y otros indicadores ambientales. Se utilizan imágenes de teledetección para identificar fuentes de contaminación al monitorear extensas áreas geográficas y evaluando alteraciones en la utilización del terreno que puedan afectar la liberación de sustancias contaminantes[56]. También se aplican técnicas de interpolación espacial y kriging espaciotemporal para estimar la distribución de la contaminación del aire [57].

En lo que respecta a la evaluación a largo plazo de la exposición a contaminantes atmosféricos, se recurre a enfoques como la modelización de dispersión, la interpolación y la modelización de regresión basada en el uso del suelo, aprovechando los datos de óptica de aerosoles recuperados por satélites [3].

La detección y rastreo de las fuentes de contaminación del entorno implican la utilización de modelos tanto físicos como empíricos que simulan la interacción de la luz con la columna de agua, lo que permite detectar cambios en la calidad del agua y patrones de contaminación [58]. El uso de datos satelitales, como los proporcionados por el satélite Sentinel-5 Precursor, resulta fundamental para observar componentes atmosféricos como el NO₂ [59]. Además, la integración de técnicas de aprendizaje automático y datos relacionados con la superficie terrestre, el clima, la población y las características geográficas permite la identificación de fuentes contaminantes y la evaluación de la contaminación ambiental[60].

Los modelos de aprendizaje automático se aplican para predecir las cantidades de sustancias contaminantes en la superficie terrestre, y se recurre a métodos de interpolación que se fundamentan en la ubicación de estaciones de control de la contaminación. Estos métodos permiten tanto la identificación de fuentes de contaminación como el seguimiento de estas, además de evaluar la contaminación ambiental en la región [61].

Destacamos la utilidad de la profundidad óptica de aerosoles (AOD) recuperada mediante teledetección como una valiosa herramienta para el monitoreo ambiental atmosférico en áreas extensas. Asimismo, se sugiere que la evaluación de

la contaminación ambiental puede mejorarse a través del uso de datos de teledetección para comprender el impacto de la exposición a PM_{2.5} en la población y para formular recomendaciones de políticas relacionadas con el control de la contaminación del [62].

Cuando se trata de supervisar la calidad del agua, se utilizan una variedad de métodos empíricos, estadísticos, semi-analíticos y de ML, incluyendo redes neuronales convolucionales, para construir modelos de recuperación de parámetros de calidad del agua. Estos métodos permiten detectar cambios en la calidad del agua y realizar un monitoreo avanzado de los recursos terrestres y los cambios ambientales [63].

Adicionalmente, en el caso de la inversión de los niveles de partículas finas PM_{2.5}, se recurre a datos de teledetección satelital y a una variedad de métodos estadísticos, como fórmulas semi-empíricas, regresión múltiple, regresión geográficamente ponderada y redes neuronales[64].

Los resultados más comunes utilizados para la identificación y rastreo de fuentes de contaminación ambiental en imágenes de teledetección abarcan desde enfoques físicos y empíricos hasta técnicas de aprendizaje automático, combinando datos satelitales, climáticos y sociodemográficos para evaluar la contaminación ambiental y el gran impacto que tiene en el medio ambiente y la salud [65].

Los estudios recientes han demostrado la eficacia de diversas estrategias. La modelización de dispersión y la interpolación espacial-temporal se han utilizado para estimar la distribución de la contaminación del aire, aprovechando datos de óptica de aerosoles obtenidos por satélites [3]. Además, el uso de datos satelitales, como los proporcionados por el satélite Sentinel-5 Precursor, resulta crucial para observar componentes atmosféricos como el NO₂ [52].

La integración de datos y técnicas de aprendizaje automático ha permitido identificar fuentes contaminantes y evaluar la contaminación ambiental [60]. Los modelos de aprendizaje automático se aplican para predecir cantidades de sustancias contaminantes en la superficie terrestre [61].

También, se ha destacado la utilidad de la profundidad óptica de aerosoles (AOD) recuperada mediante teledetección para el monitoreo ambiental atmosférico en áreas extensas [62].

Además, técnicas como las redes neuronales convolucionales se utilizan para construir modelos de recuperación de parámetros de calidad del agua y predecir la concentración de contaminantes [63]. El análisis de datos de teledetección y modelos basados en imágenes ha aplicado técnicas de aprendizaje profundo para estimar la contaminación atmosférica y los niveles de PM_{2.5} utilizando múltiples fuentes de datos.

Esta combinación de métodos ofrece soluciones innovadoras que contribuyen a un mejor entendimiento y control de la contaminación en el medio ambiente, facilitando la gestión y toma de decisiones relacionadas con la protección

del entorno y la salud pública. La tabla V muestra las referencias bibliográficas por tipo de resultados.

TABLA V
REFERENCIAS POR TIPO DE RESULTADOS

TIPOS DE RESULTADOS	ID REFERENCIA.
EFFECTIVO	[3], [54], [56], [58], [59], [57], [60], [61], [62], [63], [64], [65]

V CONCLUSIONES

Tras el análisis se observó una clara preeminencia de los algoritmos de Aprendizaje Profundo, especialmente las Redes Neuronales Convolucionales (CNN), Random Forest (RF), Máquinas de Vectores de Soporte (SVM) y Redes Neuronales Recurrentes (RNN), destacándose por su efectividad en la detección y seguimiento de fuentes de contaminación en imágenes de teledetección. Estos métodos avanzados han demostrado ser altamente eficaces en la identificación de patrones complejos, tanto espaciales como temporales, evidenciando su capacidad para abordar la complejidad de los datos multidimensionales. Si bien se reconoce la utilidad de métodos más tradicionales como la Regresión Lineal (LR) y el Análisis de Componentes Principales (PCA) en situaciones específicas, la predominancia y efectividad de los enfoques de Aprendizaje Profundo resaltan la necesidad de estrategias personalizadas y específicas para enfrentar los desafíos variados en la detección de contaminantes en entornos ambientales mediante algoritmos de Aprendizaje Automático.

REFERENCIAS

[1] Y. Li, Y. Zhang, Z. Yuan, H. Guo, H. Pan, and J. Guo, "Marine oil spill detection based on the comprehensive use of polarimetric SAR data," *Sustainability (Switzerland)*, vol. 10, no. 12, Nov. 2018, doi: 10.3390/su10124408.

[2] L. Grbčić, I. Lučin, L. Kranjčević, and S. Družeta, "A machine learning-based algorithm for water network contamination source localization," *Sensors (Switzerland)*, vol. 20, no. 9, May 2020, doi: 10.3390/s20092613.

[3] O. O. Arowosegbe et al., "Ensemble averaging using remote sensing data to model spatiotemporal PM10 concentrations in sparsely monitored South Africa," *Environmental Pollution*, vol. 310, Oct. 2022, doi: 10.1016/j.envpol.2022.119883.

[4] H. Onyeaka, C. K. Anumudu, Z. T. Al-Sharify, E. Egele-Godswill, and P. Mbaegbu, "COVID-19 pandemic: A review of the global lockdown and its far-reaching effects," *Science Progress*, vol. 104, no. 2. SAGE Publications Ltd, 2021. doi:10.1177/00368504211019854.

[5] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," *Frontiers in Public Health*, vol. 8. Frontiers Media S.A., Feb. 20, 2020. doi: 10.3389/fpubh.2020.00014.

[6] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018. Hindawi Limited, 2018. doi: 10.1155/2018/7068349.

[7] S. K. Chaturvedi, "Health Impact Assessment of Air Pollution in India During COVID-19 Lockdown by Using Satellite Remote Sensing and Deep Learning," *Nature Environment and Pollution Technology*, vol. 21, no. 4, pp. 1767–1774, Dec. 2022, doi: 10.46488/NEPT.2022.v21i04.032.

[8] A. Nishikawa-Pacher, "Research Questions with PICO: A Universal Mnemonic," *Publications*, vol. 10, no. 3, Sep. 2022, doi: 10.3390/publications10030021.

[9] C. Arienti, S. G. Lazzarini, M. Patrini, L. Puljak, A. Pollock, and S. Negrini, "The Structure of Research Questions in Randomized Controlled Trials in the Rehabilitation Field: A Methodological Study," *Am J Phys Med Rehabil*, vol. 100, no. 1, pp. 29–33, Jan. 2021, doi: 10.1097/PHM.0000000000001612.

[10] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Syst Rev*, vol. 10, no. 1, Dec. 2021, doi: 10.1186/s13643-021-01626-4.

[11] R. E. O’Dea et al., "Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension," *Biological Reviews*, vol. 96, no. 5, pp. 1695–1722, Oct. 2021, doi: 10.1111/brv.12721.

[12] S. M. O’Neill et al., "A multi-analysis approach for estimating regional health impacts from the 2017 Northern California wildfires," *J Air Waste Manage Assoc*, vol. 71, no. 7, pp. 791–814, 2021, doi: 10.1080/10962247.2021.1891994.

[13] H. Wu et al., "Prediction of Fine Particulate Matter Concentration near the Ground in North China from Multivariable Remote Sensing Data Based on MIV-BP Neural Network," *Atmosphere (Basel)*, vol. 13, no. 5, May 2022, doi: 10.3390/atmos13050825.

[14] J. Li, J. Wang, Y. Wu, Y. Cui, and S. Yan, "Remote sensing monitoring of total nitrogen and total phosphorus concentrations in the water around Chaohu Lake based on geographical division," *Front Environ Sci*, vol. 10, Oct. 2022, doi:10.3389/fenvs.2022.1014155.

[15] M. Sorek-Hamer et al., "A Deep Learning Approach for Meter-Scale Air Quality Estimation in Urban Environments Using Very High- Spatial-Resolution Satellite Imagery," *Atmosphere (Basel)*, vol. 13, no. 5, May 2022, doi: 10.3390/atmos13050696.

- [16] M. M. Rahman and G. Thurston, "A hybrid satellite and land use regression model of source-specific PM_{2.5} and PM_{2.5} constituents," *Environ Int*, vol. 163, May 2022, doi: 10.1016/j.envint.2022.107233.
- [17] G. Chen et al., "A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information," *Science of the Total Environment*, vol. 636, pp. 52–60, Sep. 2018, doi: 10.1016/j.scitotenv.2018.04.251.
- [18] X. Yan et al., "A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM_{2.5}," *Environmental Pollution*, vol. 273, Mar. 2021, doi: 10.1016/j.envpol.2021.116459.
- [19] N. Xu et al., "Accurate suitability evaluation of large-scale roof greening based on RS and GIS methods," *Sustainability (Switzerland)*, vol. 12, no. 11, Jun. 2020, doi: 10.3390/su12114375.
- [20] Q. Guo, X. Wu, X. Sang, Y. Fu, Y. Zang, and X. Gong, "An integrated study on change detection and environment evaluation of surface water," *Appl Water Sci*, vol. 10, no. 1, Jan. 2020, doi: 10.1007/s13201-019-1109-3.
- [21] Y. Wang et al., "Automatic detection of suspected sewage discharge from coastal outfalls based on Sentinel-2 imagery," *Science of the Total Environment*, vol. 853, Dec. 2022, doi:10.1016/j.scitotenv.2022.158374.
- [22] S. Kimothi et al., "Big Data Analysis Framework for Water Quality Indicators with Assimilation of IoT and ML," *Electronics (Switzerland)*, vol. 11, no. 13, Jul. 2022, doi:10.3390/electronics11131927.
- [23] Q. Zhao, X. Dong, G. Li, Y. Jin, X. Yang, and Y. Qu, "Classification and Regression Tree Models for Remote Recognition of Black and Odorous Water Bodies Based on Sensor Networks," *Sci Program*, vol. 2022, 2022, doi: 10.1155/2022/7390098.
- [24] D. Pala et al., "Deep learning to unveil correlations between urban landscape and population health," *Sensors (Switzerland)*, vol. 20, no.7, Apr. 2020, doi: 10.3390/s20072105.
- [25] L. Zang, F. Mao, J. Guo, W. Gong, W. Wang, and Z. Pan, "Estimating hourly PM₁ concentrations from Himawari-8 aerosol optical depth in China," *Environmental Pollution*, vol. 241, pp. 654–663, Oct. 2018, doi: 10.1016/j.envpol.2018.05.100.
- [26] A. Agrawal and M. R. Petersen, "Detecting arsenic contamination using satellite imagery and machine learning," *Toxics*, vol. 9, no. 12, Dec. 2021, doi: 10.3390/toxics9120333.
- [27] L. Li et al., "Ensemble-based deep learning for estimating PM_{2.5} over California with multisource big data including wildfire smoke," *Environ Int*, vol. 145, Dec. 2020, doi: 10.1016/j.envint.2020.106143.
- [28] J. Gao, J. Liu, R. Xu, S. Pandey, V. S. K. S. V. Siva, and D. Yu, "Environmental Pollution Analysis and Impact Study-A Case Study for the Salton Sea in California," *Atmosphere (Basel)*, vol. 13, no. 6, Jun. 2022, doi: 10.3390/atmos13060914.
- [29] G. Lassalle et al., "Estimating persistent oil contamination in tropical region using vegetation indices and random forest regression," *Ecotoxicol Environ Saf*, vol. 184, Nov. 2019, doi:10.1016/j.ecoenv.2019.109654.
- [30] L. Wei, Y. Zhang, Q. Lu, Z. Yuan, H. Li, and Q. Huang, "Estimating the spatial distribution of soil total arsenic in the suspected contaminated area using UAV-Borne hyperspectral imagery and deep learning," *Ecol Indic*, vol. 133, Dec. 2021, doi:10.1016/j.ecolind.2021.108384.
- [31] J. Li et al., "Estimation of ambient PM_{2.5} in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing," *Environ Int*, vol. 151, Jun. 2021, doi: 10.1016/j.envint.2021.106445.
- [32] M. Stafoggia et al., "Estimation of daily PM₁₀ and PM_{2.5} concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model," *Environ Int*, vol. 124, pp. 170–179, Mar. 2019, doi: 10.1016/j.envint.2019.01.016.
- [33] S. Shi et al., "Estimation of Heavy Metal Content in Soil Based on Machine Learning Models," *Land (Basel)*, vol. 11, no. 7, Jul. 2022, doi: 10.3390/land11071037.
- [34] Y. Xu et al., "Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM_{2.5}," *Environmental Pollution*, vol. 242, pp. 1417–1426, Nov. 2018, doi:10.1016/j.envpol.2018.08.029.
- [35] J. Li, M. Jin, and H. Li, "Exploring spatial influence of remotely sensed PM_{2.5} concentration using a developed deep convolutional neural network model," *Int J Environ Res Public Health*, vol. 16, no. 3, Jan. 2019, doi: 10.3390/ijerph16030454.
- [36] L. Yin et al., "Haze Grading Using the Convolutional Neural Networks," *Atmosphere (Basel)*, vol. 13, no. 4, Apr. 2022, doi: 10.3390/atmos13040522.
- [37] M. Wu, N. Lin, G. Li, H. Liu, and D. Li, "Hyperspectral estimation of petroleum hydrocarbon content in soil using ensemble learning method and LASSO feature extraction," *Environmental Pollutants and Bioavailability*, vol. 34, no. 1, pp. 308–320, 2022, doi: 10.1080/26395940.2022.2102543.
- [38] H. Shen, M. Zhou, T. Li, and C. Zeng, "Integration of remote sensing and social sensing data in a deep learning framework for hourly urban PM_{2.5} mapping," *Int J Environ Res Public Health*, vol. 16, no. 21, Nov. 2019, doi: 10.3390/ijerph16214102.
- [39] L. Dong, S. Li, J. Yang, W. Shi, and L. Zhang, "Investigating the performance of satellite-based models in estimating the surface PM_{2.5} over China,"

- Chemosphere, vol. 256, Oct. 2020, doi: 10.1016/j.chemosphere.2020.127051.
- [40] M. Vrekoussis et al., “Local and regional air pollution characteristics in Cyprus: A long-term trace gases observations analysis,” *Science of the Total Environment*, vol. 845, Nov. 2022, doi:10.1016/j.scitotenv.2022.157315.
- [41] Y. N. Kontos, T. Kassandra, K. Perifanos, M. Karampasis, K. L. Katsifarakis, and K. Karatzas, “Machine learning for groundwater pollution source identification and monitoring network optimization,” *Neural Comput Appl*, vol. 34, no. 22, pp. 19515–19545, Nov. 2022, doi: 10.1007/s00521-022-07507-8.
- [42] X. Jia et al., “Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field,” *Environmental Pollution*, vol. 270, Feb. 2021, doi: 10.1016/j.envpol.2020.116281.
- [43] J. Ma, R. Zhang, J. Xu, and Z. Yu, “MERRA-2 PM2.5 mass concentration reconstruction in China mainland based on LightGBM machine learning,” *Science of the Total Environment*, vol. 827, Jun. 2022, doi: 10.1016/j.scitotenv.2022.154363.
- [44] M. Z. Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, “PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data,” *Atmosphere (Basel)*, vol. 10, no. 7, Jul. 2019, doi: 10.3390/atmos10070373.
- [45] X. Li and X. Zhang, “Predicting ground-level PM2.5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning approach,” *Environmental Pollution*, vol. 249, pp. 735–749, Jun. 2019, doi: 10.1016/j.envpol.2019.03.068.
- [46] P. Muthukumar et al., “PM2.5 Air Pollution Prediction through Deep Learning Using Multisource Meteorological, Wildfire, and Heat Data,” *Atmosphere (Basel)*, vol. 13, no. 5, May 2022, doi: 10.3390/atmos13050822.
- [47] P. Muthukumar et al., “Predicting PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data,” *Air Qual Atmos Health*, vol. 15, no. 7, pp. 1221–1234, Jul. 2022, doi:10.1007/s11869-021-01126-3.
- [48] K. Huang et al., “Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain,” *Environmental Pollution*, vol. 242, pp. 675–683, Nov. 2018, doi: 10.1016/j.envpol.2018.07.016.
- [49] I. Gutiérrez-Avila et al., “Prediction of daily mean and one-hour maximum PM2.5 concentrations and applications in Central Mexico using satellite-based machine-learning models,” *J Expo Sci Environ Epidemiol*, vol. 32, no. 6, pp. 917–925, Nov. 2022, doi: 10.1038/s41370-022-00471-4.
- [50] Y. Xia et al., “Rapid assessments of light-duty gasoline vehicle emissions using on-road remote sensing and machine learning,” *Science of the Total Environment*, vol. 815, Apr. 2022, doi: 10.1016/j.scitotenv.2021.152771.
- [51] G. Liu et al., “Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data,” *Environmental Pollution*, vol. 267, Dec. 2020, doi: 10.1016/j.envpol.2020.115631.
- [52] C. Deng, L. Zhang, and Y. Cen, “Retrieval of chemical oxygen demand through modified capsule network based on hyperspectral data,” *Applied Sciences (Switzerland)*, vol. 9, no. 21, Nov. 2019, doi: 10.3390/app9214620.
- [53] J. Qun’ou, X. Lidan, S. Siyang, W. Meilin, and X. Huijie, “Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms – A case study in the Miyun Reservoir, China,” *Ecol Indic*, vol. 124, May 2021, doi:10.1016/j.ecolind.2021.107356.
- [54] L. Yin, L. Wang, W. Huang, S. Liu, B. Yang, and W. Zheng, “Spatiotemporal analysis of haze in Beijing based on the multi-convolution model,” *Atmosphere (Basel)*, vol. 12, no. 11, Nov. 2021, doi: 10.3390/atmos12111408.
- [55] H. Dai, G. Huang, J. Wang, H. Zeng, and F. Zhou, “Spatio-Temporal Characteristics of PM2.5 Concentrations in China Based on Multiple Sources of Data and LUR-GBM during 2016–2021,” *Int J Environ Res Public Health*, vol. 19, no. 10, May 2022, doi:10.3390/ijerph19106292.
- [56] Y. Lin et al., “An optimized machine learning approach to water pollution variation monitoring with time-series Landsat images,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, Oct. 2021, doi: 10.1016/j.jag.2021.102370.
- [57] M. Yu and Q. Liu, “Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations,” *Science of the Total Environment*, vol. 773, Jun. 2021, doi: 10.1016/j.scitotenv.2021.145145.
- [58] B. Fu et al., “Evaluating optically and non-optically active water quality and its response relationship to hydro-meteorology using multi-source data in Poyang Lake, China,” *Ecol Indic*, vol. 145, Dec. 2022, doi: 10.1016/j.ecolind.2022.109675.
- [59] B. Chen, Z. Song, F. Pan, and Y. Huang, “Obtaining vertical distribution of PM2.5 from CALIOP data and machine learning algorithms,” *Science of the Total Environment*, vol. 805, Jan. 2022, doi: 10.1016/j.scitotenv.2021.150338.
- [60] J. Wei et al., “Ground-Level NO2 Surveillance from Space Across China for High Resolution Using

- Interpretable Spatiotemporally Weighted Artificial Intelligence,” *Environ Sci Technol*, vol. 56, no. 14, pp. 9988–9998, Jul. 2022, doi: 10.1021/acs.est.2c03834.
- [61] M. S. Rahman et al., “How air quality and COVID-19 transmission change under different lockdown scenarios? A case from Dhaka city, Bangladesh,” *Science of the Total Environment*, vol. 762, Mar. 2021, doi: 10.1016/j.scitotenv.2020.143161.
- [62] N. Amoroso, R. Cilli, T. Maggipinto, A. Monaco, S. Tangaro, and R. Bellotti, “Satellite data and machine learning reveal a significant correlation between NO₂ and COVID-19 mortality,” *Environ Res*, vol. 204, Mar. 2022, doi: 10.1016/j.envres.2021.111970.
- [63] Y. Chi et al., “Machine learning-based estimation of ground-level NO₂ concentrations over China,” *Science of the Total Environment*, vol. 807, Feb. 2022, doi: 10.1016/j.scitotenv.2021.150721.
- [64] O. Haffner, E. Kučera, P. Drahoš, J. Cigánek, A. Kozáková, and B. Urminská, “Lemna minor bioassay evaluation using computer image analysis,” *Water (Switzerland)*, vol. 12, no. 8, Aug. 2020, doi: 10.3390/w12082207.
- [65] B. Chen et al., “Spatiotemporal dynamics and exposure analysis of daily PM_{2.5} using a remote sensing-based machine learning model and multi-time meteorological parameters,” *Atmos Pollut Res*, vol. 12, no. 2, pp. 23–31, Feb. 2021, doi: 10.1016/j.apr.2020.10.005.