

Aplicación de la inteligencia artificial en el monitoreo ambiental: modelos predictivos de elementos potencialmente tóxicos (PTEs) mediante machine learning

Artificial intelligence applications in environmental monitoring: machine learning-based prediction of potentially toxic elements (PTEs)

Teresa Salazar-Rojas¹, Guillermo Calvo-Brenes²

Salazar-Rojas, T; Calvo-Brenes, G. Aplicación de la inteligencia artificial en el monitoreo ambiental: modelos predictivos de elementos potencialmente tóxicos (PTEs) mediante machine learning. *Tecnología en Marcha*. Vol. 39 N° especial sobre Inteligencia Artificial. Febrero, 2026. Pág. 250-263.

 <https://doi.org/10.18845/tm.v39i5.8507>



- 1 Chemistry School, Costa Rica Institute of Technology. Costa Rica.
 tsalazar@itcr.ac.cr
 <https://orcid.org/00000000223663638>
- 2 Chemistry School, Costa Rica Institute of Technology. Costa Rica.
 gcalvo@itcr.ac.cr
 <https://orcid.org/0000-0002-7021-3509>

Palabras claves

Máquina de soporte de vectores (SVM); elementos potencialmente tóxicos (PTEs); propiedades magnéticas; plantas; polvo de carretera; machine learning.

Resumen

La inteligencia artificial (IA), y en particular el aprendizaje automático (*machine learning*), ha emergido como una herramienta clave para mejorar el monitoreo ambiental, permitiendo desarrollar modelos predictivos capaces de estimar contaminantes a partir de variables fácilmente medibles. Este estudio aplicó técnicas de aprendizaje automático, específicamente Máquina de Vectores de Soporte (SVM), para predecir concentraciones de elementos potencialmente tóxicos (PTEs) en biomonitores vegetales y polvo de carretera de la Gran Área Metropolitana de Costa Rica. Se recolectaron 160 muestras de hojas de *Casuarina equisetifolia* y *Cupressus lusitanica*, junto con 80 muestras de polvo de carretera, durante campañas en 2020 y 2021. Las muestras se analizaron para determinar propiedades magnéticas y concentraciones de metales (Fe, Cu, Cr, Ni, Pb, V, Zn y Cd). Los modelos SVM, comparados con regresiones lineales simples y múltiples, mostraron mejor desempeño predictivo en la mayoría de los metales, con valores de R^2 superiores a 0.7 para Fe, Cu, Cr, V y Zn, y menores para Pb y Ni. *C. equisetifolia* presentó los mejores resultados en términos de ajuste (R^2) y menor error (RMSE, MAE), superando a *C. lusitanica*; el polvo de carretera, por su parte, mostró una mayor variabilidad debido a sus múltiples fuentes de contaminación. Los resultados demostraron que la integración de la IA en el monitoreo biomagnético facilita la identificación de relaciones complejas entre las propiedades magnéticas y la contaminación por PTEs, confirmando su eficacia como herramienta rápida y de bajo costo para la evaluación ambiental.

Keywords

Support Vector Machine (SVM); PTEs; magnetic properties; plants; street dust; machine learning.

Abstract

Artificial intelligence (AI), and particularly Machine Learning, has emerged as a key tool for improving environmental monitoring, enabling the development of predictive models capable of estimating pollutants from easily measurable variables. This study applied machine learning techniques, specifically Support Vector Machine (SVM), to predict the concentrations of potentially toxic elements (PTEs) in plant biomonitors and road dust from the Greater Metropolitan Area of Costa Rica. A total of 160 leaf samples from *Casuarina equisetifolia* and *Cupressus lusitanica*, along with 80 road dust samples, were collected during field campaigns in 2020 and 2021. The samples were analyzed to determine their magnetic properties and metal concentrations (Fe, Cu, Cr, Ni, Pb, V, Zn, and Cd). The SVM models, compared to simple and multiple linear regressions, showed better predictive performance for most metals, with R^2 values above 0.7 for Fe, Cu, Cr, V, and Zn, and lower values for Pb and Ni. *C. equisetifolia* exhibited the best performance in terms of fit (R^2) and lower error (RMSE, MAE), surpassing *C. lusitanica*. In contrast, road dust showed greater variability associated with its multiple pollution sources. The results demonstrated that the integration of AI into biomagnetic monitoring facilitates the identification of complex relationships between magnetic properties and PTEs pollution, confirming its effectiveness as a fast and low-cost tool for environmental assessment.

Introducción

La inteligencia artificial (IA) se ha posicionado como una de las tecnologías más transformadoras del siglo XXI, con un impacto transversal en campos tan diversos como la ciencia, la salud, la educación, la industria y la gestión ambiental [1], [2]. Su capacidad para procesar grandes volúmenes de datos, reconocer patrones complejos y generar predicciones precisas ha revolucionado la forma en que comprendemos y gestionamos los sistemas naturales y antrópicos [3]. Los recientes avances tecnológicos en IA han impulsado un creciente interés en su aplicación en el campo de la detección y monitoreo de contaminantes ambientales, ampliando las posibilidades de análisis y gestión de la calidad ambiental [1], [4].

Como rama de la informática, la IA se centra en el desarrollo de algoritmos y sistemas computacionales capaces de realizar actividades como la percepción, el razonamiento, el aprendizaje y la toma de decisiones, que tradicionalmente requerían de la inteligencia humana [5], [6]. Los avances en aprendizaje automático (machine learning, ML) —una de las ramas más desarrolladas de la IA— han permitido crear modelos predictivos que integran datos provenientes de sensores, satélites, imágenes, variables meteorológicas y registros de emisiones, superando limitaciones de cobertura espacial, temporal y costo [6], [7], [8], [9].

Por otra parte, el monitoreo biomagnético ha surgido rápidamente como una herramienta valiosa para la evaluación de la contaminación atmosférica urbana debido a su alta resolución espacial, complementando los sistemas de monitoreo tradicionales. Con el rápido avance de la urbanización, el problema de la contaminación por material particulado (PM) en las ciudades se ha vuelto cada vez más crítico. La relación entre la exposición prolongada al $PM_{2.5}$ en exteriores y el aumento de la mortalidad prematura ha sido ampliamente documentada. El $PM_{2.5}$, debido a su tamaño diminuto, puede transportar una variedad de sustancias tóxicas y peligrosas, entre ellas elementos potencialmente tóxicos (PTEs) como algunos metales [10], [11], [12] representando un riesgo significativo para la salud humana y las plantas. El magnetismo ambiental es cada vez más preferido por los investigadores que estudian cambios ambientales modernos, gracias a sus capacidades de análisis rápido, su carácter no contaminante y su compatibilidad ambiental [13], [14]. En este contexto, el monitoreo biomagnético, que analiza las propiedades magnéticas de los contaminantes acumulados pasivamente en organismos vegetales, ha emergido como un enfoque prometedor, dado que existe una correlación entre estas propiedades magnéticas y el contenido de metales pesados [15], [16].

En este marco, nuestro trabajo emplea técnicas de aprendizaje automático (machine learning) como herramientas predictivas para estimar concentraciones de elementos potencialmente tóxicos (PTEs) en matrices ambientales. Específicamente, se implementaron modelos de Máquinas de Vectores de Soporte (SVM) para capturar relaciones complejas entre las variables de entrada y los niveles de contaminación. A través del uso de modelos de SVM demostramos cómo los algoritmos de IA pueden capturar relaciones complejas entre variables físicas fácilmente medibles y la contaminación atmosférica por metales, generando modelos robustos, rápidos y de bajo costo operativo. Este enfoque se plantea como un método de cribado ambiental que complementa los sistemas tradicionales de monitoreo, expandiendo las posibilidades de vigilancia continua de contaminantes en regiones urbanas y periurbanas.

La integración de IA en el monitoreo ambiental representa, por tanto, una evolución hacia sistemas más inteligentes, eficientes y sostenibles. Su adopción responsable puede facilitar la transición hacia políticas públicas basadas en datos, fortaleciendo la capacidad de respuesta frente a la contaminación y mejorando la salud ambiental y humana.

Metodología

Sitio de Muestreo

El sitio de muestreo fue la Gran Área Metropolitana (GAM) de Costa Rica, la cual debe su importancia a que aunque abarca el 3,8% del territorio nacional, el 70% del parque vehicular, el 62% de la población, el 65% del parque empresarial formal y el 82% de las ventas [17].

Los sitios de muestreo, Fig. 1, se seleccionaron considerando la densidad vehicular obtenida del Anuario de Información de Tránsito 2019 [18]. Se seleccionaron ocho sitios de muestreo para cada monitor y, además, se seleccionaron cuatro puntos de muestreo adicionales para validar los resultados de los modelos (Cuadro 1).

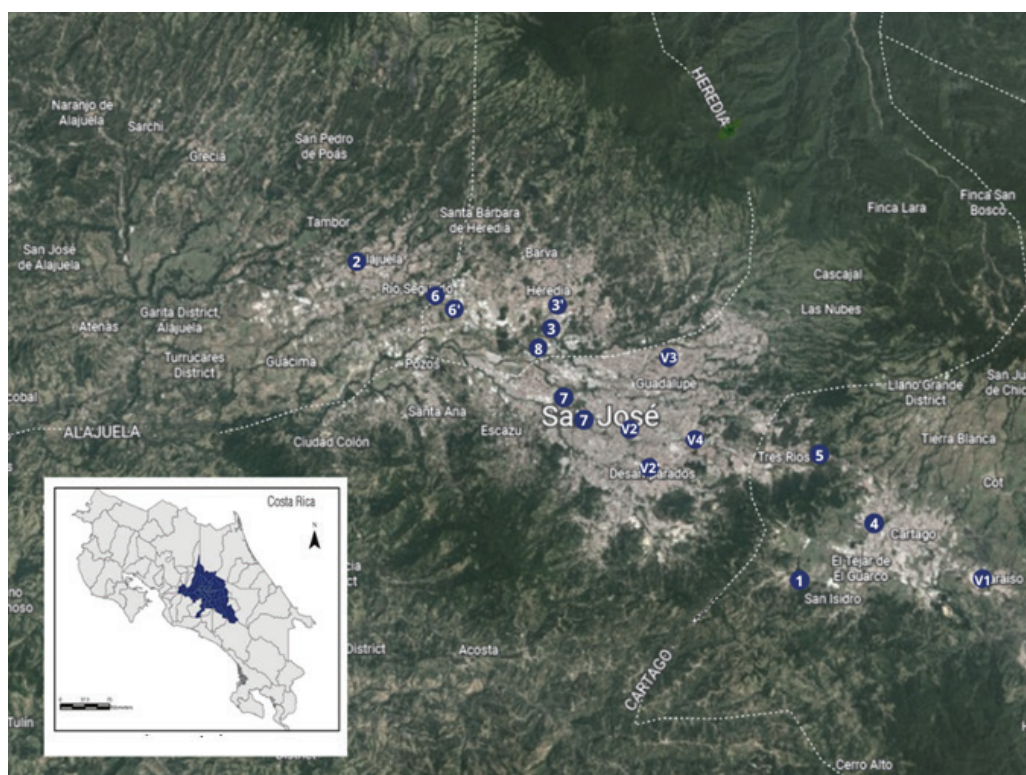


Figura 1. Mapa del sitio de estudio.

Diseño de muestreo

La muestra de estudio estuvo compuesta por un total de 160 muestras de hojas de plantas y 80 muestras de polvo de carretera, recolectados durante dos campañas realizadas entre febrero y marzo, una en 2020 y otra en 2021.

Para las plantas se tomó follaje maduro del dosel lateral orientado hacia el costado de la carretera, entre 1 y 5 plantas por sitio, a una altura de 1,5 a 2,0 m, considerando la altura normal de respiración humana y con el fin de evitar la influencia de partículas del suelo urbano [19]. Las mismas se almacenaron en bolsas plásticas y se trasladaron en una hielera. En el laboratorio, las muestras se secaron a $(55 \pm 1)^\circ\text{C}$ hasta alcanzar peso constante y luego se molieron para su posterior análisis en el laboratorio, siguiendo a [20]. Las especies de plantas se seleccionaron después de realizar un censo en enero de 2020, considerando criterios de abundancia, accesibilidad, longevidad, proximidad a la carretera y resultados obtenidos en estudios

previos. Los bioindicadores seleccionados fueron *Cupressus lusitanica* (Cupressaceae) [21] y *Casuarina equisetifolia* (Casuarinaceae) [22], ambas especies perennes, con una duración del follaje de 2 años o más (Cuadro 1).

Las muestras de polvo de carretera se recolectaron en los mismos sitios y durante el mismo período de muestreo que las plantas. Las muestras se tomaron directamente sobre la carretera, realizando un barrido de aproximadamente un metro cuadrado en el borde de la vía. Las muestras de polvo se almacenaron en bolsas plásticas y se trasladaron en una hielera. En el laboratorio, las muestras se secaron a temperatura ambiente y luego se tamizaron para su posterior análisis en el laboratorio, siguiendo a [20].

Cuadro 1. Densidad vehicular de los sitios de muestreo y de los monitores utilizados.

Punto	Densidad Vehicular (Vehículos/día)	Provincia	Cantón (Distrito)	Monitor
1	0	Cartago	El Guarco (Tablón)	C.I., C.e. y PC
2	17101	Alajuela	Central (San José)	C.I., C.e. y PC
3	32138	Heredia	Barva (San Pablo)	C.e. y PC
	29198			C.I.
4	45326	Cartago	San Nicolás (Taras)	C.I., C.e. y PC
5	62008	Cartago	La Unión (Tres Ríos)	C.I., C.e. y PC
6	77671	Alajuela	Central (Río Segundo)	C.I.
	73982	Heredia	Belén	C.e. y PC
7	88585	San José	Central (Hatillo)	C.I., C.e. y PC
8	104558	San José	Central (La Uruca)	C.I., C.e. y PC
V1	3639	Cartago	Paraíso (Central)	C.I., C.e. y PC
V2	17866	San José	Central (Vasconia)	C.e.
	23597	San José	San Francisco (Medal)	C.I. y PC
V3	31999	San José	Moravia (San Vicente)	C.I., C.e. y PC
V4	44228	San José	Curridabat (Cipreses)	C.I., C.e. y PC

Plantas: *Casuarina equisetifolia* (C.e.). *Cupressus lusitanica* (C.I.). Polvo de carretera (PC).

Análisis Químico

Los análisis químicos se realizaron en réplicas de muestras de hojas sin lavar y de polvo de carretera, con el fin de determinar el contenido de los metales Cu, Cr, Ni, Pb, V, Zn, Fe y Cd. Para la digestión de las muestras se aplicó el método EPA 3051 de Digestión Ácida Asistida por Microondas para sólidos, utilizando un sistema Mars 6 (CEM) [23]. Las determinaciones se efectuaron mediante espectrofotometría de absorción atómica (Perkin Elmer, modelo AAnalyst 800), con el apoyo de un horno de grafito y un generador de hidruros (Perkin Elmer, modelo FIAS 100) para la detección de trazas. Como estándar de referencia se empleó IV-Stock-10 de Inorganic Ventures.

Mediciones magnéticas

En el análisis de las propiedades magnéticas se siguió el protocolo descrito por Dearing [24]. Las muestras secas fueron trituradas con un mortero y luego compactadas en un recipiente plástico de 8 cm³. La medición de la susceptibilidad magnética a baja frecuencia (klf a 0,46

kHz) y a alta frecuencia (khf a 46,0 kHz) se realizó utilizando un sensor de frecuencia dual para muestras individuales Bartington MS2B y un medidor de susceptibilidad magnética Bartington MS3. Estos resultados se utilizaron para determinar:

$$\chi_{lf} = \frac{\kappa_{lf}}{\rho} \quad (1)$$

donde ρ es la densidad del material en $\text{kg}\cdot\text{m}^{-3}$.

El porcentaje de susceptibilidad magnética dependiente de la frecuencia:

$$\chi_{fd\%} = \left[\left(\frac{\kappa_{lf} - \kappa_{hf}}{\kappa_{lf}} \right) \right] \times 100 \quad (2)$$

Estos parámetros magnéticos permitieron identificar la concentración de material magnético y la presencia de partículas ultrafinas (≤ 30 nm) con características ferrimagnéticas (SP) en las hojas de los biomonitores respectivamente.

Formulación y validación de los modelos

La formulación y validación de los modelos utilizó el software RStudio, versión 1.4.1717, junto con los paquetes *tidyverse*, *caret* y *e1071* [25]. Se desarrollaron cuatro tipos de modelos para cada uno de los biomonitores y para el polvo de carretera, obteniendo en todos los casos como datos de salida la concentración de metales (Cuadro 2). Los modelos seleccionados para cada metal y monitor se eligieron entre al menos cincuenta ensayos de modelos.

Cuadro 2. Tipo de modelo desarrollado para la obtención de la concentración de PTEs.

Modelo	Tipo	Entrada (Propiedad magnética)
I	Máquina de Regresión lineal simple (MLR-S)	χ_{lf}
II	Máquina de Vectores de Soporte simple (SVM-S)	χ_{lf}
III	Máquina de Regresión lineal múltiple (MLR-M)	χ_{lf} y $\chi_{fd\%}$
IV	Máquina de Vectores de Soporte múltiple (SVM-M)	χ_{lf} y $\chi_{fd\%}$

Máquina de Regresión lineal múltiple y simple (MLR-S y MLR-M)

Se generaron modelos de regresión lineal simple y múltiple mediante *machine learning*, basados en parámetros magnéticos y concentraciones de metales pesados, del tipo:

$$HM_c = (\beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3) + e \quad (3)$$

Donde β_0 es el intercepto, β_1 , β_2 y β_3 son los coeficientes de regresión asociados a cada parámetro magnético Z_i , y e es el error asociado al modelo [26].

Máquina de vectores de soporte simple y múltiple (SVM-S y SVM-M)

Los modelos generados mediante máquinas de vectores de soporte (SVM) corresponden a algoritmos de aprendizaje supervisado utilizados para tareas de clasificación y regresión. En el caso de la regresión por vectores de soporte, la línea que mejor ajusta los datos se denomina

hiperplano [27]. En estos modelos, las variables de entrada se transforman a un espacio de características de alta dimensión mediante una función kernel, lo que permite abordar relaciones no lineales dentro de un marco lineal. En este espacio de alta dimensión, la SVM aproxima el conjunto de datos mediante una función lineal descrita por la ecuación:

$$y = f(x, \omega) = \sum_{i=1}^m \omega_i \phi(x_i) + b \quad (4)$$

donde $\phi(x_i)$ representa las variables de entrada después de la transformación kernel, y ω_i y b son los coeficientes estimados al minimizar la función de riesgo regularizada [28]. Se evaluaron diferentes funciones kernel (lineal y RBF), seleccionándose aquella que presentó el mejor desempeño según validación cruzada. Los modelos SVM se implementaron en R mediante el paquete `e1071`, utilizando los valores por defecto del algoritmo. En particular, se empleó un kernel RBF con parámetro de regularización $C = 1$ y $\gamma = 1/p$, donde p corresponde al número de variables predictoras.

Los modelos fueron validados por medio de dos metodologías. El primer método consistió en realizar una validación cruzada utilizando un enfoque de machine learning, conocido como Set Approach o división de datos 80/20. En este los datos fueron divididos aleatoriamente en dos conjuntos, para el estudio el 80% para el conjunto fue utilizado para el entrenamiento del modelo y el 20% para el conjunto de prueba de este [29]. El segundo método de validación consistió en tomar una muestra para cada biomonitor y polvo de carretera en cuatro nuevos puntos de muestreo (Cuadro 1), siguiendo los mismos procedimientos de muestreo que se utilizaron para construir los modelos.

Para evaluar el desempeño de los modelos, se utilizó el coeficiente de determinación ajustado (R^2) comparando las concentraciones elementales predichas con las observadas. Además, se emplearon el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE) para evaluar la eficiencia predictiva del modelo.

Resultados y discusión

Rendimiento de los modelos

La selección del modelo óptimo para cada metal se basó en el valor del R^2 , mediante la evaluación del índice de mejora entre modelos sucesivos [6]. En la práctica, un R^2 alto confirma que el modelo es fiable para seguir los cambios en las concentraciones de los metales, asegurando que los resultados estadísticos realmente se reflejen en predicciones químicas exactas. Al comparar los valores de R^2 obtenidos para los diferentes modelos (Cuadro 3), en el caso de *C. equisetifolia*, al pasar de los modelos MLR-S y MLR-M a los modelos SVM-S y SVM-M, se obtuvo un índice de mejora positivo en seis de los siete metales analizados. En *C. lusitanica* y en el polvo de carretera, la mejora se presentó en todos los metales estudiados. En general, para todos los monitores y metales, el cambio de MLR a SVM resultó en un mejor desempeño, por lo que este tipo de modelo fue el seleccionado, variando entre las versiones simple y múltiple según el metal analizado.

Cuadro 3. Índice de cambio del valor de R^2 al pasar de los modelos MLR simple y múltiples a los modelos SVM simple y múltiples.

Monitor		R^2 MLR-S	R^2 SVM-S	Índice de mejora	R^2 MLR-M	R^2 SVM-M	Índice de mejora
<i>Casuarina equisetifolia</i>	Fe	0.92	0.94	2%	0.92	0.92	0%
	Cu	0.52	0.74	42%	0.55	0.66	21%
	Pb	0.13	0.41	222%	0.18	0.44	146%
	Cr	0.82	0.80	-3%	0.83	0.84	1%
	Ni	0.19	0.34	77%	0.22	0.35	59%
	V	0.86	0.86	1%	0.85	0.81	-5%
	Zn	0.49	0.65	33%	0.48	0.55	15%
<i>Cupressus lusitanica</i>	Fe	0.71	0.75	5%	0.75	0.76	1%
	Cu	0.29	0.41	43%	0.29	0.54	85%
	Pb	0.24	0.38	55%	0.24	0.34	38%
	Cr	0.43	0.59	36%	0.44	0.48	10%
	Ni	0.22	0.34	56%	0.17	0.36	111%
	V	0.17	0.27	54%	0.19	0.29	50%
	Zn	0.50	0.67	35%	0.20	0.33	69%
Polvo de carretera	Fe	0.029	0.079	170%	0.040	0.171	326%
	Cu	0.002	0.030	1619%	0.026	0.051	94%
	Pb	0.021	0.024	16%	0.029	0.082	187%
	Cr	0.003	0.107	3650%	0.086	0.272	215%
	Ni	0.025	0.110	335%	0.057	0.276	381%
	V	0.002	0.120	4713%	0.004	0.273	6076%
	Zn	0.002	0.109	6530%	0.137	0.288	110%

En la Fig. 2. en donde la línea discontinua negra ($y = x$) indica el ideal: rendimiento idéntico en entrenamiento y prueba, puede observarse que los modelos seleccionados para cada metal de los tres biomonitores presentan los puntos sobre la línea sugiriendo un mejor desempeño en prueba que en entrenamiento, siendo esto importante ya que los puntos bajo la línea podrían indicar ligero sobreajuste de los modelos, siendo esta una de las precauciones a considerar al trabajar con machine learning. Además, se puede observar que para ambas plantas el Fe, Cu, Zn y Cr presentaron los valores más altos y consistentes, mientras que Ni y Pb se alejaron más de la línea en R^2 . en cuanto al polvo de carretera este presentó los valores más bajos de R^2 . En el análisis de correlación entre las propiedades magnéticas y el contenido de metales, las muestras de plantas también presentaron los coeficientes de correlación de Pearson (r) más altos y estadísticamente significativos entre los metales analizados [16]

El valor bajo R^2 obtenido para el polvo de carretera se explica por las múltiples fuentes de metales que se acumulan en los mismos, ya que, en la cuneta, donde se muestreo el polvo se acumulan restos metálicos procedentes del tráfico, aguas residuales grises de las viviendas y negocios cercanos y fuentes naturales (deposición atmosférica seca y húmeda, resuspensión del suelo superficial [30].

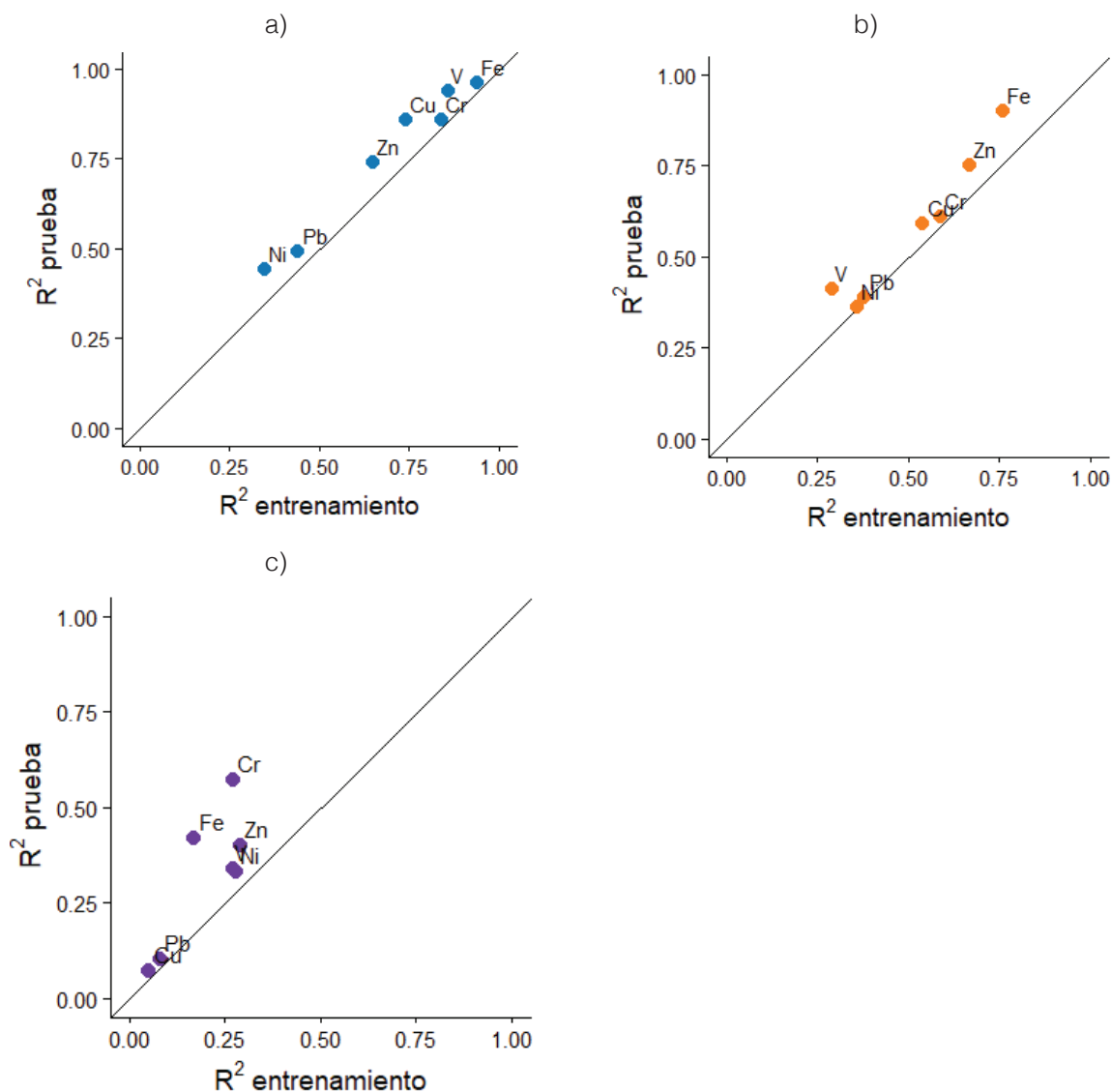


Figura 2. Rendimiento de modelos de los metales mediante R^2 de los datos de entrenamiento vs los de prueba para a) *C. equisetifolia*, b) *C. lusitanica* y c) polvo de carretera.

Comparación entre los RMSE y MAE de entrenamiento y prueba

En cuanto a los errores RMSE y MAE (Fig. 3), la similitud entre los datos de entrenamiento y prueba para los tres monitores indica una buena capacidad de generalización. En términos prácticos, el MAE nos indica el error promedio que el modelo comete en la cuantificación de cada metal, mientras que el RMSE confirma la ausencia de errores grandes o atípicos. El hecho de que ambos valores sean bajos y cercanos entre sí asegura que el modelo es estable y que no existe sobreajuste (overfitting), validando su precisión para analizar muestras nuevas.

Los metales con menores errores fueron Cr y V en *C. equisetifolia* y *C. lusitanica*, así como Cr y Ni en el polvo de carretera. Sin embargo, este último mostró errores mucho mayores en comparación con los obtenidos en las plantas.

En cuanto a los metales, el Fe presentó los valores de error más altos, incluyendo los datos de validación, en línea con lo reportado en estudios previos[31], [32]. Este comportamiento es consistente con su carácter ambientalmente estable, por lo que el Fe se emplea comúnmente como elemento normalizador [33].

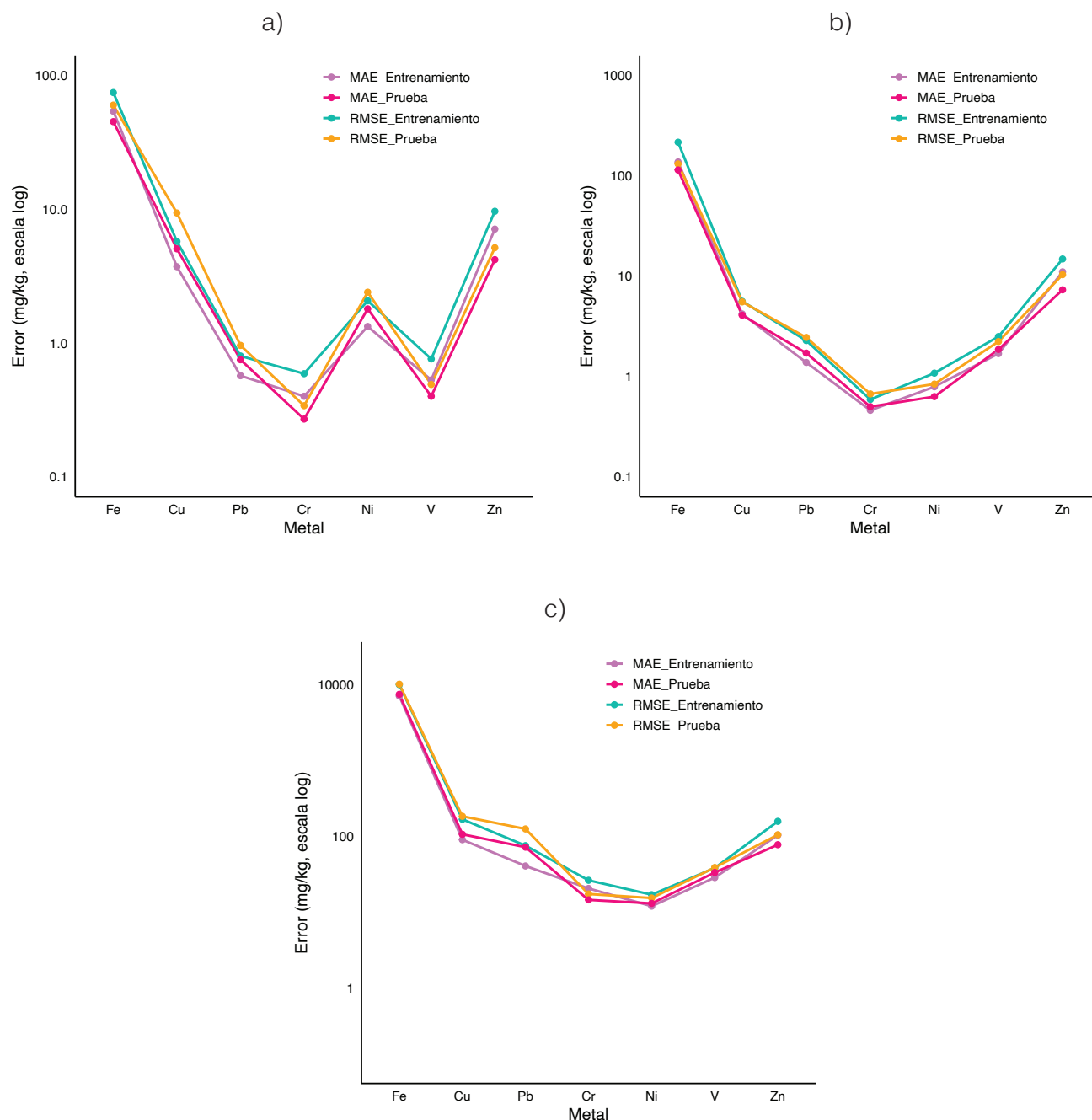


Figura 3. Error de predicción (RMSE y MAE) para cada metal de los datos de entrenamiento vrs los de prueba para a) *Casuarina equisetifolia*, b) *Cupressus lusitanica* y c) polvo de carretera.

Comparación entre los RMSE y MAE de validación con puntos nuevos

Al comparar los valores de RMSE y MAE para ambas plantas, se observa en la Figura 4a y 4b que estos mantienen el mismo patrón que los datos de entrenamiento del modelo. Los metales con los valores más bajos de error son Cr, V y Ni, mientras que el valor más alto corresponde a Fe. Con respecto al polvo de carretera, Fig. 4c, concuerda con los valores más bajos de RMSE y MAE para los metales Ni, V y Cr, sin embargo, los valores de error fueron bastante mayores a los obtenidos por las plantas. Esta observación se justifica por lo previamente señalado sobre la multiplicidad de fuentes de contaminación a las que está expuesto y que contribuyen a la composición del polvo de carretera.

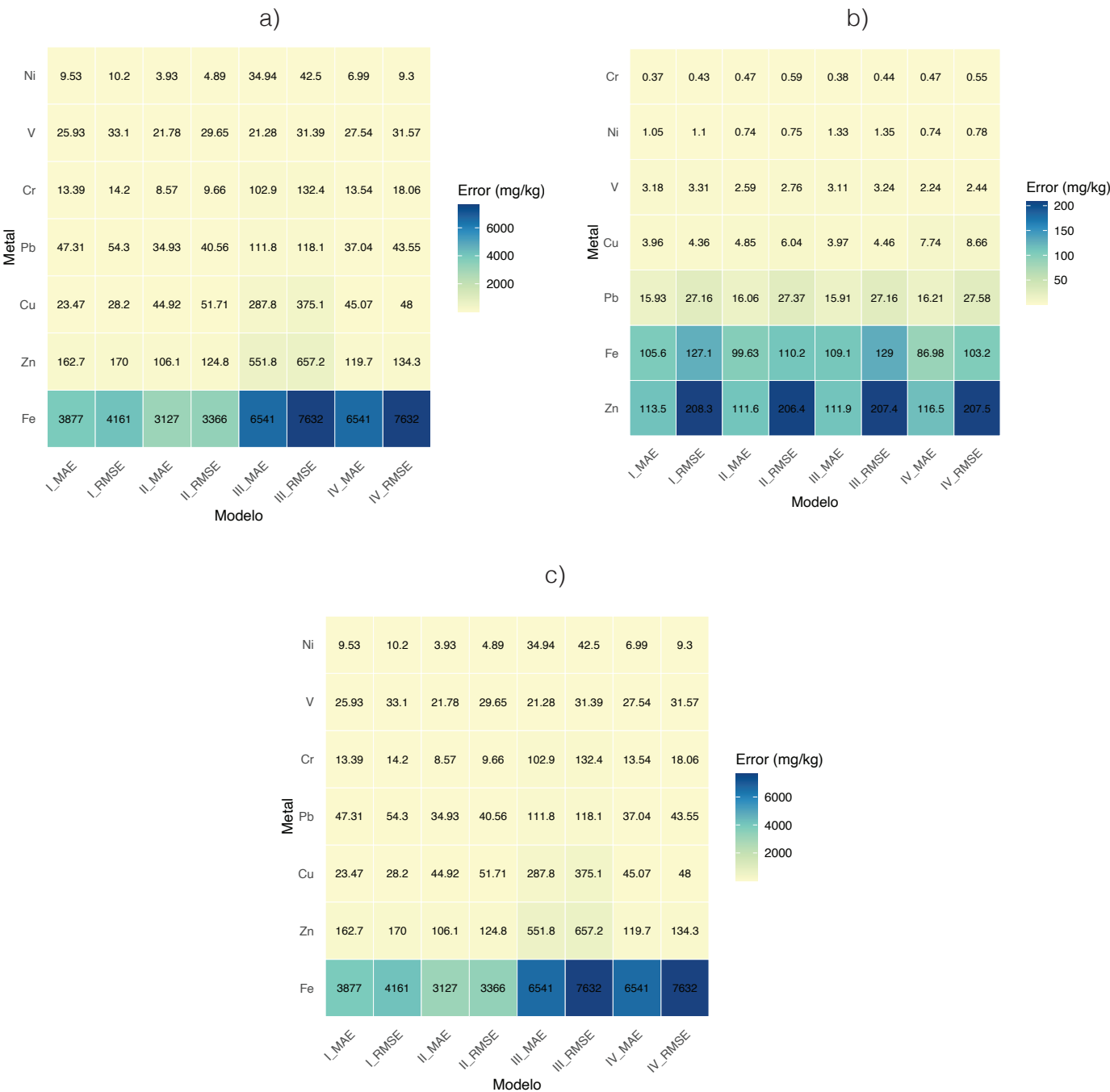


Figura 4. Error de predicción (RMSE y MAE) para cada metal en puntos de validación en puntos nuevos para a) *C. equisetifolia*, b) *C. lusitanica* y c) polvo de carretera.

Es fundamental considerar que la capacidad de generalización de los modelos desarrollados está supeditada a la naturaleza de la matriz química y al rango de concentraciones evaluado. Estos factores intrínsecos sugieren que la transferencia del modelo a nuevos escenarios experimentales requeriría necesariamente una reoptimización de los hiperparámetros C (penalización de errores) y γ (alcance de la influencia del kernel) para evitar el subajuste o el sobreajuste. Asimismo, debe reconocerse una limitación inherente a la arquitectura de las SVM, su dependencia del volumen de datos. Si bien el algoritmo es altamente eficiente para las dimensiones muestrales de esta investigación, el incremento exponencial en el costo computacional ante conjuntos de datos masivos representa un desafío para su escalabilidad, lo que podría exigir el empleo de modelos con menor complejidad computacional o algoritmos de aprendizaje optimizados para grandes volúmenes de datos.

Conclusiones

Los mejores valores ajustados de R^2 tanto en las etapas de entrenamiento como de prueba del modelo predictivo para *Casuarina equisetifolia*, *Cupressus lusitanica* y el polvo de carretera se obtuvieron con el modelo de Máquina de Vectores de Soporte (SVM), con variaciones entre el uso de una sola variable (x_{lf}) y múltiples (x_{lf} y $x_{fd\%}$).

Los resultados mostraron tanto en la etapa de entrenamiento como en la de prueba, que las concentraciones de Fe, Cu, Cr, V y Zn fueron bien estimadas por los modelos de predicción SVM, con valores R^2 ajustados > 0.7 y las concentraciones de Pb y Ni fueron estimados con menor precisión, con valores R^2 ajustados < 0.7 .

Los resultados del monitor *C. equisetifolia* fueron los más acertados para determinar la concentración de PTEs a partir de sus propiedades magnéticas, con los mejores valores de R^2 ajustado y RMSE y MAE más bajos, siendo por ello superior al *C. lusitanica* y al polvo de carretera, para el monitoreo de la contaminación del aire por PTEs.

Estos resultados destacan la importancia de la inteligencia artificial (IA) en el monitoreo ambiental, ya que los modelos de *machine learning* permitieron identificar relaciones complejas entre las propiedades magnéticas y la concentración de metales pesados, mejorando la capacidad predictiva frente a métodos convencionales.

Agradecimientos

Los autores agradecen al Instituto Tecnológico de Costa Rica (ITCR) por su apoyo financiero y administrativo. Asimismo, al Centro de Investigación en Protección Ambiental (CIPA), al Centro de Investigación y de Servicios Químicos y Microbiológicos (CEQIATEC) (Costa Rica), al Instituto de Geofísica de la Universidad Nacional Autónoma de México (UNAM, Morelia), y al Instituto Meteorológico Nacional (IMN) por su valioso apoyo al proyecto.

Cumplimiento de normas éticas

Los bioindicadores fueron recolectados de acuerdo con las normas éticas establecidas por el Instituto Tecnológico de Costa Rica.

Referencias

- [1] X. Zhang, K. Shu, S. Rajkumar, and V. Sivakumar, "Research on deep integration of application of artificial intelligence in environmental monitoring system and real economy," *Environ. Impact Assess. Rev.*, vol. 86, Jan. 2021, doi: 10.1016/j.eiar.2020.106499.

- [2] E. Alotaibi and N. Nassif, "Artificial intelligence in environmental monitoring: in-depth analysis," *Discover Artificial Intelligence*, vol. 4, no. 1, Dec. 2024, doi: 10.1007/s44163-024-00198-1.
- [3] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, "Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions," *Hygiene and Environmental Health Advances*, vol. 12, Dec. 2024, doi: 10.1016/j.heha.2024.100114.
- [4] S. M. Popescu *et al.*, "Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management," *Front. Environ. Sci.*, vol. 12, 2024, doi: 10.3389/fenvs.2024.1336088.
- [5] E. C. McClure *et al.*, "Artificial Intelligence Meets Citizen Science to Supercharge Ecological Monitoring," *Patterns*, vol. 1, no. 7, Oct. 2020, doi: 10.1016/j.patter.2020.100109.
- [6] T. Salazar-Rojas, F. R. Cejudo-Ruiz, and G. Calvo-Brenes, "Comparison Between Machine Linear Regression (MLR) and Support Vector Machine (SVM) as Model Generators for Heavy Metal Assessment Captured in Biomonitoring and Road Dust," *Environmental Pollution*, vol. 314, p. 120227, 2022, doi: <https://doi.org/10.1016/j.envpol.2022.120227>.
- [7] S. Ferebee, "Predicting Tomorrow: A Review of Machine Learning's Role in Shaping Environmental Forecasts," *Premier Journal of Science*, vol. 2, Nov. 2024, doi: 10.70389/PJS.100008.
- [8] Q. Dai, M. Zhou, H. Li, X. Qian, M. Yang, and F. Li, "Biomagnetic monitoring combined with support vector machine: a new opportunity for predicting particle-bound-heavy metals," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020, doi: 10.1038/s41598-020-65677-8.
- [9] Y. Xu *et al.*, "Integrating machine learning for enhanced spatial prediction and risk assessment of soil heavy metal(loid)s," *Environmental Pollution*, vol. 383, Oct. 2025, doi: 10.1016/j.envpol.2025.126919.
- [10] P. I. Angulo-Bejarano, J. Puente-Rivera, R. Cruz-Ortega, and E. Hanus-Fajerska, "plants Metal and Metalloid Toxicity in Plants: An Overview on Molecular Aspects Metal and Metalloid Toxicity in," *Plants: An Overview on Molecular Aspects. Plants*, vol. 10, p. 635, 2021, doi: 10.3390/plants.
- [11] T. Salazar-Rojas, F. R. Cejudo-Ruiz, and G. Calvo-Brenes, "Magnetic and Chemical Testing in Plants, Road Dust and Soil, as Indicators of Atmospheric Pollution," *Water Air Soil Pollut.*, vol. 235, no. 9, Sep. 2024, doi: 10.1007/s11270-024-07333-9.
- [12] S. Stankovic, P. Kalaba, and A. R. Stankovic, "Biota as toxic metal indicators," *Environ. Chem. Lett.*, vol. 12, no. 1, pp. 63–84, 2014, doi: 10.1007/s10311-013-0430-6.
- [13] G. Chen-ming, W. Bo, and S. Xiao-han, "Biomagnetic monitoring of urban atmospheric pollution: A review of magnetic signatures from different types of plants," *Science of the Total Environment*, vol. 963, Feb. 2025, doi: 10.1016/j.scitotenv.2025.178518.
- [14] T. Salazar-Rojas, S. Murillo-Murillo, R. Ulate-Molina, F. R. Cejudo-Ruiz, and G. Calvo-Brenes, "Spatial assessment of potentially toxic elements (PTEs) pollution in the vicinity of a cement plant through magnetic and chemical testing in plants and soil," *Environmental Challenges*, vol. 21, Dec. 2025, doi: 10.1016/j.envc.2025.101313.
- [15] M. Arsyad, V. A. Tiwow, M. J. Rampe, and H. L. Rampe, "Relationships of Magnetic Properties and Heavy Metals Content of Guano in Bat Cave, South Sulawesi, Indonesia," *Karbala International Journal of Modern Science*, vol. 8, no. 3, pp. 406–414, 2022, doi: 10.33640/2405-609X.3254.
- [16] T. Salazar-Rojas, F. R. Cejudo-Ruiz, and G. Calvo-Brenes, "Assessing magnetic properties of biomonitoring and road dust as a screening method for air pollution monitoring," *Chemosphere*, vol. 310, no. September 2022, 2023, doi: 10.1016/j.chemosphere.2022.136795.
- [17] Nation State Program, "Seguimiento al desarrollo humano sostenible," 2019. [Online]. Available: <https://estadonacion.or.cr/informes/>
- [18] Ministry of Public Construction and Transportation, "ANUARIO DE INFORMACIÓN DE TRÁNSITO 2019," Costa Rica, 2020.
- [19] M. A. E. Chaparro, M. A. E. Chaparro, A. G. Castañeda Miranda, H. N. Böhnelt, and A. M. Sinito, "An interval fuzzy model for magnetic biomonitoring using the specie *Tillandsia recurvata* L.," *Ecol. Indic.*, vol. 54, pp. 238–245, 2015, doi: 10.1016/j.ecolind.2015.02.018.
- [20] B. Aguilar *et al.*, "Ficus benjamina leaves as indicator of atmospheric pollution: A reconnaissance study," *Studia Geophysica et Geodaetica*, vol. 56, no. 3, pp. 879–887, Jun. 2012, doi: 10.1007/s11200-011-0265-1.
- [21] S. Zamudio and E. Carranza, "Cupressaceae," *Flora del Bajío y de Regiones Adyacentes*, vol. 29, Pátzcuaro, Michoacán, 1994. Accessed: Mar. 27, 2025. [Online]. Available: <http://inecolbajio.inecol.mx/floradelbajio/documentos/fasciculos/ordinarios/Cupressaceae%2029.pdf>

- [22] F. Rojas-Rodríguez and G. Torres-Córdoba, "Casuarina," *Revista Forestal Mesoamericana Kurú*, vol. 10, no. 25, pp. 32–33, 2013, doi: <https://doi.org/10.18845/rfmk.v10i25.1375>.
- [23] U.S. EPA, "Method 3051A Microwave Assisted Acid Digestion of Sediments, Sludges, Soils, and Oils,," no. February, USA, pp. 1–30, 2007. Available: <https://www.epa.gov/sites/default/files/2015-12/documents/3051a.pdf>
- [24] J. A. Dearing, "Environmental Magnetic Susceptibility: Using the Bartington MS2 System,," *Environmental Magnetic Susceptibility*, p. 52, 1999, Accessed: Mar. 27, 2025. [Online]. Available: https://gmw.com/wp-content/uploads/2019/03/JDearing-Handbook-OM0409.pdf?srsitid=AfmBOooQ9qCZFrk_srScVdcwAG-0c4Yt1k6uz-PM5ODYwv2XiGXpM3BiR
- [25] RStudio Team, "RStudio: Integrated Development for R. RStudio, PBC.,," 2020, *R Foundation for Statistical Computing, Boston, USA*: 1.4.1717. [Online]. Available: <http://www.rstudio.com/>.
- [26] E. Acuña, "Regresión Aplicada usando R," 2015, [Online]. Available: https://www.researchgate.net/publication/335796371_REGRESION_APLICADA_USANDO_R
- [27] A. Raj, "Unlocking the True Power of Support Vector Regression,," Towards Data Science. Accessed: Dec. 17, 2021. [Online]. Available: <https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0>
- [28] H. Li, J. Wang, Q. Wang, C. Tian, X. Qian, and X. Leng, "Magnetic Properties as a Proxy for Predicting Fine-Particle-Bound Heavy Metals in a Support Vector Machine Approach,," *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6927–6935, 2017, doi: 10.1021/acs.est.7b00729.
- [29] H. Li, J. Wang, Q. Wang, C. Tian, X. Qian, and X. Leng, "Magnetic Properties as a Proxy for Predicting Fine-Particle-Bound Heavy Metals in a Support Vector Machine Approach,," *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6927–6935, 2017, doi: 10.1021/acs.est.7b00729.
- [30] G. Wang *et al.*, "Magnetic properties of street dust in Shanghai, China and its relationship to anthropogenic activities,," *Environmental Pollution*, vol. 255, p. 113214, 2019, doi: 10.1016/j.envpol.2019.113214.
- [31] A. Schmidt, R. Yarnold, M. Hill, and M. Ashmore, "Magnetic susceptibility as proxy for heavy metal pollution: A site study,," *J. Geochem. Explor.*, vol. 85, no. 3, pp. 109–117, 2005, doi: 10.1016/j.gexplo.2004.12.001.
- [32] S. G. Lu, S. Q. Bai, and Q. F. Xue, "Magnetic properties as indicators of heavy metals pollution in urban topsoils: A case study from the city of Luoyang, China,," *Geophys. J. Int.*, vol. 171, no. 2, pp. 568–580, 2007, doi: 10.1111/j.1365-246X.2007.03545.x.
- [33] O. O. Okedeyi, S. Dube, O. R. Awofolu, and M. M. Nindi, "Assessing the enrichment of heavy metals in surface soil and plant (*Digitaria eriantha*) around coal-fired power plants in South Africa,," *Environmental Science and Pollution Research*, vol. 21, no. 6, pp. 4686–4696, 2014, doi: 10.1007/s11356-013-2432-0.

Declaración sobre uso de Inteligencia Artificial (IA)

Los autores aquí firmantes declaramos que no se utilizó ninguna herramienta de IA para la conceptualización, traducción o redacción de este artículo.