



**TECNOLÓGICO
NACIONAL DE MÉXICO®**



Tecnológico Nacional de México

Instituto Tecnológico de Culiacán

Ingeniería en sistemas computacionales

Tópicos de Inteligencia Artificial

11:00 – 12:00

Proyecto de Investigación

Alumno:

Zavala Carmona Miguel Angel

12/09/2025

Link al repositorio de GitHub: <https://github.com/MiguelZC20/Topicos-de-IA.git>

Mapeo hiperlocal de calidad del aire y temperatura mediante muestreo móvil y modelos de inteligencia artificial.

1.- Introducción

La calidad del aire y la temperatura urbana son factores ambientales críticos que inciden directamente en la salud pública, la productividad y la calidad de vida en las ciudades. La evidencia global y regional muestra que la exposición prolongada y aguda a contaminantes como las partículas finas (PM_{2.5}) y gases como NO₂ se asocia con un aumento de mortalidad por enfermedades respiratorias y cardiovasculares. En contextos urbanos, además, la distribución espacial de contaminantes no es homogénea, es decir, factores locales como el tráfico, la morfología urbana, uso del suelo y cobertura vegetal generan heterogeneidades a escala de cuadra que rara vez captan las estaciones oficiales que miden los índices de la calidad del aire.

Frente a esta limitación, el presente proyecto plantea explorar el potencial de integrar instrumentación de respuesta rápida y técnicas de inteligencia artificial para producir mapas hiperlocales de contaminantes en un área piloto con el objetivo de generar evidencia técnica que apoye decisiones de mitigación focalizada.

La literatura reciente aporta una base metodológica sólida para este enfoque, con estudios que equiparon vehículos con instrumentación rápida demostrando la capacidad del muestreo móvil para revelar heterogeneidades a escala de cuadra. Asimismo investigaciones que muestran que la calibración de sensores de bajo costo con modelos de inteligencia artificial puede alcanzar una precisión que los hace útiles como complementos a las estaciones oficiales. Finalmente, comparativas entre modelos de inteligencia artificial indican que la selección del mejor enfoque depende del objetivo y el horizonte considerado.

2.- Objetivos

2.1.- Objetivo general

Desarrollar y validar un protocolo de investigación para el mapeo hiperlocal de la temperatura urbana y la calidad del aire mediante campañas de muestreo móvil complementadas por sensores temporales y técnicas de inteligencia artificial, con el fin del generar mapas de alta resolución espacial que permitan identificar hotspots de contaminación y evaluar el potencial de intervenciones locales basadas en los resultados.

2.2.- Objetivos específicos

Diseñar un protocolo de muestreo móvil.

- Definir rutas piloto, horarios representativos y puntos de colocación de sensores temporales considerando la configuración urbana y zonas de actividad local.

Identificar y compilar fuentes de datos locales para calibración y contextualización.

- Reunir series históricas y en tiempo real de estaciones oficiales de sistemas de monitoreo de la calidad del aire y otros repositorios relevantes.

Recolectar datos piloto mediante campañas móviles y sensores temporales en zonas seleccionadas de la ciudad.

- Ejecutar pasadas en distintos días y horarios, y desplegar sensores estáticos en puntos estratégicos para generar pares de calibración y series locales.

Implementar procedimientos de control de calidad y calibración orientados a las condiciones locales.

- Entrenar y evaluar modelos de calibración usando pares sensores low-cost/estación oficial para corregir sesgos producidos por temperatura, humedad y condiciones operativas propias de la ciudad.

Construir mapas hiperlocales y evaluar su precisión espacial y temporal.

- Probar métodos (por ejemplo, XGBoost con covariables urbanas) para generar mapas de PM2.5 y temperatura, reportando errores e incertidumbres.

Analizar causas locales y proponer intervenciones de mitigación

- Usar explicabilidad y covariables (uso del suelo, distancias, vegetación) para sugerir intervenciones locales que reduzcan el impacto en la salud.

3.- Justificación

La contaminación del aire urbano y el aumento de la temperatura en las ciudades constituyen problemas ambientales y de salud pública de primer orden. La exposición continua a contaminantes como PM2.5, NO2 y ozono se asocia con el aumento de la mortalidad por enfermedades respiratorias y cardiovasculares, y la combinación de olas de calor con episodios de alta contaminación agrava la vulnerabilidad de poblaciones sensibles. Además, las ciudades presentan una heterogeneidad espacial marcada, es decir, la calidad del aire y la temperatura pueden variar de forma significativa a escala de cuadra debido a factores como tráfico, configuración urbana, cobertura vegetal y fuentes puntuales de emisión (Apte et al, 2017).

A pesar de la existencia de redes de monitoreo oficiales, su densidad espacial es insuficiente en muchas ciudades, incluida Culiacán, lo que produce huecos informativos y dificulta la identificación de hotspots donde la población está más expuesta. En este escenario, la inteligencia artificial (IA) y las técnicas de modelado “spatio-temporal” ofrecen herramientas prometedoras para fusionar fuentes heterogéneas (estaciones oficiales, sensores low-cost, datos satelitales y covariables urbanas) y así producir estimaciones más precisas y de mayor resolución espacial.

4.- Alcance

El proyecto se llevará a cabo en la ciudad de Culiacán, Sinaloa, durante un periodo aproximado de cuatro meses, y seguirá un enfoque por fases que incluye: diseño y preparación, campañas móviles de muestreo, preprocesamiento y calibración de datos, modelado y validación, generación y análisis de propuestas de intervención, y entrega de resultados.

El estudio estará delimitado a un área piloto acotada dentro de la ciudad, seleccionada por su representatividad urbana según criterios como lo son la presencia de avenidas de alta intensidad vehicular, zonas comerciales, variación en la densidad de edificación y cobertura vegetal, con el fin de capturar la heterogeneidad espacial típica de la ciudad.

5.- Desarrollo

5.1.- Marco Referencial

La calidad del aire es un factor determinante para la salud y el bienestar de la población, especialmente en zonas urbanas donde el crecimiento industrial y vehicular incrementa la concentración de contaminantes (López et al., 2025).

Según la Organización Mundial de la Salud (UNEP, 2022):

El 99% de la población mundial respira aire insalubre y la contaminación atmosférica causa 7 millones de muertes prematuras cada año. El PM2.5, que se refiere al conjunto de partículas con un diámetro igual o inferior a 2,5 micrómetros, representa la mayor amenaza para la salud y a menudo se utiliza como métrica en las normas legales de calidad del aire.

La presente investigación se sitúa dentro del conocimiento académico e institucional existente sobre calidad del aire, temperatura urbana y aplicaciones de inteligencia artificial (IA) en el monitoreo y la calidad del aire urbano. Estudios y revisiones recientes han demostrado que la fusión de datos provenientes de estaciones oficiales, sensores de bajo costo e imágenes satelitales, junto con técnicas de machine learning y modelos spatio-temporales, mejoran la cobertura espacial y la capacidad predictiva de PM2.5 y otros contaminantes en entornos urbanos.

Un estudio realizado en California, Estados Unidos, equipó vehículos (Google Street View cars) con instrumentación de respuesta rápida y muestreó repetidamente todas las calles de un área urbana de 30 km², demostrando que la monitorización móvil permite revelar heterogeneidades a escala de cuadra (hotspots) que quedarían ocultas con estaciones fijas. Además, el trabajo analiza cuántas pasadas son necesarias para obtener medidas estables y aconseja repetir muestreos para promediar la variabilidad temporal local (Apte et al, 2017).

Los estudios sobre sensores de bajo costo revelan que, sin calibración, sus lecturas suelen presentar sesgos importantes y sensibilidad a condiciones ambientales (temperatura, humedad), lo que limita su uso directo en investigación aplicada o en soporte a decisiones. Sin embargo, trabajos experimentales recientes demuestran que modelos de machine learning (por ejemplo, Decision Tree, XGBoost) aplicados como calibradores pueden mejorar dramáticamente la precisión de esos sensores (Ravindra et al, 2024).

En cuanto a la predicción y el modelado, comparativas recientes entre métodos clásicos y deep learning muestran que no existe un único “mejor” modelo para todas las condiciones. El rendimiento depende del contaminante, del horizonte de predicción y de la calidad/duración de las series.

Un estudio comparativo aplicado a PM2.5 y PM10 en estaciones reales en Emiratos Árabes Unidos Abuouelezz et al. (2025) encontró que:

Support Vector Regression (SVR) ofreció los mejores resultados para PM2.5 en horizontes muy cortos (1–2 horas) en términos de MAPE (18% – 28%), mientras que

CNN obtuvo mejor desempeño para PM10 a horizontes de 1 hora y que modelos tipo Prophet rindieron especialmente bien en horizontes diarios/semanales.

Estas diferencias recomiendan evaluar varios enfoques y seleccionar o combinar modelos según el objetivo, ya sea predicción inmediata para alertas o estimación espacial para mapas.

5.2.- Marco contextual

La calidad del aire puede compararse al estado del tiempo, es decir, puede cambiar rápidamente, incluso en cuestión de horas.

De acuerdo con el centro de investigación de la NASA:

Para medir e informar sobre la calidad del aire, la EPA utiliza el Índice de Calidad del Aire (ICA) de Estados Unidos. El ICA se calcula midiendo cada uno de los seis contaminantes principales del aire en una escala de “Bueno” a “Peligroso”, para producir un valor numérico combinado de la ICA de 0 a 500 (NASA, 2025).

En México, en el ámbito estatal y municipal existen las redes oficiales de monitoreo que proveen registros útiles (ProAire/SMCA) y muestran el índice de la calidad del aire con una escala similar a la de Estados Unidos, pero su cobertura espacial suele dejar vacíos en barrios y corredores específicos que habitualmente no capturan heterogeneidades a escala de cuadra. Es por ello, que el mapeo hiperlocal mediante campañas móviles y sensores temporales con técnicas de inteligencia artificial puede complementar la información oficial y revelar hotspots de exposición.

Culiacán, capital del estado de Sinaloa, es una ciudad que combina zonas de alta actividad vehicular y comercial con áreas residenciales, además de actividades agroindustriales en sus alrededores.

El proyecto se inscribe en la necesidad de entregar información accionable para autoridades municipales y comunidades locales, es decir, identificar hotspots de contaminación, estimar la exposición de la población y ofrecer escenarios de mitigación focalizados.

5.3.- Marco Teórico

5.3.1.- Conceptos, teorías y modelos

Los trabajos recientes que integran muestreo móvil, sensores de bajo costo, productos satelitales y algoritmos de aprendizaje automático ofrecen un soporte teórico y empírico robusto para diseñar protocolos que produzcan mapas urbanos de alta resolución espacial y herramientas de predicción operacional.

Estaciones oficiales

Las estaciones oficiales de monitoreo de la calidad del aire ofrecen datos calibrados y continuos, pero su calidad espacial suele ser baja por costos operativos y regulaciones, por lo tanto, carecen de la resolución necesaria para identificar heterogeneidades a escala de cuadra.

Sensores low-cost y muestreo móvil

Los sensores de bajo costo (low-cost) permiten ampliar la cobertura espacial y temporal, pero presentan problemas documentados, como lo pueden ser sesgos, sensibilidad a temperatura y humedad, y variabilidad con el tiempo. El muestreo móvil (instrumentar vehículos, bicicletas) permite muestrear de forma sistemática áreas amplias y detectar hotspots que las estaciones puntuales no captan, sin embargo, se exigen protocolos estrictos de montaje, velocidad y repetición para controlar sesgos inducidos por el movimiento y la estructura local (Apte et al, 2017).

Calibración de sensores

Dado entonces que los sensores low-cost padecen de sesgos y sensibilidad ambiental, la literatura recomienda la calibración local mediante modelos supervisados utilizando pares “sensor low-cost y estación oficial de referencia”. Ravindra et al. (2024) documentan que modelos de Machine Learning (árboles de decisión, boosting) pueden corregir sistemáticamente sesgos y reducir errores, siempre que se disponga de colocaciones representativas y procedimientos de validación adecuados.

Los modelos de calibración típicos usan como entradas la lectura cruda del sensor, variables ambientales (temperatura, humedad), y variables operacionales (velocidad para muestreo móvil, hora del día). Algoritmos como Random Forest, XGBoost o regresión lineal multivariada son opciones razonables, sin embargo, es crítico evaluar el rendimiento con métricas (RMSE, MAE, R^2) y aplicar monitoreo de variaciones reentrenamientos periódicos (Ravindra et al, 2024).

Modelado espacial

Los modelos supervisados con covariables como Random Forest o XGBoost predicen una variable objetivo (por ejemplo, PM_{2.5}) a partir de un conjunto de covariables (índice de vegetación, distancias, densidad edificatoria, meteorología), siendo flexibles, rápidos y permitiendo interpretar la importancia relativa de características.

En los modelos “spatio-temporales” y deep learning para capturar dinámica temporal y dependencias espaciales simultaneas, se emplean arquitecturas como la convolucional combinada con memoria a corto y largo plazo. Siendo potentes para la predicción a horizontes de horarios o de varios días, pero requiriendo series más densas y potencia computacional. Abuouelezz et al. (2025) muestran en su comparativa que el rendimiento relativo de modelos clásicos contra deep learning depende del horizonte y de la variable, por lo que la selección debe ser empírica y orientada al objetivo (alertas o mapas espaciales).

Validación y métricas de evaluación

Una limitación recurrente es la pérdida de rendimiento al aplicar modelos fuera de la región/condiciones donde fueron entrenados. Por ello, es esencial incorporar validación espacial y validación temporal para evitar el sobreoptimismo en las métricas. Por su parte, las métricas estándar incluyen RMSE, MAE y R^2 para predicción continua.

De acuerdo con Zhou et al (2025) en su investigación sobre el aprendizaje profundo para la predicción de PM2.5:

El error cuadrático medio (RMSE) es una métrica comúnmente utilizada para medir la diferencia entre los valores reales y los predichos en el análisis de regresión. El error absoluto medio (MAE) mide la magnitud promedio de los errores entre los valores predichos y reales. El coeficiente de determinación (R^2) es una medida de la proporción de variabilidad en la variable dependiente que se explica por las variables independientes en un modelo de regresión. Varía de 0 a 1; los valores más altos indican un mejor ajuste entre los valores predichos y los reales.

5.4.- Metodología

Se puede definir a la presente investigación como un estudio de campo experimental con diseño exploratorio y comparativo de métodos. Busca recolectar datos hiperlocales mediante campañas móviles y sensores temporales, calibrar sensores de bajo costo con estaciones de referencia usando técnicas de inteligencia artificial, generar mapas espaciales y predicciones robustas.

Fase 1: Diseño y preparación

Se definirá exactamente donde se hará el estudio (área piloto en Culiacán), que se va a medir y como se va a organizar el trabajo. Se eligen los polígonos representativos (avenidas principales, zonas comerciales, barrios residenciales y parques), se trazan las rutas de muestreo, se solicitan permisos si son necesarios y se revisa el equipo (sensores, GPS y respaldo informático).

Fase 2: Campañas móviles de muestreo

Se ejecutarán las salidas a campo con los sensores montados en vehículo, bicicleta o a pie. Cada ruta se recorre varias veces en distintos horarios y cada pasada registra lecturas de PM2.5/PM10 junto con la ubicación GPS y condiciones (temperatura, humedad). Es importante mantener montaje y velocidad lo más constantes posible y anotar incidencias (obras, humo, atascos).

Fase 3: Despliegue de sensores temporales (estáticos)

Al inicio de las campañas se colocarán varios sensores estáticos en puntos estratégicos del área piloto. Esos sensores permanecen funcionando durante toda la campaña y servirán para dos cosas: validar las lecturas móviles (compararlas) y crear pares de datos con las estaciones oficiales.

Fase 4: Preprocesamiento de datos

Una vez recogidos, los datos se pasarán a la computadora y se organizarán sincronizando las marcas de tiempo, enlazando lecturas con las coordenadas del GPS y eliminando lecturas claramente erróneas. Además, cada punto se enriquecerá con información adicional que ayuda al modelado, por ejemplo, el índice de vegetación, distancia a la vía más cercana, tipo de uso del suelo y condiciones meteorológicas.

Fase 5: Calibración de sensores low-cost

Dado que los sensores low-cost pueden tener sesgos, se creará un conjunto de pares que reúne lecturas del sensor y lecturas simultáneas de la estación de referencia. Con esos pares se entrenarán modelos supervisados (por ejemplo, XGBoost o Random Forest) que aprendan a corregir la lectura cruda del sensor usando temperatura, humedad y hora del día como entradas. Se evaluará la mejora con métricas (RMSE, MAE, R^2).

Fase 6: Modelado espacial y generación de mapas hiperlocales

Con las lecturas calibradas se probarán técnicas para convertir puntos en mapas. Se trabajará con dos enfoques complementarios: métodos geoestadísticos como Kriging, que generan mapas y entregan una medida de incertidumbre en cada celda y modelos supervisados (XGBoost/Random Forest) que usan variables espaciales (índice de vegetación, distancias, densidad urbana) para predecir en cada celda de una malla urbana (por ejemplo de 100–250 m).

Fase 7: Validación espacial y temporal

Para saber que tan fiables son las predicciones se aplicarán pruebas de validación, como por ejemplo, validación temporal (entrenar con un conjunto de días y predecir días posteriores) y validación espacial (dejar fuera áreas completas del entrenamiento y evaluar allí). Estas pruebas miden si el modelo generaliza a nuevos tiempos y lugares dentro de la ciudad. Además, se reportarán métricas estándar (RMSE, MAE, R^2) y se analizará el desempeño por tipo de zona (avenida, residencial, parque).

Fase 8: Simulaciones de intervención

Con el modelo validado se crearán escenarios hipotéticos para estimar el impacto de medidas locales, por ejemplo, qué ocurre si se reduce el tráfico en cierta avenida en un 20%, o si se aumenta la vegetación en unas manzanas. Para ello se modificarán las variables (tráfico o índice de vegetación) en las celdas afectadas y se recalcularán las predicciones. La diferencia indicará la reducción estimada de PM2.5.

Fase 9: Gestión de datos y comunicación de resultados

Toda la cadena de datos y documentación generada se subirá a un repositorio, integrando anexos y descripciones que permitan a otros reproducir el estudio.

6.- Agenda

Semana 1 – Preparación inicial

- Revisión y selección del área piloto, prueba de sensores y GPS.

Semana 2 – Diseño de rutas y logísticas

- Trazado de rutas representativas y planificación de tiempos.

Semana 3 – Gestión de permisos

- Gestionar permisos para colocaciones temporales.

Semana 4 – Prueba piloto

- Salida piloto corta para detectar problemas operativos y ajustes en montaje.

Semana 5 – Campaña móvil 1 y despliegue de estáticos

- Despliegue de sensores estáticos en puntos estratégicos y realización de pasadas (mañana, tarde, horas no pico) por las rutas planificadas.

Semana 6 – Campaña móvil 2 (repetición)

- Nuevas pasadas para cubrir variabilidad temporal y acumulación de pares. Además de monitoreo de estado de los sensores estáticos.

Semana 7 – Campaña móvil 3 (recolección complementaria)

- Complementación de pasadas en días y horarios faltantes.

Semana 8 – Revisión de campo y cierre de recolección

- Revisión de la calidad de los datos y decidir si se requieren pasadas adicionales. Seguimiento de preparar la transferencia de datos al equipo de análisis.

Semana 9 – Preprocesamiento

- Limpieza de datos, sincronización temporal e identificación de outliers (valores atípicos). Complementado con enriquecimiento espacial (índice de vegetación, distancias, uso del suelo).

Semana 10 – Construcción del dataset

- Emparejamiento entre los sensores estáticos/móviles y estaciones oficiales.

Semana 11 – Entrenamiento de calibradores

- Entrenar y evaluar modelos de calibración (XGBoost, RF).

Semana 12 – Aplicación de calibrador

Aplicar calibrador a todo el dataset y exportarlo. Adicional se puede comparar el dataset pre/post calibración

Semana 13 – Modelado espacial preliminar

Entrenamiento de Kriging y XGBoost para la generación de mapas iniciales.

Semana 14 – Validación robusta

Ejecutar validación temporal y espacial. Ajuste final de modelos según resultados.

Semana 15 – Simulación y análisis de escenarios

Ejecutar escenarios hipotéticos y generar mapas comparativos con la estimación de impacto de intervenciones.

Semana 16 – Redacción y entrega final

Preparación de informe técnico y entrega de datos.

7.- Conclusiones

El análisis de la bibliografía revisada y el diseño metodológico propuesto permitieron sintetizar hallazgos y visualizar el impacto potencial de la intervención local. En primer lugar, la evidencia mostro que la combinación de muestreo móvil, sensores temporales de bajo costo y técnicas de inteligencia artificial es capaz de revelar heterogeneidades espaciales relevantes a escala de cuadra que las redes de monitoreo oficiales no captan, lo que facilita la identificación de “hotspots” de contaminación y zonas con mayores temperaturas locales. En segundo lugar, los estudios sobre calibración indicaron que los sensores de bajo costo, si se calibran mediante modelos supervisados (por ejemplo, XGBoost o Random Forest) pueden alcanzar una precisión operativa adecuada para complementar a la estaciones oficiales, ampliando su cobertura con un costo relativamente bajo. En tercer lugar, la literatura comparativa sobre modelos de predicción y mapeo resaltaron que no existe una única solución óptima; algunos ofrecen mejores estimaciones, otros explotan covariables urbanas con robustez y rapidez, y otros más proveen ventajas para predicciones horarias.

A partir de estos hallazgos, la propuesta desarrollada aporta una mejora concreta respecto a las prácticas existentes al priorizar la calibración robusta, cuantificación de incertidumbre y explicabilidad de resultados.

Se llego a la conclusión de que el impacto potencial en el medio ambiente y la gestión urbana es múltiple, es decir, producir mapas hiperlocales permitirá priorizar intervenciones donde reduzcan la exposición humana; las simulaciones de escenarios ofrecerán estimaciones cuantificadas de reducción de contaminantes, facilitando decisiones costo-efectivas; y la disponibilidad de datos calibrados y reproducibles fortalecerá la capacidad de monitoreo. En resumen, si se aplican con rigor las fases propuestas, la iniciativa proporcionará un marco metodológico replicable para otras ciudades, contribuyendo a una reducción medible de riesgos ambientales para la población.

Referencias

¿Cómo se mide la calidad del aire? (22 de Septiembre de 2022). Obtenido de UNEP:
<https://www.unep.org/es/noticias-y-reportajes/reportajes/como-se-mide-la-calidad-del-aire>

¿Qué es la calidad del aire? (08 de Julio de 2025). Obtenido de NASA:
<https://ciencia.nasa.gov/ciencias-terrestres/que-es-la-calidad-del-aire/>

Abuouelezz, W., Ali, N., ZeyarAung, Altunaiji, A., Shah, S. B., & Gliddon, D. (2025). Exploring PM2.5 and PM10 ML forecasting models: a comparative study in the UAE. *Scientific Reports*.

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., . . . Hamburg, S. P. (2017). High-Resolution Air Pollution Mapping with Google Street View Cars:. *Environmental Science & Technology*, 6999–7008.

Ravindra, K., Kumar, S., A. K., & Mor, S. (2024). Enhancing accuracy of air quality sensors with machine learning to augment large-scale monitoring networks. *npj climate and atmospheric science*.

Sistema Nacional de Información de la Calidad del Aire, SINAICA. (s.f.). Obtenido de SINAICA, INECC: <https://sinaica.inecc.gob.mx/>

Zhou, S., Wang, W., Zhu, L., Qiao, Q., & Kang, Y. (2024). Deep-learning architecture for PM2.5 concentration prediction: A review. *Environmental Science and Ecotechnology*.