

Tecnológico Nacional de México

**Plataforma Adaptativa de Inteligencia Artificial
para la Gestión Integrada y Predictiva de la
Calidad del Agua**

Materia:
Tópicos de Inteligencia Artificial

Integrantes:
Dávila Bejarano Víctor Jesús
Flores Medina Martin

Profesor:
Mora Félix Zuriel Dathan

12 de Septiembre del 2025

Repositorio de GitHub
<https://github.com/MartinMFM/topicosIA>

Contents

1	Introduction	4
2	Objetivo General	5
3	Objetivos Específicos	5
4	Justificación	5
5	Alcance	6
6	Antecedentes	6
6.1	Planteamiento General y Ventajas	6
6.2	Modelos de IA para Predicción y Clasificación del ICA	7
6.3	Tendencias y Modelos Híbridos	7
6.4	Fuentes de Datos Innovadoras: Teledetección	7
6.5	Aplicaciones Avanzadas: Rastreo de Fuentes de Contaminación	8
7	Propuesta	9
7.1	Problema Ambiental Abordado	9
7.2	Soluciones Actuales de IA y su Efectividad	9
7.3	Solución de IA Propuesta	9
7.4	Impacto Potencial de la Solución en el Medio Ambiente	10
8	Desarrollo	10
8.1	Arquitectura Conceptual de la Plataforma Integrada	10
8.1.1	Visión General del Sistema de Extremo a Extremo	10
8.1.2	Flujo de Datos y Lógica de Interacción	11
8.2	Módulo de Adquisición de Datos en Tiempo Real (IoT)	11
8.2.1	Selección de Parámetros y Sensores	11
8.2.2	Diseño del Nodo Sensor (Hardware)	12
8.2.3	Protocolos de Comunicación	12
8.3	Canalizaciones de Pre-procesamiento de Datos (Data Pipelines) .	12
8.4	Núcleo de Inteligencia Artificial: Modelado Predictivo y Adaptativo	12
8.4.1	Modelo Predictivo Basado en RNA para el Índice de Calidad del Agua (ICA)	13
8.4.2	Sensor Virtual Inteligente para la Detección de Contaminación Microbiológica	14
8.4.3	Mecanismos de Adaptación y Simulación de Escenarios .	15
8.5	Desarrollo del Modelo y Criterios de Validación	15
8.5.1	Protocolo para la Partición de Datos	15
8.5.2	Métricas de Evaluación del Rendimiento	16
8.5.3	Criterios para la Selección del Modelo Óptimo	16
8.6	Interfaz de Usuario y Sistema de Visualización	16
8.6.1	Diseño del Panel de Control (Dashboard)	16

8.6.2	Componentes de Visualización	17
8.6.3	Sistema de Alarmas Multicapa	17
9	Agenda de Presentación	18
10	Conclusiones	19
	Referencias	20

1 Introduction

La degradación de la calidad del agua es uno de los desafíos ambientales más críticos a nivel global. El incremento de la población y la actividad industrial y agrícola han provocado un acelerado deterioro de los recursos hídricos [1, 2]. La gestión tradicional de estos recursos, basada en muestreos periódicos y análisis de laboratorio, resulta a menudo costosa, lenta y reactiva, siendo insuficiente para anticipar y mitigar eventos de contaminación de manera eficaz, especialmente en sistemas de Agua Potable Rural (APR) con presupuestos limitados y sistemas sin vigilancia continua [3].

En los últimos años, la **Inteligencia Artificial** ha emergido como un campo prometedor para abordar esta problemática [4, 5]. Técnicas como las **Redes Neuronales Artificiales (RNA)** han demostrado una alta eficacia para modelar las complejas relaciones no lineales entre los parámetros fisicoquímicos y predecir Índices de Calidad del Agua (ICA) [1, 5]. Sin embargo, la mayoría de estas soluciones se enfocan en problemas específicos, carecen de adaptabilidad a diferentes usos del agua y no suelen estar integradas en plataformas accesibles para los gestores ambientales que no son expertos en IA [2]. Este proyecto propone el diseño conceptual de una plataforma que integra múltiples técnicas de IA para ofrecer una herramienta de gestión hídrica proactiva, adaptativa y de bajo costo.

2 Objetivo General

Desarrollar una plataforma conceptual basada en técnicas de inteligencia artificial que permita el monitoreo, análisis predictivo y gestión adaptativa de la calidad del agua, integrando modelos de aprendizaje automático y esquemas de visualización intuitivos, con el fin de anticipar eventos de contaminación y apoyar la toma de decisiones en la gestión sostenible de los recursos hídricos.

3 Objetivos Específicos

- Definir los parámetros críticos de calidad del agua que serán monitoreados por la plataforma (pH, turbidez, conductividad, oxígeno disuelto, entre otros).
- Definir la arquitectura de un modelo de Red Neuronal Artificial (RNA) para la predicción del Índice de Calidad del Agua (ICA) y la simulación de escenarios de contaminación.
- Diseñar la arquitectura del módulo de adquisición de datos en tiempo real (IoT), especificando los componentes de hardware (sensores, microcontrolador, SBC) y los protocolos de comunicación para la transmisión de datos.
- Diseñar una interfaz de usuario amigable y accesible que contemple la visualización de indicadores en tiempo real, el historial de datos y las predicciones del sistema.
- Promover la gestión sostenible del recurso hídrico mediante la adopción de tecnologías inteligentes accesibles y de bajo costo, que reduzcan riesgos sanitarios y ambientales.

4 Justificación

El agua es un recurso vital para la vida y el desarrollo socioeconómico. Su disponibilidad y calidad están directamente relacionadas con la salud pública, la seguridad alimentaria y el equilibrio de los ecosistemas. Sin embargo, el incremento de la población, las actividades industriales y agrícolas, así como la urbanización no planificada, han generado un deterioro progresivo en la calidad de los cuerpos de agua [6]. Las metodologías tradicionales de monitoreo hídrico, basadas en muestrerios periódicos y análisis de laboratorio, presentan limitaciones importantes: suelen ser costosas, requieren de personal especializado y no permiten anticipar eventos críticos de contaminación en tiempo real [3]. Esto se agrava en comunidades rurales o con recursos limitados, donde la ausencia de vigilancia continua aumenta la vulnerabilidad frente a riesgos sanitarios. En este contexto, la **Inteligencia Artificial** se presenta como una herramienta estratégica capaz de transformar la gestión del recurso hídrico. Estudios recientes

muestran que técnicas como las Redes Neuronales Artificiales, la lógica difusa y los modelos híbridos permiten predecir índices de calidad del agua y detectar contaminantes con alta precisión [2, 1].

Es precisamente en esta intersección entre la necesidad y la oportunidad donde se enmarca este proyecto, a través del diseño de una plataforma **integrada, adaptativa y de bajo costo**. Al aprovechar las ventajas de la IA para el monitoreo y la gestión predictiva, se busca ofrecer una herramienta que permita optimizar la toma de decisiones, reducir riesgos sanitarios y ambientales, y generar un impacto positivo tangible en las comunidades con mayores limitaciones.

5 Alcance

El presente proyecto se limita al desarrollo de un diseño conceptual y a la revisión teórica de las herramientas de inteligencia artificial aplicadas al monitoreo y gestión de la calidad del agua. No contempla la implementación de prototipos físicos ni el desarrollo de software operativo, sino que se enfoca en la construcción de un marco metodológico y en la propuesta de una arquitectura que pueda servir de base para futuros trabajos de desarrollo e implementación práctica.

6 Antecedentes

El uso de la Inteligencia Artificial ha demostrado ser una herramienta poderosa y versátil para la evaluación, monitoreo y predicción de la calidad del agua, superando muchas de las limitaciones de los métodos tradicionales. La literatura reciente explora diversas facetas de esta aplicación, desde la predicción de índices de calidad hasta el rastreo de fuentes de contaminación, utilizando una variedad de modelos y fuentes de datos.

6.1 Planteamiento General y Ventajas

Los métodos convencionales para evaluar la calidad del agua, que dependen de la recolección de muestras y análisis de laboratorio, son a menudo costosos, lentos y no permiten un monitoreo continuo en tiempo real [6, 7]. La combinación del Internet de las Cosas (IoT) y la IA surge como una alternativa eficaz [7]. Los dispositivos IoT, como sensores, capturan datos en tiempo real, mientras que las técnicas de IA, especialmente las Redes Neuronales Artificiales (ANN), analizan estos datos para modelar y predecir el comportamiento de sistemas complejos y no lineales [7]. Este enfoque no solo es más rápido y eficiente, sino que también puede operar con datos incompletos, a diferencia de las fórmulas empíricas tradicionales que fallan si falta la medición de algún parámetro [6].

6.2 Modelos de IA para Predicción y Clasificación del ICA

La predicción del Índice de Calidad del Agua (ICA) es una de las aplicaciones más comunes de la IA en este campo.

- Se ha comprobado la eficacia de las redes neuronales para calcular el ICA en ríos. En estudios de caso, modelos como la Red Neuronal Feedforward (FNN) —con una capa oculta y 5 neuronas han logrado un coeficiente de correlación (R) de 0.99 en la fase de entrenamiento [6].
- Se han utilizado enfoques duales que combinan diferentes modelos para tareas específicas. Por ejemplo, un sistema ANFIS (Sistema de Inferencia Neuro-Difuso Adaptativo) ha sido empleado para predecir el valor numérico del ICA, mientras que una Red Neuronal Feedforward (FFNN) se utilizó para clasificar la calidad del agua en categorías. Este método ha demostrado ser muy exitoso, alcanzando un coeficiente de regresión del 96.17% en la predicción y una precisión del 100% en la clasificación [8].
- En comparativas entre diferentes arquitecturas, como ANN y LSTM (Memoria a Corto-Largo Plazo), se ha concluido que el modelo LSTM puede ofrecer mayor precisión para la predicción de la calidad del agua, alcanzando una exactitud del 95% [9].

6.3 Tendencias y Modelos Híbridos

Un análisis sistemático de la literatura de la última década (2011-2020) revela varias tendencias clave [10]. Los modelos ANFIS y ANN son los más utilizados en la investigación sobre calidad del agua superficial. Sin embargo, los modelos híbridos, como Wavelet-ANN (W-ANN) y Wavelet-ANFIS (W-ANFIS), han demostrado ser los más precisos [10]. El estudio también encontró que la Demanda Bioquímica de Oxígeno (DBO) es el parámetro más investigado y que la mayoría de los estudios provienen de Irán y el sudeste asiático [10].

6.4 Fuentes de Datos Innovadoras: Teledetección

Más allá de los sensores in-situ, la IA permite el uso de fuentes de datos innovadoras como la teledetección. Se han utilizado imágenes satelitales de Landsat 8 para estimar el ICA del río Hudson [11]. En dichos estudios, se correlacionaron los índices espectrales de las imágenes con 13 parámetros de calidad del agua medidos en campo y luego se usaron modelos de IA para predecir el ICA a partir de los datos satelitales [11]. El modelo MARS (Multivariate Adaptive Regression Spline) demostró ser el más eficaz [11]. Este enfoque es particularmente valioso por ser una solución práctica y económica que reduce la necesidad de muestreos físicos costosos y laboriosos [11].

6.5 Aplicaciones Avanzadas: Rastreo de Fuentes de Contaminación

Un avance significativo más allá de la simple predicción es el rastreo de las fuentes de contaminación [12]. Se han propuesto sistemas de IA que utilizan una combinación de correlación cruzada, reglas de asociación (Apriori) y una red de memoria a corto-largo plazo (LSTM) para identificar las características de la contaminación y rastrear su origen industrial [12]. Estos sistemas han podido rastrear con precisión las fuentes puntuales de futuros cambios en la calidad del agua con una exactitud de entre 95% y 100% en diferentes cuencas [12]. Esta capacidad representa un salto cualitativo, pasando de un monitoreo pasivo a una gestión ambiental proactiva y fiscalizadora.

7 Propuesta

Con base en la revisión de la literatura y la identificación de las brechas existentes, se presenta una propuesta formal para el desarrollo conceptual de la plataforma.

7.1 Problema Ambiental Abordado

La degradación de los recursos hídricos superficiales es un problema ambiental crítico, agravado por la presión demográfica y las actividades industriales y agrícolas. Las metodologías de monitoreo tradicionales son insuficientes para una gestión eficaz, ya que son costosas, lentas y reactivas [6, 7]. Esta ineficiencia pone en riesgo la salud pública y el equilibrio de los ecosistemas, afectando de manera desproporcionada a comunidades con recursos limitados que carecen de sistemas de vigilancia continua.

7.2 Soluciones Actuales de IA y su Efectividad

La Inteligencia Artificial ha demostrado ser una solución robusta y eficaz para el monitoreo y predicción de la calidad del agua. La literatura científica confirma que modelos como las Redes Neuronales Artificiales (ANN) pueden predecir Índices de Calidad del Agua (ICA) con altísima precisión ($R > 0.99$) [6] y clasificar el estado del agua con una exactitud de hasta el 100% [8]. Adicionalmente, aplicaciones avanzadas han logrado integrar datos de teledetección para monitoreo a gran escala [11] e incluso han desarrollado sistemas capaces de rastrear las fuentes industriales de contaminación en tiempo real con una precisión superior al 95% [12].

7.3 Solución de IA Propuesta

Nuestra propuesta es el diseño conceptual de una **Plataforma Adaptativa de Inteligencia Artificial para la Gestión Integrada y Predictiva de la Calidad del Agua**. A diferencia de las soluciones aisladas, esta plataforma integra cuatro capas funcionales (adquisición IoT, comunicación, procesamiento en la nube y presentación) en un sistema coherente.

La principal mejora radica en su **enfoque integrado y multifuncional**:

- **Núcleo de IA Dual:** Combina un modelo de Red Neuronal Artificial (RNA) para la predicción numérica precisa del ICA con un "sensor virtual" basado en lógica difusa para la estimación en tiempo real del riesgo microbiológico, un parámetro crítico que los métodos tradicionales no pueden medir instantáneamente [3].
- **Plataforma de Bajo Costo:** Utiliza hardware asequible y de código abierto (Arduino, Raspberry Pi) y protocolos de comunicación eficientes (MQTT), haciendo la tecnología accesible para entidades con presupuestos limitados [7].

- **Enfoque en la Usabilidad:** Incorpora un dashboard interactivo diseñado para gestores de recursos hídricos, no para expertos en IA, con visualizaciones claras, mapas geoespaciales y un sistema de alertas multicapa.
- **Capacidad de Simulación Estratégica:** Permite la simulación de escenarios "qué pasaría si", transformando la plataforma en una herramienta de planificación proactiva y respuesta a emergencias [1].

7.4 Impacto Potencial de la Solución en el Medio Ambiente

El impacto potencial de esta plataforma es multifacético y significativo.

Transición a una Gestión Proactiva: Permitirá a los gestores pasar de reaccionar a crisis de contaminación a **anticiparlas y prevenirlas**, gracias a las predicciones en tiempo real y al sistema de alerta temprana.

Protección de la Salud Pública: Al detectar riesgos microbiológicos y químicos de manera instantánea, se pueden tomar medidas inmediatas para evitar la propagación de enfermedades transmitidas por el agua.

Democratización del Monitoreo Avanzado: Su diseño de bajo costo hará posible que comunidades rurales y regiones con recursos limitados implementen por primera vez un sistema de vigilancia de la calidad del agua robusto y continuo.

Optimización de Recursos y Sostenibilidad: Al proporcionar datos precisos y predicciones fiables, la plataforma ayudará a los responsables a tomar decisiones más informadas sobre inversiones en infraestructura y aplicación de regulaciones, promoviendo una gestión más sostenible de los recursos hídricos.

8 Desarrollo

8.1 Arquitectura Conceptual de la Plataforma Integrada

8.1.1 Visión General del Sistema de Extremo a Extremo

La plataforma se basa en una arquitectura modular de cuatro capas interconectadas: Adquisición de Datos (IoT), Comunicación, Procesamiento en la Nube (IA) y Presentación. Este diseño garantiza la escalabilidad, robustez y mantenibilidad del sistema [3].

Las cuatro capas funcionales son:

1. **Capa 1: Adquisición de Datos (IoT):** Red de nodos de sensores inteligentes desplegados en puntos de muestreo estratégicos para la recolección autónoma de datos fisicoquímicos en tiempo real.

2. **Capa 2: Comunicación y Transporte de Datos:** Infraestructura de red inalámbrica (3G/4G o Wi-Fi) para la transmisión segura y eficiente de los datos desde los nodos IoT hasta la plataforma central.
3. **Capa 3: Procesamiento en la Nube y Núcleo de IA:** Infraestructura en la nube responsable del almacenamiento, pre-procesamiento y análisis predictivo de los datos mediante modelos de inteligencia artificial.
4. **Capa 4: Presentación y Soporte a la Decisión:** Interfaz de usuario web (dashboard) que traduce los resultados de la IA en visualizaciones intuitivas, informes y un sistema de alertas proactivo para los gestores.

8.1.2 Flujo de Datos y Lógica de Interacción

El flujo de información es secuencial y automatizado para maximizar la eficiencia. El proceso comienza en el nodo sensor, donde un microcontrolador (Arduino UNO) recolecta los datos de los sensores. Estos datos son enviados a un computadora de placa única (Raspberry Pi) en el mismo nodo, que realiza un pre-procesamiento local (agregación, validación, estampado de tiempo y geolocalización).

Posteriormente, los datos se empaquetan y transmiten utilizando el protocolo **MQTT (Message Queuing Telemetry Transport)**, seleccionado por su ligereza y robustez ante conexiones intermitentes [3]. Los mensajes son enviados a un *broker* MQTT en la nube (IBM Cloud), que actúa como punto de entrada.

En la nube, los datos se almacenan en bases de datos de series temporales y de documentos. Simultáneamente, activan el **Núcleo de Inteligencia Artificial**, que ejecuta los modelos predictivos en tiempo real. Finalmente, la interfaz de usuario consulta los resultados procesados para actualizar el *dashboard* y, si se superan ciertos umbrales, activar el sistema de alertas.

8.2 Módulo de Adquisición de Datos en Tiempo Real (IoT)

8.2.1 Selección de Parámetros y Sensores

La selección de parámetros se basa en la literatura científica y normativas como el NSF WQI [5, 1]. Estos parámetros constituyen la base de entrada para el modelo de predicción del Índice de Calidad del Agua.

Parámetros Primarios (Núcleo del Modelo Predictivo):

- **Oxígeno Disuelto (OD):** Indicador clave de la salud del ecosistema.
- **Demand Bioquímica de Oxígeno (DBO):** Medida de contaminación orgánica.
- **Nitratos:** Indicador de contaminación por nutrientes.
- **Coliformes Fecales:** Parámetro microbiológico esencial.
- **pH (Potencial de Hidrógeno):** Mide la acidez o alcalinidad.

- **Turbidez:** Medida de la claridad del agua.

Estos seis parámetros forman el vector de entrada del modelo ANN para la predicción del NSF WQI [5].

8.2.2 Diseño del Nodo Sensor (Hardware)

El diseño del nodo busca un equilibrio entre bajo costo, robustez y fiabilidad. El microcontrolador Arduino UNO, seleccionado por su bajo costo, robustez y amplio soporte comunitario, actúa como colector de datos primario, interactuando directamente con los sensores. La computadora de placa única Raspberry Pi 4 funciona como el gateway del nodo, gestionando la conexión a redes, el pre-procesamiento de datos y la ejecución del cliente MQTT para la transmisión segura de datos [3]. Finalmente, se contempla una solución de bajo consumo para la fuente de alimentación, con opción a alimentación solar, junto con una carcasa con certificación IP67 para resistir condiciones ambientales adversas.

8.2.3 Protocolos de Comunicación

Se utilizará el protocolo MQTT por ser ligero y eficiente, lo que lo hace ideal para redes IoT con ancho de banda limitado o intermitente [3]. Además, su mecanismo de Calidad de Servicio (QoS) garantiza la entrega de mensajes. La conectividad se logrará principalmente mediante un módem 3G/4G o, de manera alternativa, a través de Wi-Fi en aquellas ubicaciones que cuenten con infraestructura de red disponible.

8.3 Canalizaciones de Pre-procesamiento de Datos (Data Pipelines)

Para asegurar la fiabilidad de los modelos de IA, se implementará una canalización de pre-procesamiento automatizada que incluye:

1. **Limpieza de Datos:** Algoritmos para la detección y manejo de valores atípicos, imputación de datos faltantes y aplicación de filtros (ej. media móvil) para reducir el ruido del sensor [2].
2. **Normalización de Datos:** Los datos se normalizarán (ej. min-max) para escalar todas las variables a un rango uniforme como '[-1, 1]', un paso necesario para optimizar el entrenamiento de la RNA [5].
3. **Selección y Validación de Características:** Se utilizarán técnicas como matrices de correlación para validar la relevancia de las variables de entrada, optimizando el conjunto de características para el modelo.

8.4 Núcleo de Inteligencia Artificial: Modelado Predictivo y Adaptativo

El núcleo de IA integra un conjunto de técnicas para proporcionar una visión completa y proactiva de la calidad del agua.

8.4.1 Modelo Predictivo Basado en RNA para el Índice de Calidad del Agua (ICA)

Para la predicción del Índice de Calidad del Agua (ICA), se optó por un modelo basado en Redes Neuronales Artificiales (RNA) debido a la naturaleza del problema. La relación entre los diversos parámetros fisicoquímicos y un índice compuesto como el ICA es inherentemente compleja, multivariada y no lineal. Las RNA son excepcionalmente adecuadas para modelar estas dinámicas, ya que su principal fortaleza radica en la capacidad de aprender patrones complejos directamente de los datos, sin necesidad de definir explícitamente las relaciones matemáticas subyacentes. Esto las posiciona como una herramienta superior a los modelos estadísticos tradicionales en este dominio [2, 5].

- **Arquitectura:** Se implementará un modelo de **Perceptrón Multicapa (MLP)**, una de las arquitecturas de RNA más robustas y estudiadas. Basándose en los resultados exitosos de investigaciones previas en la estimación del ICA en ríos, se adoptará una arquitectura de tres capas como punto de partida: una capa de entrada, una capa oculta y una capa de salida, como se muestra en la Figura 1. La configuración inicial, sujeta a un proceso de optimización de hiperparámetros, será de **6-12-1**: seis neuronas en la capa de entrada (correspondientes a los seis parámetros primarios), doce neuronas en la capa oculta (un número que ha demostrado un buen equilibrio entre capacidad de aprendizaje y generalización), y una única neurona en la capa de salida [5].
- **Entrenamiento:** El proceso de aprendizaje de la red se llevará a cabo mediante un algoritmo de retropropagación. Específicamente, se utilizará el algoritmo de **Levenberg-Marquardt**, que ha demostrado en la práctica ser uno de los más eficientes y rápidos para el entrenamiento de redes MLP de tamaño moderado, convergiendo a una solución óptima con menos iteraciones que otros métodos como el descenso de gradiente. Las funciones de activación, que introducen la no linealidad necesaria para aprender relaciones complejas, serán de tipo **tangente sigmoidea (tansig)** tanto para la capa oculta como para la capa de salida, conforme a las configuraciones validadas en estudios de referencia [5].
- **Salida:** La neurona de la capa de salida proporcionará una predicción numérica directa del valor del **NSF WQI** en una escala de 0 a 100. Este valor cuantitativo será posteriormente mapeado a una escala cualitativa para una interpretación más sencilla por parte de los usuarios, utilizando las categorías estándar: Muy mala (0-25), Mala (26-50), Media (51-70), Buena (71-90) y Excelente (91-100) [5].

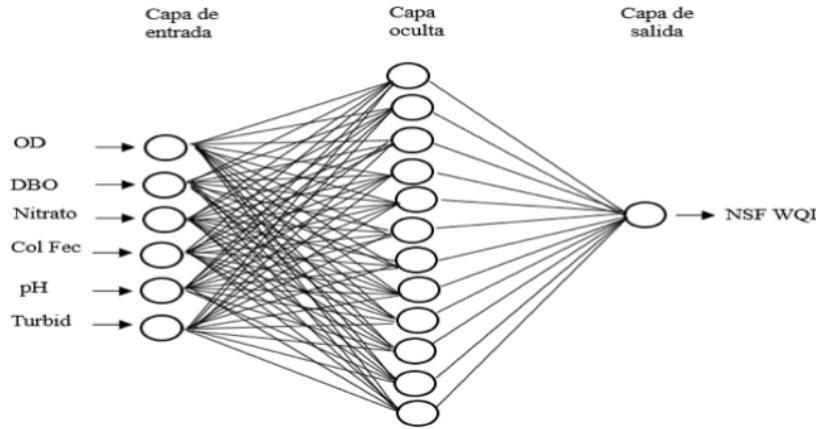


Figure 1: Arquitectura de la Red Neuronal Artificial (RNA) para la estimación del Índice de Calidad del Agua (NSF WQI) [5].

8.4.2 Sensor Virtual Inteligente para la Detección de Contaminación Microbiológica

El monitoreo de la contaminación microbiológica es fundamental para la salud pública, pero su análisis directo mediante cultivo de muestras presenta series limitaciones. Este proceso es lento (tarda entre 24 y 48 horas), costoso y requiere personal especializado, lo que lo hace inviable para una vigilancia continua y en tiempo real, especialmente en sistemas con recursos limitados como los de Agua Potable Rural [3]. Para superar esta barrera, la plataforma implementará un "sensor virtual" inteligente. Este componente, basado en IA, está diseñado para ofrecer una estimación del riesgo microbiológico de forma instantánea y a un costo significativamente menor.

- **Metodología:** La técnica seleccionada para este sensor virtual es un sistema de lógica difusa de tipo Mamdani. Este enfoque es particularmente adecuado porque está diseñado para manejar la incertidumbre, la ambigüedad y la imprecisión, que son características inherentes al proceso de inferir un estado complejo (como la presencia de contaminación biológica) a partir de mediciones indirectas y a menudo correlacionadas [3].
- **Entradas y Funciones de Pertenencia:** El sistema difuso utilizará como entradas los valores normalizados de cuatro parámetros clave: **Turbidez, ORP, Temperatura y TDS**. Para cada una de estas entradas, se definirán funciones de pertenencia (por ejemplo, triangulares o trapezoidales) que traducen el valor numérico nítido a un grado de pertenencia

a conjuntos difusos lingüísticos. Por ejemplo, para la variable Temperatura, los conjuntos difusos podrían ser 'Fría', 'Tibia' y 'Caliente', y un valor de 18°C podría tener un grado de pertenencia de 0.8 a 'Tibia' y 0.2 a 'Fría' [3].

- **Base de Reglas y Salida:** El conocimiento del dominio se codificará en una base de reglas heurísticas en formato "SI-ENTONCES". Estas reglas emulan el razonamiento de un experto humano. Un ejemplo de regla podría ser: "**SI** la *Turbidez* es *Alta* **Y** la *Temperatura* es *Tibia* **Y** el *ORP* es *Bajo* **ENTONCES** el *Riesgo Microbiológico* es *Alto*". El motor de inferencia del sistema evaluará todas las reglas relevantes y combinará sus resultados para producir una salida. La salida será un valor numérico de riesgo de contaminación en una escala de 0 a 10, que a su vez se clasificará en categorías lingüísticas como 'Bajo', 'Medio', 'Alto', 'Indeterminado' (para datos contradictorios) y 'Otras Partículas' (cuando la turbidez se atribuye a sólidos y no a biomasa) [3].

8.4.3 Mecanismos de Adaptación y Simulación de Escenarios

- **Modelos Híbridos:** El diseño permitirá la futura integración de modelos híbridos (ej. RNA-ARIMA) para mejorar la precisión predictiva, combinando la capacidad de modelado no lineal con el análisis de series temporales [2].
- **Simulación de Escenarios "Qué Pasaría Si":** Una vez entrenado, el modelo RNA servirá como motor de simulación. Los usuarios podrán introducir eventos hipotéticos (ej. un vertido) para predecir el impacto en el ICA, transformando la plataforma en una herramienta de planificación estratégica [1].
- **Optimización Adaptativa:** Se explorará el uso de **Algoritmos Genéticos (AG)** para optimizar los hiperparámetros de la RNA y para refinar la base de reglas del sistema de lógica difusa, permitiendo que el sistema mejore su rendimiento de forma autónoma con el tiempo [2, 4].

8.5 Desarrollo del Modelo y Criterios de Validación

Se seguirá un protocolo riguroso para garantizar la fiabilidad del modelo predictivo.

8.5.1 Protocolo para la Partición de Datos

Para evitar el sobreajuste y asegurar la generalización del modelo, el conjunto de datos se dividirá de forma aleatoria en tres subconjuntos, siguiendo la distribución validada en la literatura [5]:

- **Conjunto de Entrenamiento (70%):** Utilizado para el ajuste de los pesos y sesgos de la red neuronal.

- **Conjunto de Validación (10%):** Usado para el ajuste de hiperparámetros y la aplicación de técnicas de *early stopping* para prevenir el sobreajuste.
- **Conjunto de Prueba (20%):** Mantenido aislado hasta el final del proceso para proporcionar una evaluación imparcial del rendimiento del modelo final en datos no vistos.

8.5.2 Métricas de Evaluación del Rendimiento

Se utilizarán métricas estadísticas estándar para cuantificar el rendimiento del modelo [5]:

- **Raíz del Error Cuadrático Medio (RMSE):** Mide la magnitud promedio de los errores de predicción.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - q_i)^2}{n}}$$

- **Coeficiente de Correlación (R):** Mide la fuerza de la relación lineal entre los valores predichos y los observados. Un valor cercano a 1 indica un ajuste casi perfecto.

$$R = \frac{\sum_{i=1}^n (Q_i - \bar{Q})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2(q_i - \bar{q})^2}}$$

donde Q_i son los valores observados, q_i los predichos, y n el número de muestras.

8.5.3 Criterios para la Selección del Modelo Óptimo

El modelo óptimo será aquel que presente la mejor combinación del **RMSE más bajo** y el **coeficiente R más alto** en el **conjunto de validación**. El rendimiento final y reportado del modelo será el obtenido sobre el **conjunto de prueba**, proporcionando una medida imparcial de su capacidad de generalización. Se tomará como referencia de éxito alcanzar un $R \geq 0.94$, como en estudios previos [5].

8.6 Interfaz de Usuario y Sistema de Visualización

8.6.1 Diseño del Panel de Control (Dashboard)

La interfaz de usuario (dashboard) se diseñará para ser intuitiva y comunicar los resultados de la IA de forma clara a los gestores del agua, quienes no necesariamente son expertos en ciencia de datos.

8.6.2 Componentes de Visualización

El *dashboard* incluirá varios componentes clave:

- **Vista en Tiempo Real:** Medidores y gráficos que muestren los valores actuales de los parámetros y del ICA predicho, con actualizaciones en tiempo real.
- **Vista Histórica:** Gráficos interactivos para el análisis de tendencias, permitiendo seleccionar rangos de fechas y superponer parámetros.
- **Mapa Geoespacial:** Un mapa interactivo con la ubicación de los nodos sensores, coloreados según el estado actual de la calidad del agua para identificar rápidamente puntos críticos.
- **Generación de Informes y Exportación de Datos:** Funcionalidad para generar informes y descargar datos históricos en formatos estándar (CSV, Excel) para análisis externos o cumplimiento normativo [3].

8.6.3 Sistema de Alarmas Multicapa

Se implementará un sistema de alerta temprana automatizado para notificar sobre eventos críticos, permitiendo una respuesta proactiva. Las alarmas se clasificarán de la siguiente manera [3]:

- **Alarma Química:** Se activará si parámetros como el pH salen de su rango operativo normal.
- **Alarma Física:** Se activará si la turbidez o la conductividad superan umbrales establecidos.
- **Alarma Microbiológica:** Se activará directamente desde la salida del sensor virtual de lógica difusa, con diferentes niveles de severidad (advertencia, alerta crítica).

Las notificaciones se enviarán a través de múltiples canales como **correo electrónico** y/o **mensajería SMS** para asegurar su recepción.

9 Agenda de Presentación

Table 1: Cronograma de Actividades del Proyecto (Diseño y Desarrollo)

Fase	Actividad	Semanas
Fase 1: Planificación y Diseño Inicial (Semanas 1-4)		
Planificación	Definición detallada de objetivos, alcance y metodología.	Semana 1
Revisión Bibliográfica	Investigación sobre arquitecturas de sistemas IoT y modelos de IA para calidad del agua.	Semanas 1-2
Diseño de Arquitectura	Elaboración del diseño conceptual de la plataforma y selección de tecnologías (hardware, protocolos, nube).	Semanas 3-4
Fase 2: Desarrollo e Implementación (Semanas 5-13)		
Desarrollo del Hardware (IoT)	Ensamblaje y prototipado del nodo sensor físico (Arduino, Raspberry Pi, sensores).	Semanas 5-6
Desarrollo de Software (Nodo IoT)	Programación del firmware para la recolección, pre-procesamiento y transmisión de datos (MQTT).	Semanas 7-8
Desarrollo del Backend	Configuración de la infraestructura en la nube (Bases de Datos, Servicios IoT) y desarrollo de las canalizaciones de datos.	Semanas 9-10
Desarrollo del Núcleo de IA	Implementación y entrenamiento de los modelos de IA (Red Neuronal para ICA y Lógica Difusa para sensor virtual).	Semanas 10-12
Desarrollo del Frontend	Implementación del dashboard, componentes de visualización (gráficos, mapas) y sistema de alertas.	Semanas 11-13
Fase 3: Integración y Pruebas (Semanas 14-15)		
Pruebas e Integración	Pruebas unitarias de cada componente (IoT, Backend, IA, Frontend) y pruebas de integración del sistema completo.	Semana 14
Validación y Ajustes	Validación del flujo de datos de extremo a extremo y ajustes finales del sistema.	Semana 15
Fase 4: Documentación y Entrega Final (Semanas 15-16)		
Documentación Final	Redacción del informe final del proyecto, incluyendo análisis de resultados y conclusiones.	Semanas 15-16

10 Conclusiones

Este proyecto demuestra que es posible dejar atrás las limitaciones de los métodos tradicionales de monitoreo del agua. Estos métodos suelen ser costosos, lentos y, por su naturaleza reactiva, a menudo se aplican cuando el problema ya ha escalado. Aquí la propuesta es distinta: una plataforma que une IoT con inteligencia artificial para adelantarse a los riesgos.

La innovación no está en usar un solo algoritmo, sino en la manera en que todo el sistema se estructura. Se combinan equipos accesibles como Arduino y Raspberry Pi con dos módulos de IA: por un lado, una red neuronal que predice con buena precisión el Índice de Calidad del Agua (ICA), y por otro, un sensor virtual basado en lógica difusa que puede detectar riesgos microbiológicos casi en tiempo real. El resultado es una herramienta potente y, al mismo tiempo, fácil de implementar.

Además de vigilar, la plataforma abre la puerta a algo más: probar escenarios hipotéticos. En otras palabras, permite responder preguntas como “¿qué pasaría si entra un contaminante en el río?” o “¿cómo cambiaría la calidad si aumentan las lluvias?”. Esto convierte el sistema en un apoyo real para la toma de decisiones, ya que las políticas de mitigación se diseñan con datos y no con suposiciones.

No obstante, su mayor valor radica en la capacidad de llevar esta tecnología a quienes más la necesitan. Al reducir costos y complejidad, se vuelve una alternativa viable para comunidades con recursos limitados, como los sistemas de Agua Potable Rural (APR). En estos lugares, contar con alertas tempranas puede marcar la diferencia entre prevenir un riesgo sanitario o enfrentarlo cuando ya es tarde.

References

- [1] C. G. López and E. C. Artola. “Una herramienta de inteligencia artificial para el modelado de distintos escenarios con la determinación del Índice de Calidad del Agua (ICA) para aplicar a ríos urbanos de Argentina”. In: *XXV Workshop de Investigadores en Ciencias de la Computación* (Apr. 2023). 13 y 14 de abril de 2023, pp. 1–6.
- [2] Iván Darío López, Apolinar Figueroa, and Juan Carlos Corrales. “Un mapeo sistemático sobre predicción de calidad del agua mediante técnicas de inteligencia computacional”. In: *Revista Ingenierías Universidad de Medellín* 15.28 (2016), pp. 35–52.
- [3] Alvaro Conejeros Molina et al. “Monitoreo de calidad del agua en sistema de agua potable rural”. In: *Revista de Ingeniería Electrónica, Automática y Comunicaciones* 42.3 (2021), pp. 60–70.
- [4] Almeida et al. “Aplicaciones de la inteligencia artificial para el cálculo de los indicadores de calidad del agua”. In: *Jornadas de Investigación y Actualización en Ingeniería Civil*. Vol. 1. 1. 2015, pp. 41–44.
- [5] Lenin Quiñones Huatangari et al. “Red neuronal artificial para estimar un índice de calidad de agua”. In: *Enfoque UTE* 11.2 (2020), pp. 109–120.
- [6] R. Ismail et al. “Artificial intelligence for application in water engineering: the use of ANN to determine water quality index in rivers”. In: *Civil Engineering Journal* 10.7 (2024), pp. 2261–2274.
- [7] Hauwa Mohammed Mustafa et al. “Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review”. In: *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. 2021, pp. 968–975.
- [8] Mosleh Hmoud Al-Adhaileh and Fawaz Waselallah Alsaade. “Modelling and Prediction of Water Quality by Using Artificial Intelligence”. In: *Sustainability* 13.8 (2021), p. 4259.
- [9] Rishi Rana et al. “Artificial Intelligence for Surface Water Quality Evaluation, Monitoring and Assessment”. In: *Water* 15.22 (2023), p. 3919.
- [10] Joshua O. Ighalo, Adewale George Adeniyi, and Gonçalo Marques. “Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis”. In: *Modeling Earth Systems and Environment* (2020).
- [11] Mohammad Najafzadeh and Sajad Basirian. “Evaluation of River Water Quality Index Using Remote Sensing and Artificial Intelligence Models”. In: *Remote Sensing* 15.9 (2023), p. 2359.
- [12] Puze Wang et al. “Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants”. In: *Science of the Total Environment* 693 (2019), p. 133440.