

Monitorización Inteligente de Calidad Hídrica y Red de Alertas en Acueductos Rurales con Aprendizaje Automático

CONVOCATORIA COLOMBIA INTELIGENTE: CIENCIA Y TECNOLOGÍAS
CUÁNTICAS E INTELIGENCIA ARTIFICIAL PARA LOS TERRITORIOS |
Convocatoria 966

Tabla de Contenido

1. Generalidades del Proyecto	3
2. Resumen Ejecutivo	3
3. Planteamiento del Problema y Justificación	4
4. Marco Teórico y Estado del Arte	5
4.1. <i>Introducción al Dominio</i>	5
4.2. <i>Revisión de la Literatura (Literature Review)</i>	5
4.3. <i>Tecnologías y Enfoques Actuales (State of the Art)</i>	6
4.4. <i>Brechas de Conocimiento y Oportunidades (Knowledge Gaps & Opportunities)</i>	7
5. Objetivos	8
6. Metodología Propuesta	10
7. Plan de Ejecución y Gestión	11
7.1. <i>Cronograma de Actividades</i>	11
7.2. <i>Matriz de Riesgos</i>	13
8. Resultados e Impactos Esperados	14
8.1. <i>Resultados Esperados (Entregables)</i>	14
8.2. <i>Impactos Esperados</i>	15
9. Referencias Bibliográficas	15
9. <i>Referencias Bibliográficas (APA 7th Edition Format)</i>	15

1. Generalidades del Proyecto

Título: Monitorización Inteligente de Calidad Hídrica y Red de Alertas en Acueductos Rurales con Aprendizaje Automático

Convocatoria: CONVOCATORIA COLOMBIA INTELIGENTE: CIENCIA Y TECNOLOGÍAS CUÁNTICAS E INTELIGENCIA ARTIFICIAL PARA LOS TERRITORIOS | Convocatoria 966

Programa: Por definir

Entidad/Persona: COTECMAR

Línea Temática: Colombia Inteligente, Ciencia, Tecnologías Cuánticas, Inteligencia Artificial, Territorios, CTel, Innovación

Duración: 24 meses

Área OCDE: Por definir

● **Descripción:** Numerosos acueductos rurales en territorios colombianos carecen de monitoreo de calidad del agua en tiempo real, lo que genera riesgos para la salud pública. Este proyecto busca desarrollar y desplegar una red de sensores IoT de bajo costo para recopilar parámetros físico-químicos del agua. Los datos serán procesados mediante modelos de aprendizaje automático para detectar anomalías y predecir eventos de contaminación, brindando a las comunidades un sistema de alerta temprana.

● **Palabras Clave:** Inteligencia Artificial, IoT, Calidad del Agua, Acueductos Rurales, Machine Learning, Salud Pública

2. Resumen Ejecutivo

Las comunidades rurales de Colombia enfrentan un desafío crítico: la falta de monitoreo en tiempo real de la calidad del agua, lo que las expone a riesgos sanitarios y una gestión ineficiente de recursos vitales. Los métodos tradicionales son reactivos y no detectan la contaminación oportunamente. Nuestro proyecto aborda esta brecha con una solución innovadora: un sistema inteligente de monitoreo hídrico y red de alertas tempranas, basado en Inteligencia Artificial, que busca transformar la gestión del agua y cerrar brechas tecnológicas en el país.

Para lograr esta visión, desplegaremos una red modular de sensores IoT de bajo costo en acueductos rurales piloto. Paralelamente, desarrollaremos y validaremos modelos de Aprendizaje Automático robustos para la detección de anomalías y la predicción de contaminación con alta precisión. Estos componentes se integrarán en un sistema de alerta temprana que promoverá la apropiación comunitaria. Nuestra metodología híbrida Agile-Lean Startup asegura un desarrollo iterativo, centrado en el usuario y adaptable a las particularidades del entorno rural.

Los resultados esperados incluyen un prototipo funcional (TRL 7) del sistema integral, modelos ML validados y publicaciones científicas. El impacto técnico-científico avanzará el estado del arte en IA para gestión hídrica. A nivel económico, mejoraremos la eficiencia del agua y generaremos potencial de emprendimiento regional. Social y ambientalmente, protegeremos la salud pública, empoderaremos a las comunidades con herramientas tecnológicas y conocimientos, y facilitaremos una gestión hídrica más sostenible.

Este proyecto es crucial para fortalecer la investigación aplicada, cerrar brechas tecnológicas y construir un futuro más inteligente y equitativo para las comunidades rurales de Colombia, marcando un hito en la gestión proactiva de un recurso esencial.

3. Planteamiento del Problema y Justificación

La calidad del agua potable es un pilar fundamental para la salud pública y el desarrollo sostenible, especialmente en las comunidades rurales de Colombia, donde el acceso a sistemas de monitoreo avanzados es históricamente limitado. A nivel global, la supervisión tradicional de la calidad hídrica se ha basado en muestreos manuales y análisis de laboratorio periódicos, métodos que, si bien ofrecen precisión, son inherentemente reactivos. Esta aproximación carece de la capacidad para proporcionar datos en tiempo real y detectar eventos de contaminación de manera oportuna, lo que resulta en una mitigación insuficiente de riesgos sanitarios inmediatos y una gestión ineficiente de un recurso vital. En el contexto colombiano, esta problemática se agudiza en los acueductos rurales, donde la ausencia de vigilancia continua expone a poblaciones vulnerables a riesgos directos para su salud.

Como la revisión de la literatura indica, si bien el advenimiento de la Internet de las Cosas (IoT) y el Aprendizaje Automático (ML) ha revolucionado el monitoreo ambiental, con estudios como los de Sharanya et al. (2024) y Kalaivanan y Vellingiri (2024) demostrando la efectividad de la integración de IoT y ML para la detección de anomalías y la predicción en entornos urbanos y de laboratorio, persisten brechas críticas en el contexto rural. Existe una **brecha fundamental** en la falta de soluciones de monitoreo en tiempo real de bajo costo y adaptadas específicamente para acueductos rurales, ya que la mayoría de las implementaciones se centran en infraestructuras urbanas con recursos diferentes. Asimismo, se identifica una **brecha crucial** en la adaptación y validación de modelos de Aprendizaje Automático para la detección de anomalías y la predicción de contaminación en datos de calidad de agua de entornos rurales, dado que los patrones de contaminación y las características del agua en estas zonas pueden variar drásticamente y los modelos existentes no generalizan adecuadamente. Finalmente, hay una **brecha significativa** en la integración efectiva de estos sistemas tecnológicos con redes de alerta temprana que sean operativas y accesibles para las comunidades rurales, asegurando que la tecnología se traduzca en acción comunitaria efectiva y empoderamiento local.

Este proyecto surge como la respuesta directa e innovadora a estas brechas identificadas, proponiendo una solución integral de vanguardia. Mediante el diseño y despliegue de una red modular de sensores IoT de bajo costo, energéticamente eficiente y de fácil mantenimiento, se abordará la necesidad de recopilación de datos en tiempo real adaptada al entorno rural. Complementariamente, se investigarán, desarrollarán y validarán modelos de Aprendizaje Automático optimizados para la variabilidad y escasez de datos inherente a estos contextos, capaces de detectar anomalías y predecir eventos de contaminación con alta precisión. La integración de estas tecnologías en un sistema de alerta temprana operativo y accesible para las

comunidades rurales transformará un enfoque tradicionalmente reactivo en uno proactivo y predictivo, elevando la capacidad de respuesta y la protección de la salud pública.

La pertinencia de este proyecto es innegable en el contexto actual colombiano, alineándose estratégicamente con la misión de Minciencias de cerrar brechas tecnológicas y generar impacto territorial medible. Representa una solución disruptiva que no solo fortalecerá la investigación aplicada en Inteligencia Artificial para la gestión hídrica, sino que también mejorará directamente la calidad de vida de miles de habitantes rurales, fortaleciendo su soberanía hídrica. Este esfuerzo es oportuno y crítico para equipar a las comunidades rurales con herramientas tecnológicas de punta, fomentar el empoderamiento comunitario y contribuir a un desarrollo ambiental, social y económico sostenible en las regiones, marcando un paso fundamental hacia un Colombia más inteligente y equitativo.

4. Marco Teórico y Estado del Arte

4.1. Introducción al Dominio

La calidad del agua es un pilar fundamental para la salud pública y el desarrollo sostenible, especialmente en comunidades rurales donde el acceso a sistemas de monitoreo avanzados es limitado. Tradicionalmente, la supervisión de la calidad hídrica se ha basado en muestreos manuales y análisis de laboratorio periódicos, métodos que, si bien son precisos, adolecen de la falta de capacidad para proporcionar datos en tiempo real y detectar supervisión de la calidad hídrica se ha basado en muestreos manuales y análisis de laboratorio periódicos, métodos que, si bien son precisos, adolecen de la falta de capacidad para proporcionar datos en tiempo real y detectar eventos de contaminación de manera oportuna. Este enfoque reactivo es insuficiente para mitigar riesgos sanitarios inmediatos y gestionar eficientemente los recursos hídricos.

El advenimiento de la Internet de las Cosas (IoT) ha revolucionado la capacidad de recopilar datos ambientales de manera continua y a bajo costo, permitiendo la implementación de redes de sensores que transmiten información en tiempo real. Al integrar estos sistemas de monitoreo con técnicas de Aprendizaje Automático (Machine Learning - ML), es posible trascender la mera recolección de datos para desarrollar capacidades predictivas y de detección de anomalías. Los modelos de ML pueden analizar patrones complejos en los datos de los sensores, identificar desviaciones que sugieran contaminación y predecir eventos futuros, transformando el monitoreo de la calidad del agua en un sistema proactivo y predictivo. La aplicación de estas tecnologías en acueductos rurales promete un cambio paradigmático hacia una gestión hídrica más segura, eficiente y equitativa, al permitir sistemas de alerta temprana que protejan a las poblaciones vulnerables.

4.2. Revisión de la Literatura (Literature Review)

La investigación reciente ha demostrado un creciente interés en la integración de IoT y Aprendizaje Automático para la monitorización de la calidad del agua, abordando las limitaciones de los métodos tradicionales.

Sharanya et al. (2024) propusieron un Sistema Inteligente de Monitoreo de Calidad del Agua y Detección de Fugas (IWQLDS) que integra IoT y ML para mejorar la gestión del agua en sistemas urbanos. Su trabajo destaca cómo los métodos convencionales carecen de datos en tiempo real y de capacidades predictivas, lo que conduce a respuestas tardías ante la contaminación. El IWQLDS despliega una red de sensores IoT para recopilar continuamente datos de parámetros críticos del agua, procesándolos con algoritmos avanzados de ML para detectar anomalías, predecir problemas futuros e iniciar acciones correctivas. Este enfoque representa un avance significativo al ofrecer monitoreo y análisis en tiempo real, mejorando la detección de fugas y problemas de calidad del agua, y facilitando intervenciones proactivas.

En el ámbito de la detección de anomalías, **Kalaivanan y Vellingiri (2024)** desarrollaron un enfoque híbrido de aprendizaje automático para mejorar la detección de anomalías en la calidad del agua potable. Su técnica, basada en un autoencoder y una máquina de aprendizaje extremo (AEELM), utiliza el autoencoder para la selección de características y el algoritmo ELM para la clasificación. Evaluado con un conjunto de datos del río Cauvery, este enfoque superó a otros modelos con una precisión del 83%, demostrando su efectividad en la identificación de contaminaciones. Este estudio subraya la importancia de modelos ML robustos para manejar la complejidad inherente a los datos de calidad del agua y la creciente contaminación.

Mehr (2020) exploró la predicción y detección de anomalías en la calidad del agua mediante un aprendizaje jerárquico explicable a través del intercambio de parámetros. Aunque el resumen detallado no está disponible, el título sugiere un enfoque en la interpretabilidad de los modelos de ML, un aspecto crucial para la confianza y la adopción de estos sistemas en aplicaciones críticas como la monitorización del agua. La capacidad de entender por qué un modelo predice una anomalía es tan importante como la predicción misma, especialmente en contextos de salud pública.

Desde una perspectiva de salud pública, un estudio sobre la implementación de métodos de aprendizaje automático para el análisis y clasificación de la calidad del agua superficial en Java Central (PMID: 41182597, 2024) subraya el papel crítico del monitoreo de la calidad del agua en la protección ambiental y la salud pública. Propone y evalúa un marco de clasificación predictiva utilizando SVM y XGBoost para evaluar los niveles de contaminación del agua, destacando cómo el ML puede ofrecer una visión profunda y una clasificación precisa de la calidad del agua.

Finalmente, la investigación sobre la alerta temprana de proliferaciones de cianobacterias nocivas basada en monitoreo *in situ* de alta frecuencia y modelado de aprendizaje automático (PMID: 40925059, 2024) demuestra la viabilidad de sistemas de alerta temprana. Este estudio, centrado en el lago Müggelsee (Alemania), utiliza datos *in situ* de alta frecuencia para predecir proliferaciones de cianobacterias, que representan una seria amenaza para la salud humana. La capacidad de generar alertas tempranas mediante ML permite la implementación de medidas de control preventivas, un componente esencial para proteger la salud pública y la biodiversidad.

4.3. Tecnologías y Enfoques Actuales (State of the Art)

El estado del arte en la monitorización inteligente de la calidad hídrica se caracteriza por la convergencia de la Internet de las Cosas (IoT) y el Aprendizaje Automático (ML), impulsando sistemas más autónomos, predictivos y en tiempo real. Las tecnologías y enfoques dominantes incluyen:

1. Redes de Sensores IoT de Bajo Costo: La base de estos sistemas reside en el despliegue de redes de sensores capaces de medir parámetros físico-químicos clave del agua (pH, turbidez, temperatura, conductividad, oxígeno disuelto, etc.). La tendencia actual se enfoca en sensores de bajo costo que pueden ser implementados a gran escala, especialmente en áreas remotas o rurales, donde las limitaciones presupuestarias son significativas. Estos sensores están conectados a módulos de comunicación (Wi-Fi, LoRaWAN, NB-IoT, 5G) para transmitir datos a plataformas en la nube. Aunque la investigación específica en sensores de bajo costo para calidad del agua en zonas rurales es un área activa, el concepto general de sensores IoT de bajo costo para monitoreo ambiental es un campo maduro (Ferrer-Cid et al., 2022).

2. Plataformas de Recopilación y Gestión de Datos en la Nube: Los datos generados por las redes de sensores IoT se recopilan, almacenan y gestionan en plataformas basadas en la nube. Estas plataformas no solo garantizan la escalabilidad y accesibilidad de los datos, sino que también proporcionan las herramientas necesarias para la preprocesamiento, integración y análisis de grandes volúmenes de información.

3. Algoritmos de Aprendizaje Automático para Detección de Anomalías y Predicción:

- **Detección de Anomalías:** Se utilizan algoritmos como Autoencoders, Extreme Learning Machines (ELM), Support Vector Machines (SVM), Random Forest (RF) y Extreme Gradient Boosting (XGBoost) para identificar patrones inusuales en los datos que puedan indicar eventos de contaminación (Kalaivanan & Vellingiri, 2024; Fadhilah et al., 2025). La investigación también aborda el desafío de datos desequilibrados y faltantes, comunes en entornos reales (Fadhilah et al., 2025).

- **Predicción y Pronóstico:** Modelos de series temporales basados en ML (como redes neuronales recurrentes, LSTM o GRU) y Deep Learning se emplean para predecir la calidad futura del agua y la probabilidad de eventos de contaminación. Estos modelos pueden integrar datos históricos, meteorológicos e hidrológicos para mejorar la precisión (PMID: 40639737, 2024).

- **Aprendizaje Jerárquico y Explicable:** Existe un interés creciente en modelos de ML que no solo predicen, sino que también proporcionan explicaciones sobre sus predicciones, lo que es crucial para la toma de decisiones informada en la gestión de riesgos (Mehr, 2020).

4. Sistemas de Alerta Temprana: La integración de la monitorización IoT con modelos predictivos de ML ha permitido el desarrollo de sistemas de alerta temprana. Estos sistemas están diseñados para activar notificaciones automáticas a las partes interesadas (operadores, autoridades, comunidades) cuando se detectan anomalías o se pronostican eventos de contaminación, facilitando una respuesta rápida y mitigando impactos en la salud pública (PMID: 40925059, 2024; Springer, 2023).

5. Desafíos en la Implementación: A pesar de los avances, la implementación de estos sistemas enfrenta desafíos como la calidad de los datos IoT (Zubair et al., 2019), la necesidad de modelos de ML robustos para datos ruidosos o incompletos, la gestión de la energía en dispositivos remotos (Kantareddy et al., 2019) y la ciberseguridad en entornos industriales y rurales (Frey et al., 2019).

4.4. Brechas de Conocimiento y Oportunidades (Knowledge Gaps & Opportunities)

A pesar de los avances significativos en la monitorización de la calidad del agua mediante IoT y Aprendizaje Automático, existen brechas de conocimiento y oportunidades claras que justifican la innovación propuesta

por este proyecto, especialmente en el contexto de acueductos rurales en territorios como Colombia.

Una brecha fundamental es la **falta de soluciones de monitoreo en tiempo real de bajo costo y adaptadas específicamente para acueductos rurales**. La mayoría de la investigación y las implementaciones existentes se centran en entornos urbanos o industriales con infraestructuras y recursos significativamente diferentes. Los acueductos rurales a menudo carecen de acceso a energía constante, conectividad de red robusta y personal técnico especializado, lo que hace que las soluciones "plug-and-play" o de alto costo sean inviables. Esto genera una oportunidad para desarrollar una red de sensores IoT que no solo sea de bajo costo, sino también energéticamente eficiente y de fácil mantenimiento, diseñada para las condiciones operativas y ambientales particulares de estas áreas.

Otra brecha crucial radica en la **adaptación y validación de modelos de Aprendizaje Automático para la detección de anomalías y predicción de contaminación en datos de calidad de agua de entornos rurales**. Los patrones de contaminación, las características del agua y los factores ambientales pueden variar drásticamente entre regiones urbanas y rurales. Los modelos existentes, aunque efectivos en sus contextos de entrenamiento, pueden no generalizar bien a la diversidad de contaminantes y condiciones hidrológicas presentes en acueductos rurales. Esto crea una oportunidad para investigar y desarrollar modelos de ML que sean robustos ante la variabilidad de los datos rurales, capaces de manejar la escasez o intermitencia de datos y que puedan identificar de manera precisa y temprana eventos de contaminación relevantes para la salud pública local.

Finalmente, existe una brecha en la **integración efectiva de estos sistemas tecnológicos con redes de alerta temprana que sean operativas y accesibles para las comunidades rurales**. No basta con detectar y predecir la contaminación; la información debe ser comunicada de manera clara, oportuna y a través de canales adecuados para que las comunidades puedan tomar acciones preventivas. La oportunidad radica en diseñar un sistema de alerta que considere las capacidades de comunicación locales, la alfabetización tecnológica de los usuarios y que empodere a las comunidades para responder eficazmente a las amenazas a la calidad del agua, cerrando el ciclo entre la detección tecnológica y la acción comunitaria. Este proyecto busca abordar estas brechas mediante el desarrollo y despliegue de una solución integral que no solo monitoree y prediga, sino que también active una red de alertas diseñada para impactar directamente la salud pública en acueductos rurales.

5. Objetivos

Objetivo General

Fortalecer la Investigación Aplicada y el Desarrollo Tecnológico en Inteligencia Artificial para la gestión hídrica en acueductos rurales, mediante el diseño e implementación de un sistema inteligente de monitoreo de calidad del agua y red de alertas tempranas, generando soluciones disruptivas con impacto ambiental, social y económico medible que contribuyan al desarrollo regional y al cierre de brechas tecnológicas en Colombia.

Objetivos Específicos

- 1. Objetivo:** Diseñar y desplegar una red de monitoreo IoT de bajo costo en acueductos rurales.

- **Específico (S):** Diseñar, desarrollar y desplegar una red modular y de bajo costo de sensores IoT para el monitoreo continuo y en tiempo real de al menos cinco (5) parámetros físico-químicos esenciales de la calidad del agua (ej. pH, turbidez, conductividad, temperatura, oxígeno disuelto) en tres (3) acueductos rurales pilotos seleccionados en territorios del país.
- **Medible (M):** La red de sensores debe operar con una fiabilidad de transmisión de datos del 95% y los costos de implementación del hardware por nodo no deben superar el \$X USD, en los 3 acueductos rurales pilotos.
- **Alcanzable (A):** El diseño modular y la selección de tecnologías de bajo costo, junto con estrategias de eficiencia energética y comunicación resiliente, hacen que el despliegue de la red sea viable en el contexto rural. El proyecto cuenta con un equipo multidisciplinario con experiencia en IoT y desarrollo de hardware.
- **Relevante (R):** Este objetivo es fundamental para abordar la brecha de la falta de monitoreo en tiempo real y sentar las bases para la recopilación de datos necesaria para los modelos de Aprendizaje Automático, contribuyendo directamente al objetivo general de establecer un sistema inteligente de monitoreo.
- **Plazo (T):** Completado para el **mes 8** de ejecución del proyecto.

2. **Objetivo:** Desarrollar y validar modelos de Aprendizaje Automático para la detección de anomalías y predicción de contaminación.

- **Específico (S):** Investigar, desarrollar y validar robustos modelos de Aprendizaje Automático capaces de detectar anomalías en la calidad del agua con alta precisión y predecir eventos de contaminación con un horizonte mínimo de 24 horas, optimizados para la variabilidad y escasez de datos inherente a entornos de acueductos rurales.
- **Medible (M):** Los modelos de ML deben alcanzar una precisión (accuracy) en la detección de anomalías superior al 85% y un F1-score > 0.75 en la predicción de eventos de contaminación.
- **Alcanzable (A):** La investigación en arquitecturas de ML adaptadas a datos limitados (como modelos semi-supervisados o de transferencia de aprendizaje) y la validación con expertos en calidad del agua, hacen que estos objetivos de rendimiento sean alcanzables dentro de un proyecto de I+D.
- **Relevante (R):** Este objetivo es crucial para transformar el enfoque reactivo actual en uno proactivo y predictivo, abordando la brecha en la adaptación de modelos de ML para entornos rurales y siendo el corazón inteligente del sistema propuesto en el objetivo general.
- **Plazo (T):** Completado y validado en entorno simulado para el **mes 15** de ejecución del proyecto.

3. **Objetivo:** Implementar un sistema de alerta temprana integrado y promover la apropiación comunitaria.

- **Específico (S):** Implementar y validar un sistema de alerta temprana de contaminación hídrica, integrado con los modelos ML, y establecer mecanismos de comunicación y capacitación que aseguren la apropiación tecnológica y la capacidad de respuesta oportuna de las comunidades locales beneficiarias y los operadores de los acueductos rurales.

- **Medible (M):** El sistema debe lograr una tasa de acción (respuestas efectivas y verificables) ante el 70% de las alertas generadas por los modelos, y capacitar a un mínimo de 30 actores locales (operadores, líderes comunitarios, miembros del semillero) en el uso y mantenimiento del sistema.
- **Alcanzable (A):** La metodología de co-diseño centrada en el usuario, la capacitación continua y la alianza con organizaciones locales y regionales facilitarán la implementación y la apropiación por parte de las comunidades, superando barreras de adopción.
- **Relevante (R):** Este objetivo cierra la brecha de la integración efectiva de la tecnología con la acción comunitaria, asegurando que el sistema inteligente genere un impacto social directo y empodere a las comunidades, alineándose con el impacto social del objetivo general.
- **Plazo (T):** Implementación para el **mes 20** y validación final para el **mes 22** del proyecto.

4. **Objetivo:** Generar conocimiento científico y fortalecer capacidades de investigación y formación.

- **Específico (S):** Generar conocimiento científico y tecnológico aplicable, fortalecer las capacidades de investigación de la IES y de sus aliados, e impulsar la formación de capital humano especializado, incluyendo la vinculación activa de semilleros de investigación, para el desarrollo sostenible de tecnologías de IA en el país.
- **Medible (M):** Publicar al menos dos (2) artículos científicos en revistas indexadas (Scopus/WOS) y generar un (1) prototipo funcional (TRL 5-7), validado en el entorno rural. El semillero vinculado deberá producir al menos una (1) ponencia en evento nacional/internacional o un (1) informe técnico de I+D.
- **Alcanzable (A):** Como un proyecto de I+D liderado por una IES con alianzas estratégicas, la generación de publicaciones científicas, el desarrollo de un prototipo y la formación de talento son resultados esperados y viables con la asignación de recursos y el cronograma propuestos.
- **Relevante (R):** Este objetivo contribuye directamente al fortalecimiento de la investigación aplicada y el desarrollo tecnológico en IA, al cierre de brechas tecnológicas y a la formación de capital humano, pilares del objetivo general y de los impactos esperados del proyecto.
- **Plazo (T):** Las publicaciones e informes deben ser sometidas y el prototipo finalizado para el **mes 24** de ejecución del proyecto, con vinculación activa del semillero a lo largo de toda la ejecución.

6. Metodología Propuesta

Framework Seleccionado: Marco Metodológico Híbrido Agile-Lean Startup

El proyecto adoptará un Marco Metodológico Híbrido Agile-Lean Startup, ideal para abordar la complejidad de un proyecto de I+D+i que integra hardware (IoT), software (Aprendizaje Automático) y un fuerte componente social. Esta metodología es la más adecuada porque permite la gestión flexible de requisitos y la adaptación continua a los desafíos técnicos y las particularidades de los entornos rurales (Objetivo 1 y 2). Su enfoque iterativo y centrado en el usuario es crucial para co-diseñar y validar la solución con las comunidades, asegurando la pertinencia, usabilidad y apropiación del sistema de alertas (Objetivo 3), lo que es fundamental para lograr el impacto social y tecnológico disruptivo del objetivo general del proyecto.

Fases Principales de la Metodología:

- **Fase 1: Planificación y Definición de Requisitos** - Establecimiento del equipo, diagnóstico detallado de las condiciones de los acueductos rurales piloto y la comunidad, y definición de las especificaciones funcionales y no funcionales del sistema, sentando las bases para las iteraciones posteriores.
- **Fase 2: Diseño y Prototipado Iterativo de Hardware IoT** - Desarrollo en ciclos cortos de los nodos de sensores de bajo costo, incluyendo diseño, prototipado, pruebas de laboratorio y calibración inicial, con un enfoque en la eficiencia energética y la robustez para el entorno rural.
- **Fase 3: Desarrollo y Validación Iterativa de Modelos de Aprendizaje Automático** - Recopilación y preprocesamiento de datos, investigación y desarrollo de arquitecturas de ML, entrenamiento, simulación y validación continua de los modelos para detección de anomalías y predicción de contaminación.
- **Fase 4: Integración del Sistema y Despliegue Piloto en Campo** - Ensamblaje progresivo del hardware IoT, la plataforma de datos en la nube y los modelos ML, seguido de la instalación y pruebas de campo del sistema integral en los acueductos rurales piloto, validando la funcionalidad en un entorno relevante.
- **Fase 5: Validación Final, Apropiación Social y Cierre** - Monitoreo continuo del sistema en campo, ajustes finos y re-entrenamiento de modelos, implementación de talleres de capacitación y mecanismos de transferencia de conocimiento para asegurar la apropiación comunitaria y la sostenibilidad del sistema, culminando con la divulgación de resultados y publicaciones.

7. Plan de Ejecución y Gestión

Cronograma de Actividades

7.1. Cronograma de Actividades

Fase	Actividad / Hito Clave	Entregable Principal	Duración Estimada (Semanas)
Fase 1: Planeación y Levantamiento de Requisitos	<i>Establecimiento del equipo, diagnóstico detallado de las condiciones de los acueductos rurales piloto y la comunidad, y definición de las especificaciones funcionales y no funcionales del sistema.</i>		8
	1.1. Constitución del equipo de trabajo y formalización de alianzas estratégicas.	Acta de conformación de equipo y acuerdos de alianza	2
	1.2. Diagnóstico detallado de acueductos rurales piloto y levantamiento de requisitos.	Informe de diagnóstico y requisitos funcionales/no funcionales	4
	1.3. Definición de especificaciones técnicas del sistema IoT y ML.	Documento de especificaciones técnicas (Hardware, Software, ML)	3

	1.4. Planificación inicial y vinculación activa del semillero de investigación.	Plan de trabajo del semillero y cronograma de actividades	2
Fase 2: Diseño y Prototipado Iterativo de Hardware IoT	<i>Desarrollo en ciclos cortos de los nodos de sensores de bajo costo, incluyendo diseño, prototipado, pruebas de laboratorio y calibración inicial, con un enfoque en la eficiencia energética y la robustez para el entorno rural.</i>		24
	2.1. Diseño detallado y selección de componentes para nodos IoT modulares.	Diseño electrónico y mecánico (CAD), lista de materiales (BOM)	6
	2.2. Prototipado y ensamblaje de los nodos de sensores de bajo costo.	Prototipos de nodos IoT ensamblados	8
	2.3. Desarrollo e integración de módulos de comunicación y gestión energética.	Módulos de comunicación y energía integrados y probados	6
	2.4. Pruebas de laboratorio, calibración y optimización inicial de prototipos IoT.	Informe de pruebas de laboratorio y calibración (Fiabilidad > 95%)	8
Fase 3: Desarrollo y Validación Iterativa de Modelos de Aprendizaje Automático	<i>Recopilación y preprocesamiento de datos, investigación y desarrollo de arquitecturas de ML, entrenamiento, simulación y validación continua de los modelos para detección de anomalías y predicción de contaminación.</i>		40
	3.1. Recolección de datos (históricos y de campo) y preprocesamiento.	Base de datos limpia y estructurada para ML	10
	3.2. Investigación y selección de arquitecturas de ML para detección de anomalías y predicción.	Informe de revisión de arquitecturas ML y propuesta de modelos	10
	3.3. Desarrollo y entrenamiento inicial de modelos ML.	Modelos ML entrenados (Versión 1.0)	12
	3.4. Simulación y validación de modelos ML en entorno controlado.	Informe de validación de modelos (Accuracy > 85%, F1-score > 0.75)	8
Fase 4: Integración del Sistema y Despliegue Piloto en Campo	<i>Ensamblaje progresivo del hardware IoT, la plataforma de datos en la nube y los modelos ML, seguido de la instalación y pruebas de campo del sistema integral en los acueductos rurales piloto, validando la funcionalidad en un entorno relevante.</i>		36
	4.1. Integración de hardware IoT con plataforma de datos en la nube.	Plataforma en la nube recibiendo datos IoT en tiempo real	8
	4.2. Desarrollo de la interfaz de usuario y sistema de alerta temprana.	Interfaz de usuario (dashboard) y módulo de alertas funcional	8
	4.3. Despliegue e instalación de la red de sensores IoT en acueductos piloto.	Nodos IoT instalados y operativos en 3 acueductos rurales	10
	4.4. Pruebas de campo y validación inicial del sistema integral (TRL 5-6).	Informe de pruebas de campo y evaluación de TRL (5-6)	10

Fase 5: Validación Final, Apropiación Social y Cierre	<i>Monitoreo continuo del sistema en campo, ajustes finos y re-entrenamiento de modelos, implementación de talleres de capacitación y mecanismos de transferencia de conocimiento para asegurar la apropiación comunitaria y la sostenibilidad del sistema, culminando con la divulgación de resultados y publicaciones.</i>		28
	5.1. Monitoreo continuo del sistema en campo y recolección de datos.	Base de datos de monitoreo extendida y validada	12
	5.2. Re-entrenamiento y validación final de modelos ML con datos de campo reales.	Modelos ML optimizados y validados (TRL 7)	8
	5.3. Talleres de capacitación y apropiación social con comunidades y operadores.	Informes de capacitación (mínimo 30 actores capacitados)	10
	5.4. Elaboración de publicaciones científicas, informes técnicos y prototipo final.	2 Artículos científicos sometidos, 1 prototipo funcional (TRL 7), Manual de Uso	12

Matriz de Riesgos

7.2. Matriz de Riesgos

#	Riesgo Potencial	Probabilidad	Impacto	Estrategia de Mitigación
1	Precisión de Modelos ML Insuficiente <i>Relacionado con: Fase 3 (Desarrollo y Validación Iterativa de Modelos de Aprendizaje Automático), Fase 5 (Re-entrenamiento y validación final).</i>	Media	Alto	Implementar ciclos de desarrollo de ML ágiles con reentrenamiento continuo y validación por expertos. Explorar arquitecturas de ML robustas (ej., semi-supervisadas, transferencia de aprendizaje) para manejar la complejidad de los patrones de contaminación y la heterogeneidad de datos rurales.
2	Dificultades en Despliegue y Mantenimiento de Nodos IoT <i>Relacionado con: Fase 2 (Diseño y Prototipado Iterativo de Hardware IoT), Fase 4 (Despliegue Piloto en Campo).</i>	Media	Alto	Diseñar nodos energéticamente eficientes (con paneles solares/baterías autónomas) y con comunicación resiliente (ej., LoRaWAN, NB-IoT). Establecer un programa robusto de capacitación a la comunidad y a la Organización Local-Regional para mantenimiento básico y vigilancia, asegurando la continuidad operativa y protección física.
3	Baja Adopción o Apropiación Comunitaria del Sistema <i>Relacionado con: Fase 1 (Planeación y Levantamiento de Requisitos), Fase 4 (Integración del Sistema y Piloto en Campo), Fase 5 (Apropiación Social y Cierre).</i>	Media	Alto	Involucrar activamente a las comunidades y operadores desde el co-diseño de la interfaz de alertas. Realizar talleres de capacitación prácticos, transparentes y en lenguaje local, enfatizando los beneficios y el funcionamiento ético de la IA. Crear canales de retroalimentación continuos para ajustar el sistema a sus necesidades.

4	Calidad y Disponibilidad Insuficiente de Datos para ML <i>Relacionado con: Fase 3 (Desarrollo y Validación Iterativa de Modelos de Aprendizaje Automático - Actividad 3.1).</i>	Media	Alto	Implementar protocolos estrictos de recolección de datos de campo en los pilotos, complementando con datos históricos. Explorar técnicas de aumento de datos (data augmentation) y el uso de datos sintéticos. Priorizar el desarrollo de modelos de ML que sean robustos ante la escasez o variabilidad de datos, como los modelos semi-supervisados o de transferencia de aprendizaje.
5	Retrasos en la Adquisición de Componentes Críticos <i>Relacionado con: Fase 2 (Diseño y Prototipado Iterativo de Hardware IoT - Actividades 2.1, 2.2).</i>	Media	Medio	Identificar y calificar múltiples proveedores alternativos para componentes clave. Realizar pedidos con antelación y mantener un inventario estratégico de piezas críticas. Considerar la flexibilidad en el diseño para usar componentes de código abierto o fácilmente disponibles que minimicen la dependencia de un único proveedor o región.

8. Resultados e Impactos Esperados

8.1. Resultados Esperados (Entregables)

- **Red de Sensores IoT de Bajo Costo Desplegada:** Una red modular y energéticamente eficiente de sensores IoT, diseñada, desarrollada y desplegada en tres (3) acueductos rurales piloto, capaz de monitorear al menos cinco parámetros físico-químicos esenciales de la calidad del agua con una fiabilidad de transmisión del 95% y un costo de implementación por nodo no superior al \$X USD. (Corresponde al Objetivo Específico 1).
- **Modelos de Aprendizaje Automático Validados:** Un conjunto de modelos de Aprendizaje Automático robustos y optimizados para entornos rurales, validados para la detección de anomalías con una precisión superior al 85% y la predicción de eventos de contaminación con un F1-score > 0.75 y un horizonte mínimo de 24 horas. (Corresponde al Objetivo Específico 2).
- **Sistema Integrado de Alerta Temprana y Apropiación Comunitaria:** Un sistema de alerta temprana de contaminación hídrica completamente integrado con los modelos ML y la red IoT, validado en campo con una tasa de acción del 70% ante las alertas generadas, acompañado de manuales de uso y programas de capacitación para al menos 30 actores locales (operadores, líderes comunitarios, miembros del semillero). (Corresponde al Objetivo Específico 3).
- **Prototipo Funcional del Sistema de Monitoreo Inteligente (TRL 7):** Un prototipo funcional del sistema integral de monitoreo inteligente de calidad hídrica y red de alertas, validado en un entorno operativo rural, demostrando un Nivel de Madurez Tecnológica (TRL) de 7. (Corresponde al Objetivo Específico 4).

-
- **Producción Científica y Técnica:** Al menos dos (2) artículos científicos publicados en revistas indexadas (Scopus/WOS) y una (1) ponencia en evento nacional/internacional o informe técnico de I+D generado por el semillero de investigación vinculado. (Corresponde al Objetivo Específico 4).

8.2. Impactos Esperados

- **Impacto Técnico/Científico:**

El proyecto avanzará significativamente el estado del arte en la aplicación de la Inteligencia Artificial y el IoT para la gestión hídrica en contextos rurales. Se generará conocimiento científico aplicable a través de publicaciones en revistas indexadas y la creación de un prototipo funcional (TRL 7) de un sistema de monitoreo inteligente. Esto no solo validará nuevas metodologías para la detección de anomalías y predicción de contaminación en entornos con datos escasos y heterogéneos, sino que también contribuirá a cerrar la brecha tecnológica, equipando a comunidades rurales con herramientas de vanguardia para la gestión de su recurso hídrico vital.

- **Impacto Económico:**

Se espera una mejora sustancial en la eficiencia de la gestión del agua en los acueductos rurales, optimizando la asignación de recursos y reduciendo los costos asociados a la detección tardía de la contaminación. El prototipo desarrollado posee un alto potencial de escalabilidad a otros territorios, fomentando el emprendimiento local en el soporte y mantenimiento de estas tecnologías. Además, la alianza estratégica entre la IES, la Empresa Nacional y la Organización Local-Regional fortalecerá sus capacidades institucionales y generará valor a través de la co-creación y transferencia de conocimiento tecnológico, abriendo nuevas vías de desarrollo económico regional.

- **Impacto Social/Ambiental:**

Este proyecto tendrá un impacto social directo y medible en la salud pública de miles de habitantes rurales, al proporcionar un sistema de alerta temprana que permite acciones preventivas frente a la contaminación hídrica. Se empoderará a las comunidades rurales con herramientas y conocimientos para gestionar proactivamente su agua, y se capacitará a un mínimo de 30 actores locales, incluyendo semilleros de investigación, en tecnologías de IA e IoT, reduciendo la brecha de talento y promoviendo la apropiación social del conocimiento. Desde una perspectiva ambiental, el monitoreo proactivo facilitará una gestión más informada y sostenible de los recursos hídricos, protegiendo ecosistemas acuáticos y contribuyendo a la conservación de la biodiversidad, sentando bases para una bioeconomía regional más robusta.

9. Referencias Bibliográficas

9. Referencias Bibliográficas (APA 7th Edition Format)

- Fadhilah, R., Kuswanto, H., & Prastyo, D. (2025). Performance Evaluation of Classification Methods Utilizing Resampling Techniques for Water Quality Prediction on Imbalanced Data. *International Journal of Intelligent Systems and Applications in Engineering*, 13(1), 1-10. [Note: This paper is listed as 2025,

implying it's forthcoming or very recent. Citation based on abstract information.]

- Ferrer-Cid, P., Garcia-Calvete, J., Main-Nadal, A., Ye, Z., Barcelo-Ordinas, J. M., & Garcia-Vidal, J. (2022). Sampling Trade-Offs in Duty-Cycled Systems for Air Quality Low-Cost Sensors. *arXiv preprint arXiv:2206.01211*.
- Frey, M., Gündoğan, C., Kietzmann, P., Lenders, M., Petersen, H., Schmidt, T. C., Shzu-Juraschek, F., & Wählisch, M. (2019). Security for the Industrial IoT: The Case for Information-Centric Networking. *arXiv preprint arXiv:1903.02058*.
- Kalaivanan, K., & Vellingiri, J. (2024). A hybrid machine learning approach for enhanced anomaly detection in drinking water quality. *Journal of Water and Land Development*, 60(I–III), 116–123.
<https://doi.org/10.24425/jwld.2024.148560>
- Kantareddy, S. N. R., Mathews, I., Sun, S., Layurova, M., Thapa, J., Correa-Baena, J. P., Bhattacharyya, R., Buonassis, T., Sarma, S. E., & Peters, I. M. (2019). Perovskite PV-powered RFID: enabling low-cost self-powered IoT sensors. *arXiv preprint arXiv:1909.08862*.
- Mehr, A. M. (2020). Prediction and anomaly detection in water quality with explainable hierarchical learning through parameter sharing. *Dissertation, University of California, Berkeley*.
- Sharanya, U. G., Birabbi, K. M., Sahana, B. H., Kumar, D. M., Sharmila, N., & Mallikarjunaswamy, S. (2024). Design and Implementation of IoT-based Water Quality and Leakage Monitoring System for Urban Water Systems Using Machine Learning Algorithms. *Journal of Data Acquisition and Processing*, 39(1), 12-25.
- Springer. (2023). The Role of Machine Learning and IoT in Early Detection of Water Contamination. In *Recent Advances in Smart Systems and Technologies* (pp. 53-65). Springer.
https://link.springer.com/chapter/10.1007/978-981-96-3886-4_6
- Zubair, N., A, N., Hebbar, K., & Simmhan, Y. (2019). Characterizing IoT Data and its Quality for Use. *arXiv preprint arXiv:1906.09673*.
- [PMID: 40639737]. (2024). Prediction of water quality parameters and pollution exceedance analysis in typical rivers of semi-arid regions based on interpretable deep learning models. *Environmental Research*, 252(Pt 1), 118712. <https://doi.org/10.1016/j.envres.2024.118712>
- [PMID: 40925059]. (2024). Early warning of harmful cyanobacteria blooms based on high frequency in situ monitoring and intelligible machine learning modelling: The case study of Lake Müggelsee (Germany). *Water Research*, 260, 122046. <https://doi.org/10.1016/j.watres.2024.122046>
- [PMID: 41182597]. (2024). Implementing machine learning methods for in-depth analysis and classification of surface water quality in Central Java. *Heliyon*, 10(20), e31653.
<https://doi.org/10.1016/j.heliyon.2024.e31653>