

Ecosistema de internet de las cosas para la clasificación de la calidad del agua mediante aprendizaje máquina

Valentín Calzada-Ledesma, Güily Uziel Cruz-Gallo,
Alan Eduardo Stuart Cabrera-Alcalá, Jonathan López-Arellano

Instituto Tecnológico Superior de Purisima del Rincón,
Ingeniería en Informática,
México

`valentin.cl@purisima.tecnm.mx`

Resumen. En este artículo, se propone un sistema de monitoreo y clasificación de la calidad del agua utilizando Internet de las Cosas y Aprendizaje Máquina. A través de la instrumentación de un dispositivo gestionado por un microcontrolador ESP32, se obtienen datos de cuatro sensores con los que se caracteriza la calidad del agua, los parámetros medidos son: temperatura, sólidos disueltos totales, turbidez y nivel de pH. Se consolidó un conjunto de datos (que paralelamente se almacena en la nube mediante un protocolo MQTT) con 60,000 registros repartidos equitativamente en seis clases de agua diferentes. Con éste se entrenaron tres modelos de Redes Neuronales Artificiales tipo Perceptrón Multicapa para clasificar diferentes tipos de agua, y a su vez, determinar si su calidad es apta para consumo humano y/o riego, obteniendo (para este diseño experimental propuesto) una exactitud promedio superior al 95 % bajo un esquema de validación cruzada. La implementación de este tipo de tecnologías puede proporcionar información valiosa para facilitar la toma de decisiones en temas relacionados con la gestión y uso del agua.

Palabras clave: Calidad del agua, redes neuronales artificiales, internet de las cosas, reconocimiento de patrones, ESP32.

Internet of Things Ecosystem for Water Quality Classification Using Machine Learning

Abstract. This article proposes a water quality monitoring and classification system using the Internet of Things and Machine Learning. Through the instrumentation of a device managed by an ESP32 microcontroller, data is obtained from four sensors with which the water quality is characterized, the parameters measured are temperature, total dissolved solids, turbidity, and pH level. A data set was consolidated (also stored in the cloud using an MQTT protocol) with 60,000 records evenly distributed in six different water classes. With this data, three Multilayer Perceptron type Artificial Neural Network models were trained to classify different types of water, and in turn, determine whether its quality is suitable for human consumption or irrigation, obtaining (for this proposed experimental design) an average accuracy greater than 95% under a

cross-validation scheme. The implementation of this type of technology can provide valuable information to facilitate decision-making on issues related to water management and use.

Keywords: Water Quality, Artificial Neural Networks, Internet of Things, Pattern Recognition, ESP32.

1. Introducción

El agua es un recurso hídrico indispensable para la existencia de los seres vivos, y aunque es un recurso natural renovable, su uso no se limita únicamente al consumo humano, sino que también se utiliza para actividades de agricultura, ganadería y producción de energía [11]. Es innegable que el agotamiento y deterioro de las fuentes de recursos hídricos es un fenómeno que no puede ser subestimado. La sobreexplotación de acuíferos, el cambio climático y la variabilidad en los patrones de precipitación, entre otros factores, han incrementado la vulnerabilidad de las reservas de agua, llevando a situaciones críticas de escasez y estrés hídrico en distintas regiones del mundo [4]. En este contexto, México no es una excepción.

La medición de la calidad del agua es de vital importancia debido a su impacto directo en la salud humana, la conservación del medio ambiente y la sostenibilidad de los recursos hídricos. En un contexto donde el acceso a agua potable segura es fundamental para la salud y el bienestar de las comunidades, la capacidad de monitorear y clasificar la calidad del agua es crucial para prevenir enfermedades transmitidas por el agua y garantizar un suministro adecuado para consumo humano y actividades agrícolas [5].

Sin embargo, la clasificación de la calidad del agua presenta desafíos significativos, como la detección de contaminantes, la evaluación de la contaminación a lo largo del tiempo y en diferentes fuentes de agua, y la implementación de tecnologías y métodos de monitoreo adecuados para abordar la complejidad de los ecosistemas acuáticos. Sin embargo, este tipo de análisis usualmente se realizan en laboratorios especializados con equipo profesional, a los cuales puede ser difícil acceder en zonas remotas.

Por otro lado, la falta de datos actualizados y la infraestructura de monitoreo limitada son también obstáculos importantes que dificultan la toma de decisiones informadas y la implementación efectiva de medidas de protección y conservación del agua [1]. En este sentido, es imperativo desarrollar estrategias de monitoreo y clasificación de la calidad del agua, para abordar estos desafíos y asegurar la disponibilidad de agua limpia y saludable para las generaciones presentes y futuras.

Para lidiar con el problema de la falta de laboratorios especializados para el análisis de la calidad del agua, en [3] se establece que sensores de bajo costo integrados con Internet de las Cosas (IoT por sus siglas en inglés), pueden utilizarse para realizar un monitoreo ambiental del agua, dejando claro que al ser dispositivos de bajo costo es necesaria más investigación sobre su precisión y confiabilidad en comparación con equipos profesionales. Por otro lado, en [9] se discute el cómo los ecosistemas de IoT en el contexto del monitoreo de la calidad del agua, pueden aumentar la conciencia social al proporcionar datos accesibles en tiempo real, lo cual es positivo en zonas con

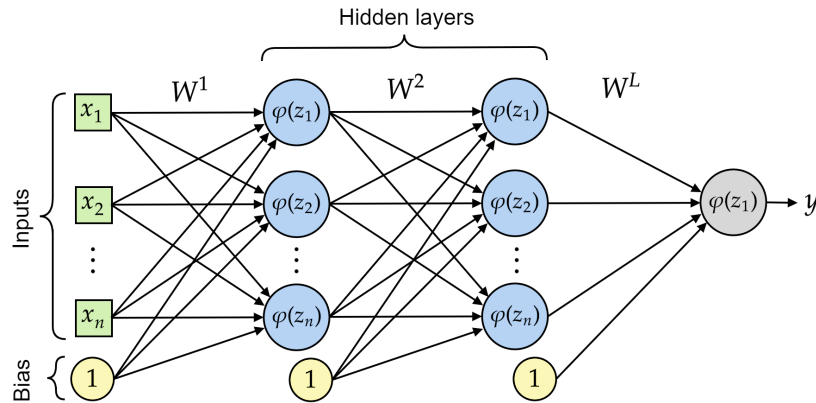


Fig. 1. Red Neuronal Artificial tipo Perceptrón Multicapa.

estrés hídrico. Además, en [12] se propone el uso de Redes Neuronales Artificiales (RNA) para detectar la calidad del agua logrando resultados satisfactorios. Asimismo, en [10] se propuso el modelo CATBoost, ofreciendo una buena precisión para tareas de clasificación de datos de calidad del agua, lo que sugiere que los algoritmos de aprendizaje máquina pueden ser un enfoque confiable para mejorar el monitoreo de la calidad del agua mediante inteligencia artificial.

Motivados por los trabajos antes mencionados, y la necesidad de implementar sistemas relacionados con la cuarta revolución industrial, en este artículo se propone un sistema de Internet de las Cosas (IoT por sus siglas en inglés) para monitorear y clasificar la calidad del agua en tiempo real, esto a través de un sistema de clasificación dirigido por una RNA tipo Perceptrón Multicapa (MLP por sus siglas en inglés). La elección de utilizar un MLP en nuestro estudio se respalda por un análisis en el estado del arte, en donde este tipo de RNAs son ampliamente utilizadas para lidiar con problemas relacionados con la calidad del agua [7, ?,?].

Además, para nuestro caso de estudio, este tipo de RNA presenta versatilidad y una alta capacidad para modelar relaciones no lineales entre variables de entrada y salida, es fácil de implementar, y los modelos de creados tienden a ser más ligeros en términos de tamaño y complejidad computacional. Este último factor es fundamental para nuestra implementación, ya que en comparación con algunas otras arquitecturas de RNAs, como las Redes Neuronales Convolucionales (CNN) o las Redes Neuronales Recurrentes (RNN), un MLP suele ser más rápido en su fase de inferencia. Lo cual es crucial para una aplicación en tiempo real de IoT con conexión a la nube.

La presente propuesta podría representar un avance tecnológico significativo hacia un futuro del uso del agua más eficiente, sostenible y benéfico para la sociedad en general. El artículo se organiza de la siguiente manera. En la Sección 2, se presentan los conceptos relacionados con el trabajo. En la Sección 3, se presenta la metodología. El diseño de experimentos se muestra en la Sección 4. Los resultados se reportan y analizan en la Sección 5. En la Sección 6, se exponen las limitaciones del trabajo seguidas del trabajo futuro. Finalmente, se muestran las conclusiones en la Sección 8.

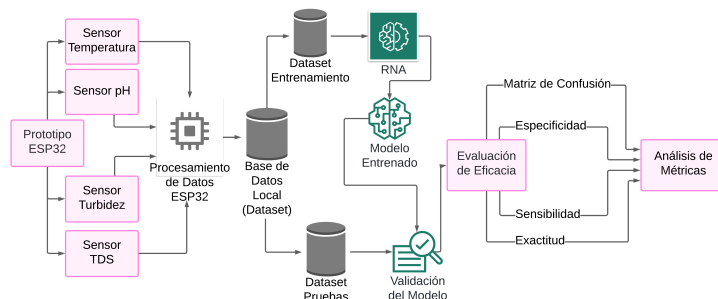


Fig. 2. Esquema general del proceso de aprendizaje máquina.

2. Marco teórico

2.1. Recursos hídricos

Los recursos hídricos cumplen un papel muy importante en nuestra vida diaria, al proporcionar agua dulce para satisfacer las diferentes actividades humanas, sustentar la biodiversidad, y además, para equilibrar el ecosistema [15]. Entre las actividades esenciales humanas se encuentran la agricultura, el consumo humano, la generación de energía, la ingeniería ganadera, y finalmente, la industria química.

Sin embargo, los recursos hídricos cada vez escasean y son vulnerables frente a la contaminación y el cambio climático, por lo que su gestión sostenible es uno de los principales retos que se abordan en la actualidad en México, para garantizar la equidad en el acceso al agua, la protección de los ecosistemas acuáticos y la resiliencia ante los desafíos futuros, como el estrés hídrico.

2.2. Redes neuronales artificiales

Una Red Neuronal Artificial (RNA) es una abstracción matemática inspirada en el funcionamiento del sistema nervioso biológico, específicamente en la forma en que las neuronas interactúan y procesan información [6]. El objetivo principal de una RNA es “aprender” a partir de datos de ejemplo, para realizar tareas específicas como el reconocimiento de patrones o la predicción.

Un tipo de RNA es el Perceptrón Multicapa (MLP por sus siglas en inglés), el cual está compuesto por un conjunto de “neuronas” organizadas en capas interconectadas llamadas capas ocultas (*hidden layers*). Desde un punto de vista matemático, un MLP puede definirse como un grafo dirigido que consta de un conjunto de nodos y de conexiones entre los mismos [6]. La operación matemática básica que realiza un MLP es la siguiente:

$$y = \varphi(Z), \quad (1)$$

donde y es la salida de la neurona, $\varphi(\cdot)$ es una función de activación, usualmente la función ReLU (*Rectified Linear Units*), y Z es un vector (columna) de pre-activaciones z_i que se calcula de la siguiente manera:

$$Z = W \cdot X + b. \quad (2)$$

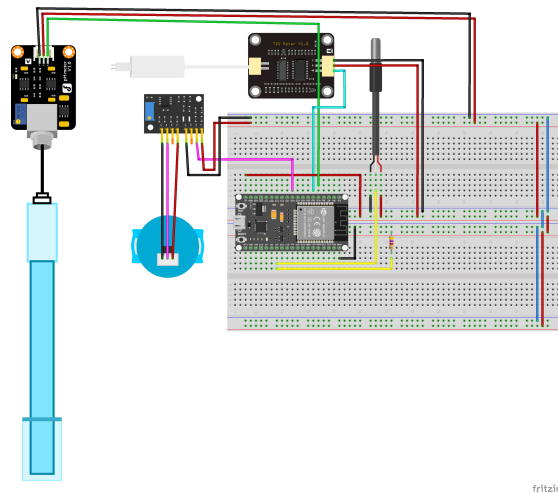


Fig. 3. Instrumentación del ecosistema IoT para el sensor de parámetros.

En este contexto, b es el sesgo (usualmente un vector de 1's), X es un vector columna de n entradas x_i , y W^l es una matriz cuadrada de pesos para cada capa $l \in [1, L]$. En donde cada fila m de la matriz, corresponde a un vector de n pesos de un perceptrón simple o neurona. En la Figura. 1 se muestra el esquema general de un MLP. El algoritmo básico para entrenar a un MLP se llama retropropagación, el cual es un método de optimización que permite ajustar los pesos W de las conexiones entre las neuronas de la RNA para minimizar una función de pérdida, lo que en última instancia mejora el rendimiento del modelo en la tarea que está realizando. Para más información, consulte [13].

2.3. Sensores para medir la calidad del agua

En México, existe la norma mexicana NOM-127-SSA1-1994: “Agua para uso y consumo humano, límites permisibles de calidad y tratamientos a que debe someterse el agua para su potabilización”, y la NOM-CCA/032-ECOL/1993: “Límites máximos permisibles de contaminantes en las aguas residuales de origen urbano o municipal para su disposición mediante riego agrícola”, en ambas normas se establecen los parámetros fundamentales para considerar si el agua es apta o no para consumo humano y para riego. De acuerdo con los parámetros establecidos en las normas, y a su vez, realizando un contraste de los sensores disponibles en el mercado, los parámetros elegidos para evaluar la calidad del agua en este trabajo son: temperatura, sólidos disueltos totales, turbidez y nivel de pH. Estos cuatro parámetros proporcionan una evaluación básica pero importante de la calidad del agua y pueden indicar la presencia de contaminantes, cambios ambientales o condiciones que afectan su idoneidad para el consumo humano. Además, son parámetros que pueden medirse de manera relativamente sencilla y económica, lo que los hace adecuados para monitoreo continuo en tiempo real o en sistemas de detección temprana de problemas de calidad del agua. A continuación, se describen cada uno de éstos.

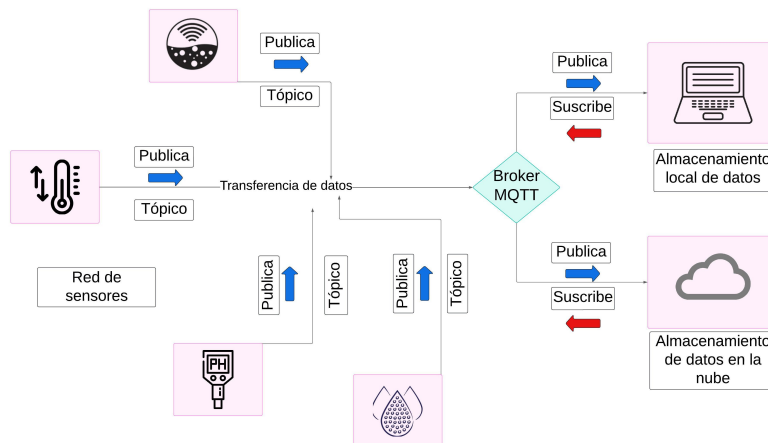


Fig. 4. Diagrama de comunicación MQTT.

- **Sensor de temperatura DS18B20:** Este sensor es capaz de realizar mediciones de la temperatura del agua [14] en un amplio rango de temperaturas desde los -55°C hasta los 125°C , con una alta precisión en sus mediciones. Una de sus características principales es que es sumergible.
- **Sensor de sólidos disueltos totales:** Este sensor indica cuantos miligramos de sólidos solubles se disuelven en un litro de agua. Esto se hace a través del parámetro TDS (Total de Sólidos Disueltos), en el cual, mientras mayor sea su valor, indicará que más sólidos solubles se encuentran dentro del agua, lo que sugiere que el agua está menos limpia [2].
- **Sensor de pH:** Este instrumento ayuda a medir la actividad de iones-hidrógeno en las soluciones basadas en agua, esto se hace a través del parámetro pH que indica la acidez o alcalinidad del agua. Para lograrlo, este sensor mide los niveles de pH usando el potencial eléctrico entre un electrodo de pH y un electrodo de referencia.
- **Sensor de turbidez TSW-20M:** Es un sensor analógico tipo óptico capaz de medir los niveles de turbiedad u opacidad que hay en los líquidos, particularmente agua [2]. Para lograr esto, el sensor detecta las partículas suspendidas en el agua, a través de la tasa de dispersión de la luz.

2.4. Internet de las cosas

En el contexto de la gestión de la calidad del agua, la integración de la tecnología de Internet de las Cosas (IoT por sus siglas en inglés) ofrece una solución innovadora para la recopilación de datos en tiempo real utilizando sensores. En esta sección se exploran tres componentes clave del ecosistema de IoT que juegan un papel fundamental para medir la calidad del agua: el protocolo de comunicación MQ Telemetry Transport (MQTT), la base de datos InfluxDB y la plataforma de integración Node-RED.

Estas herramientas trabajan en conjunto para permitir la captura y el almacenamiento de datos precisos sobre la calidad del agua. Al unir estas tecnologías, se facilita la monitorización continua y la toma de decisiones informadas para la preservación y gestión efectiva de los recursos hídricos.

El Protocolo de Comunicación MQTT, se destaca como el principal protocolo de mensajería para el Internet de las Cosas (IoT), facilitando la publicación y suscripción de datos entre dispositivos IoT a través de la red. Operando bajo el patrón de publicación/suscripción (Pub/Sub), MQTT conecta emisores y receptores mediante tópicos, permitiendo su desacoplamiento y gestionando la conexión a través de un intermediario conocido como broker MQTT. Por otro lado, InfluxDB, una base de datos desarrollada por InfluxData en Go, garantiza un almacenamiento eficiente y una rápida recuperación de datos, especialmente de series temporales, ampliando su aplicación desde la monitorización hasta el IoT y el análisis de datos de sensores. Finalmente, Node-RED emerge como una herramienta de programación innovadora, simplificando la conexión entre dispositivos de hardware, APIs y servicios en línea.

3. Metodología

En esta sección, se describe la metodología utilizada en este trabajo de investigación. De manera general, se instrumenta un dispositivo IoT que sensa cuatro parámetros sobre la calidad de seis tipos de agua diferente. Después, se consolida un conjunto de datos para entrenar un MLP, posteriormente se valida el modelo, y finalmente se obtienen diferentes métricas para conocer su desempeño. En la Figura 2 se muestra el esquema general de la metodología.

3.1. Obtención de datos

Los datos se obtuvieron a través de un dispositivo orquestado por un microcontrolador ESP32 instrumentado con cuatro sensores: el DS18B20 para medir la temperatura del agua, un sensor TDS para medir los sólidos disueltos totales, un sensor de turbidez TSW-20M, y un sensor de pH. Estos sensores se utilizaron para medir y obtener una caracterización cuantitativa de seis tipos de agua diferentes: agua limpia, agua con detergente, agua del grifo, agua con pesticida, agua con cloro y agua de pecera. En la Figura 3, se muestra la instrumentación realizada del dispositivo IoT.

Las mediciones se realizaron bajo un entorno controlado, se utilizaron recipientes de plástico para almacenar las muestras de cada tipo de agua a temperatura ambiente. Para el registro de las lecturas de los sensores se utilizó Arduino IDE, recopilando un total de 10,000 registros por clase de agua, los cuales se almacenaron en un archivo .CSV para su posterior procesamiento en el lenguaje de programación Python, asegurando así una amplia variedad de datos para el posterior entrenamiento del MLP.

Paralelamente, los datos sensados se transmiten mediante un protocolo de comunicación M2M (Machine-to-Machine) MQTT, y son almacenados en la nube utilizando un sistema de gestión de bases de datos de series temporales llamado InfluxDB.

Tabla 1. Muestras de datos obtenidos por los sensores, reescalados entre 0 y 1.

| Temperatura | TDS | Turbidez | PH | Clase | Consumo Humano | Riego |
|-------------|--------|----------|--------|-------|----------------|-------|
| 0.2104 | 0.1948 | 0.0153 | 0.3027 | A | Sí | Sí |
| 0.3687 | 0.9772 | 0.7332 | 0.7891 | B | No | No |
| 0.0421 | 0.1448 | 0.0328 | 0.6141 | C | No | Sí |
| 0.9899 | 0.5325 | 0.3911 | 0.8006 | D | No | No |
| 0.9360 | 0.1910 | 0.2616 | 0.0072 | E | No | No |
| 0.8215 | 0.3144 | 0.4183 | 0.2296 | F | No | Sí |

Este sistema se eligió debido a que puede funcionar en sistemas distribuidos, puede ser accesible a través de protocolos estándar de red, y finalmente, su nivel de seguridad permite autenticación de usuarios. La Figura 4 muestra el diagrama completo de este sistema IoT. Cabe señalarse que hasta este punto del proyecto sólo se realizó el almacenamiento de las series de tiempo, y posteriormente, con esa información se propondrán aplicaciones basadas en la nube, sin embargo, esto se enmarca como trabajo futuro.

4. Diseño de experimentos

4.1. Conjunto de datos

El conjunto de datos consta de 60,000 instancias, cuatro características (una por sensor), y seis clases diferentes de agua. Cada instancia está etiquetada con una clase (designada de la A a la F). Asimismo, se añaden dos variables objetivo (bi-clase) adicionales: “Consumo humano” y “Riego”, las cuales se basan en las normas NOM-127-SSA1-1994 y NOM-CCA/032-ECOL/1993 respectivamente (descritas en la sección Sección 2.3), ampliando la utilidad del conjunto de datos al permitir la evaluación de la calidad del agua para diferentes usos. Esto es de suma utilidad, ya que a partir del mismo conjunto de datos, se pueden entrenar diferentes modelos, uno para cada una de las variables objetivo. Este procedimiento se detalla más adelante.

Es importante mencionar que los datos obtenidos fueron reescalados entre 0 y 1 utilizando el método MinMaxScaler de la librería scikit-learn. Este reescalamiento es importante, ya que asegura que los vectores de características se encuentren en el mismo dominio, permitiendo que el proceso de entrenamiento del MLP sea adecuado. En el Tabla 1 se exponen 6 muestras de datos normalizados, su clase correspondiente, si es apto para consumo humano y si es apto para riego.

4.2. Parámetros de los algoritmos

Como se mencionó anteriormente, el conjunto de datos consta de tres variables objetivo: Clase_Agua, Consumo_Humano y Riego, por lo que se entrenó un modelo de RNA independiente para cada una de estas variables. Además, con la finalidad de conocer la robustez de las arquitecturas de las RNA propuestas, se utilizó

Tabla 2. Arquitecturas de RNA para cada variable objetivo.

| Variable objetivo | Arquitectura RNA |
|-------------------|------------------|
| Clase_Agua | 4, 5, 5, 6 |
| Consumo_Humano | 4, 8, 7, 6 |
| Riego | 4, 5, 5, 6 |

un esquema de validación cruzada con $K = 10$ folds (pliegues) [8]. A continuación, se muestra la configuración de cada una de las arquitecturas. El entrenamiento de los modelos se realizó utilizando el algoritmo de retropropagación y la función de activación ReLU (Rectified Linear Unit). Se estableció un máximo de 1000 épocas para ajustar los pesos de la RNA y minimizar la función de pérdida, en este caso, el error cuadrático medio (MSE por sus siglas en inglés).

Como se mencionó anteriormente, la evaluación del rendimiento de cada modelo se realizó mediante validación cruzada, por lo que se reportan las gráficas del comportamiento de la función de pérdida durante el proceso de entrenamiento, los resultados de exactitud para cada Fold, y la exactitud media de la validación cruzada, para cada variable objetivo. Finalmente, el modelo entrenado se guardó en un archivo utilizando la función `joblib.dump` de `scikit-learn` para su posterior uso en un contexto de IoT en tiempo real.

5. Resultados

Durante el proceso de entrenamiento y evaluación de los modelos de RNA, se obtuvieron resultados para las tres variables objetivo: `Clase_Agua`, `Consumo_Humano` y `Riego`, los cuales se reportan en el Tabla 3. Asimismo, en la Figura 5, se muestran las gráficas del comportamiento de la función de pérdida durante el proceso de entrenamiento de cada uno de los modelos. En la Figura 5, se puede observar cómo las funciones de pérdida convergen en pocas iteraciones, sugiriendo eficacia en los modelos entrenados bajo este diseño experimental.

La alta exactitud, sensibilidad y especificidad, exhibidas en la Tabla 3, indican que los modelos son capaces de capturar los patrones distintivos en los datos de los sensores y utilizarlos para realizar la clasificación de los tipos de agua, y determinar si son aptas para consumo humano y riego. Este nivel de rendimiento es prometedor y sugiere que los modelos pueden ser herramientas útiles para monitorear y evaluar la calidad del agua en diversos entornos.

Es importante mencionar que al utilizar validación cruzada, se puede obtener una estimación más precisa del rendimiento de los modelos, ayudando a evitar su sobreajuste, ya que se exponen a diferentes subconjuntos de datos durante el entrenamiento y la validación.

Con respecto a las arquitecturas propuestas, dichas configuraciones se obtuvieron después de un proceso de experimentación exhaustiva, en donde se observó que con el hecho de reducir la cantidad de neuronas en alguna de las capas ocultas, significaba un decremento de la exactitud promedio de hasta un 30 %. Por otro lado, si se incrementaba la cantidad de neuronas o la cantidad de capas ocultas, el rendimiento

Tabla 3. Resultados de la validación cruzada para las variables objetivo Clase_Agua, Consumo_Humano, y Riego.

| Fold | Clase_Agua | Consumo_Humano | Riego |
|----------------------------|-----------------|----------------|----------------|
| 1 | 1.00 | 0.99 | 1.00 |
| 2 | 1.00 | 0.99 | 1.00 |
| 3 | 1.00 | 0.99 | 1.00 |
| 4 | 1.00 | 0.99 | 1.00 |
| 5 | 1.00 | 0.99 | 1.00 |
| 6 | 1.00 | 0.99 | 1.00 |
| 7 | 1.00 | 0.99 | 1.00 |
| 8 | 1.00 | 0.99 | 1.00 |
| 9 | 1.00 | 0.99 | 1.00 |
| 10 | 1.00 | 0.99 | 1.00 |
| Exactitud media | 100.00 % | 99.07 % | 99.90 % |
| Sensibilidad media | 100.00 % | 98.95 % | 99.90 % |
| Especificidad media | 100.00 % | 98.95 % | 99.90 % |

obtenido no era significativamente mejor, aunque sí incrementaba significativamente el costo computacional. Por lo que, una arquitectura simple, para el contexto de nuestro trabajo, es mejor, ya que se adapta a las necesidades de un sistema de IoT en tiempo real.

Finalmente, se conjetura que los datos obtenidos a través de esta metodología, podrían ser linealmente separables, este hecho es algo que no se esperaba al realizar este proyecto, puesto que intuitivamente se esperaba que los datos estuvieran correlacionados. Sin embargo, explicaría el buen desempeño de la RNA, esto es algo que estudiaremos a detalle en un trabajo futuro.

6. Limitaciones

Durante el desarrollo del proyecto, se identificaron diversas limitaciones que podrían afectar el diseño y la implementación del sistema de monitoreo de la calidad del agua:

- **Volumen de datos limitado:** Para este artículo se obtuvieron un total de 60,000 registros para consolidar el conjunto de datos. Sin embargo, conjeturamos que a medida de que crezca el número de clases, probablemente habrá una disminución en el desempeño del modelo, por lo que será necesario entrenar otro modelo.
- **Inexactitud en las mediciones de los sensores:** La medición de los sensores podría ser imprecisa debido a factores como la temperatura del agua, la intensidad de la luz sobre el agua y los voltajes recibidos por los sensores, estos aspectos pueden influir en los resultados de las mediciones. Además, cambios bruscos entre diferentes tipos de agua pueden afectar la estabilidad y precisión de las lecturas de los sensores, ya que algunos de pestos pueden requerir más tiempo para adaptarse o estabilizarse.

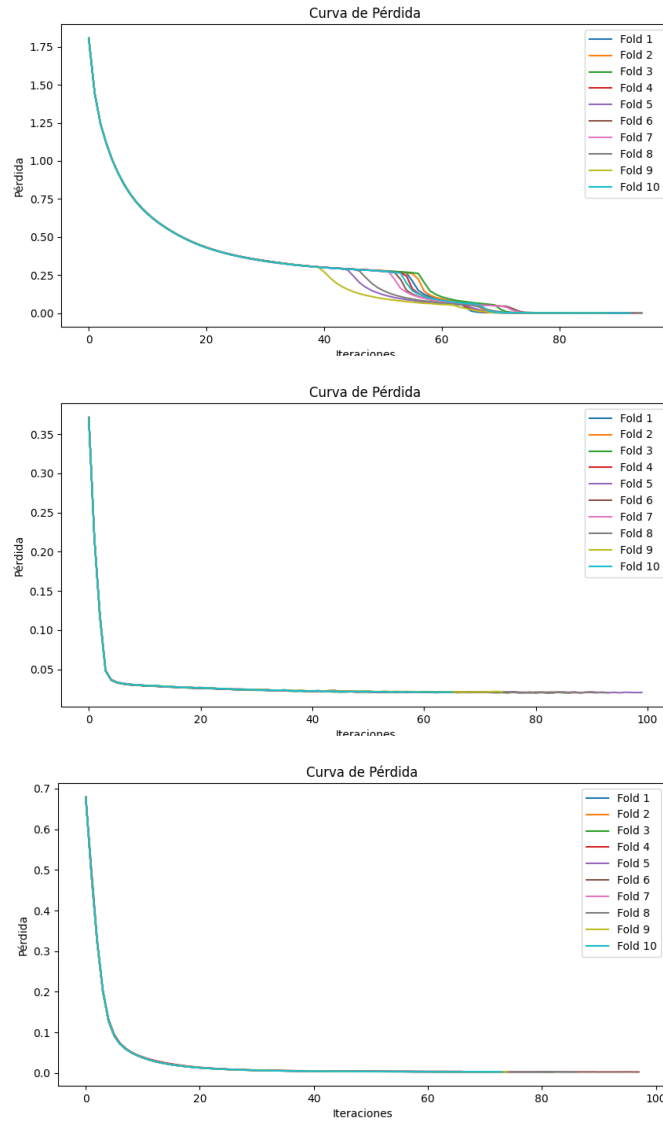


Fig. 5. Comportamiento de la función de pérdida para Clase_Agua, Consumo_Humano y Riego, respectivamente.

7. Trabajo futuro: hacia el cómputo distribuido MPI

En la actualidad, existe una gran variedad de algoritmos que requieren de muchos recursos computacionales para llevar a cabo sus tareas debido a la gran cantidad de procesamiento de datos que se manipulan. Estos tipos de algoritmos necesitan grandes capacidades de procesamiento de datos, que una computadora convencional

e incluso las que están específicamente diseñadas para abordar esas tareas no son suficientes para satisfacer las necesidades o explotar el potencial de las aplicaciones. Para llevar a cabo el entrenamiento de una RNA bajo un esquema de cómputo distribuido utilizando MPI, será necesario recopilar y preparar los datos de todos los sensores, organizándolos y adecuándolos para su procesamiento en un clúster de cómputo distribuido. Posteriormente, se implementarán técnicas de cómputo distribuido utilizando bibliotecas y herramientas especializadas, como TensorFlow o MPI. Estas herramientas permitirán distribuir la carga de trabajo entre los nodos del clúster, lo que acelerará significativamente el proceso de entrenamiento de la Red Neuronal.

8. Conclusiones

La medición precisa de la calidad del agua es crucial para garantizar la salud humana, la conservación del medio ambiente y la sostenibilidad de los recursos hídricos. A pesar de su importancia, enfrenta desafíos significativos, como la detección de contaminantes y la falta de datos actualizados y una infraestructura de monitoreo adecuada. El sistema propuesto en el presente artículo, basado en IoT y aprendizaje máquina, ofrece una solución innovadora para el monitoreo de la calidad del agua en tiempo real.

Los resultados del modelo de la RNA, bajo un esquema de validación cruzada, mostraron una exactitud, sensibilidad y especificidad superiores al 95 %, lo cual indica una alta eficacia para clasificar muestras de agua bajo el esquema experimental propuesto. La arquitectura de la RNA es de bajo costo computacional, por lo que puede implementarse sin dificultad en ecosistema de IoT en tiempo real con transferencia de datos a la nube. La implementación de este tipo de tecnologías podría beneficiar a la sociedad en general, al garantizar la disponibilidad de alimentos de alta calidad y promover la seguridad alimentaria, y a su vez, podría contribuir a un futuro de la gestión del agua más eficiente y sostenible.

Agradecimientos. Se agradece al Tecnológico Nacional de México por el financiamiento brindado que hizo posible la realización de este trabajo. Además, se extiende el reconocimiento al ITS de Purísima del Rincón por su invaluable apoyo, fundamental para culminar con éxito este proyecto.

Referencias

1. Balasooriya, B. K., Rajapakse, J., Gallage, C.: A review of drinking water quality issues in remote and indigenous communities in rich nations with special emphasis on australia. *Science of The Total Environment*, vol. 903, pp. 166559 (2023) doi: 10.1016/j.scitotenv.2023.166559
2. Conejeros-Molina, A., Hueichaqueo-Pichunman, C., Martínez-Jimenez, B. L., PlaceresRemior, A.: Monitoreo de calidad del agua en sistema de agua potable rural. *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 42, no. 3, pp. 60–70 (2021)
3. de-Camargo, E. T., Spanhol, F. A., Slongo, J. S., da Silva, M. V. R., Pazinato, J., de-Lima-Lobo, A. V., Coutinho, F. R., Pfrimer, F. W. D., Lindino, C. A., Oyamada, M. S.: Low-cost water quality sensors for IoT: A systematic review. *Sensors*, vol. 23, no. 9, pp. 4424 (2023) doi: 10.3390/s23094424

4. Du-Plessis, A.: Current and future water scarcity and stress. Water as an inescapable risk: current global water availability, quality and risks with a specific focus on South Africa, pp. 13–25 (2019) doi: 10.1007/978-3-030-03186-2_2
5. Forget, G., Sanchez-Bain, W. A.: Managing the ecosystem to improve human health: Integrated approaches to safe drinking water. *International Journal of Occupational and Environmental Health*, vol. 5, no. 1, pp. 38–50 (1999) doi: 10.1179/oeh.1999.5.1.38
6. Gardner, M. W., Dorling, S. R.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, vol. 32, no. 14–15, pp. 2627–2636 (1998) doi: 10.1016/S1352-2310(97)00447-0
7. Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A. A., Mohamed, A., Ashraf, I.: Water quality prediction using KNN imputer and multilayer perceptron. *Water*, vol. 14, no. 17, pp. 2592 (2022) doi: 10.3390/w14172592 <https://www.mdpi.com/2073-4441/14/17/2592>
8. Kärkkäinen, T.: On cross-validation for MLP model evaluation. In: *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 291–300 (2014) doi: 10.1007/978-3-662-44415-3_30
9. Miller, M., Kisiel, A., Cembrowska-Lech, D., Durlík, I., Miller, T.: IoT in water quality monitoring—are we really here? *Sensors*, vol. 23, no. 2, pp. 960 (2023) doi: 10.3390/s23020960
10. Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., Al-Shamma'a, A.: Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, vol. 48, pp. 102920 (2022) doi: 10.1016/j.jwpe.2022.102920
11. Owusu, P. A., Asumadu-Sarkodie, S., Ameyo, P.: A review of Ghana's water resource management and the future prospect. *Cogent Engineering*, vol. 3, no. 1, pp. 1164275 (2016)
12. Rodriguez-Perez, J., Leigh, C., Liqueur, B., Kermorvant, C., Peterson, E., Sous, D., Mengersen, K.: Detecting technical anomalies in high-frequency water-quality data using artificial neural networks. *Environmental Science & Technology*, vol. 54, no. 21, pp. 13719–13730 (2020) doi: 10.1021/acs.est.0c04069
13. Rojas, R.: The backpropagation algorithm. *Neural Networks: A Systematic Introduction*, pp. 149–182 (1996) doi: 10.1007/978-3-642-61068-4_7
14. Sierra García, L. A.: Diseño del sistema de medición y despliegue de temperaturas con el sensor DS18B20 mediante el protocolo de transmisión 1-WIRE. Ph.D. thesis, Universidad de San Carlos de Guatemala (2017)
15. Torregrosa, M. L., Mora, R. D., Cisneros, B. J., Michel, E. K., Austria, P. M., Cedillo, J. L. M., Viqueira, J. P., Calleros, A. R., Monjardín, L. C. R., Martelo, E. Z.: Los recursos hídricos en México. *Diagnóstico del Agua en las Américas*, pp. 309 (2012)