

# Un mapeo sistemático sobre predicción de calidad del agua mediante técnicas de inteligencia computacional

Iván Darío López\*  
Apolinar Figueroa\*\*  
Juan Carlos Corrales\*\*\*

Recibido: 29/07/2015 • Aceptado: 11/12/2015  
DOI: 10.22395/rium.v15n28a2

## **Resumen**

Dada la naturaleza renovable del agua, este recurso se ha tratado y gestionado tradicionalmente como si fuese ilimitado; sin embargo, el incremento indiscriminado de su uso ha acarreado consigo un acelerado deterioro en su calidad; es así como la predicción de la calidad del agua desempeña un papel muy importante para muchos sectores socio-económicos que dependen del uso del preciado líquido. En este estudio se realiza un mapeo sistemático de la literatura concerniente a la predicción de la calidad del agua haciendo uso de técnicas de inteligencia computacional, que incluye aquellas utilizadas para calibrar modelos predictivos en aras de mejorar su precisión. A partir de las preguntas de investigación formuladas en el mapeo sistemático es identificada una brecha orientada a la creación de un mecanismo adaptativo de predicción de calidad del agua que pueda ser aplicado en diferentes usos del agua sin que la precisión de las predicciones se vea afectada.

**Palabras clave:** calidad del agua, inteligencia computacional, predicción, sistemas adaptativos complejos

---

\* Magister en Ingeniería Telemática, estudiante de Doctorado en Ingeniería Telemática con beca de Colciencias, Investigador del Grupo de Ingeniería Telemática, Universidad del Cauca - Campus de Tulcán, Universidad del Cauca, Tel: (+57-2) 833-1346, e-mail: navis@unicauca.edu.co.

\*\* PhD. en Ciencias Biológicas, Investigador Asociado del Grupo de Estudios Ambientales, Universidad del Cauca - Carrera 2 # 1A-25, Universidad del Cauca, Tel: (+57-2) 820-9800 ext. 2607, e-mail: apolinar@unicauca.edu.co, Fax: (+57-2) 820-9860.

\*\*\* PhD. en Ciencias de la Computación, Investigador Senior del Grupo de Ingeniería Telemática, Universidad del Cauca - Campus de Tulcán, Universidad del Cauca, Tel: (+57-2) 820-9900 ext. 2129, e-mail: jcorral@unicauca.edu.co, Fax: (+57-2) 820-9900 ext. 2102

## A systematic mapping of water quality prediction using computational intelligence techniques

### **Abstract**

Due to the renewable nature of water, this resource has been treated and managed as if it were unlimited; however, increase the indiscriminate use has brought with it a rapid deterioration in quality; so as predicting water quality has a very important role for many socio-economic sectors that depend on the use of the precious liquid. In this study, a systematic literature mapping was performed about water quality prediction using computational intelligence techniques, including those used to calibrate predictive models in order to improve accuracy. Based on research questions formulated in the systematic mapping, a gap is identified oriented to creation of an adaptive mechanism for predicting water quality that can be applied in different water uses without raised the accuracy of the predictions is affected.

**Key words:** water quality, computational intelligence, forecasting, complex adaptive systems

## INTRODUCCIÓN

El agua se ha tratado y gestionado tradicionalmente como si fuese un recurso ilimitado debido a su naturaleza renovable; sin embargo, el incremento indiscriminado de su uso ha acarreado consigo un acelerado deterioro en su calidad y, en ocasiones, se ven cambios en su distribución temporal y espacial, con consecuencias que no son previstas por completo, pero de una importante trascendencia ambiental, económica, social, entre otras [1].

La calidad del agua puede definirse como el conjunto de características físicas, químicas y biológicas de los cuerpos de agua superficiales y subterráneos, sin haber realizado un tratamiento previo, es decir, sin la intervención humana [2]. Por otra parte, los expertos en el manejo del recurso hídrico indican que no existe una única definición de la calidad del agua, pues esta depende estrictamente del uso al que esté destinado el preciado líquido [3].

Las principales situaciones problemáticas sobre la calidad del agua obedecen, en gran medida, al crecimiento de la población y de las actividades económicas, lo que requiere monitorización y control constantes que permitan tomar las acciones necesarias para abordar esta problemática y disminuir su impacto en los procesos naturales y sociales, especialmente en la salud humana [4]. Sin embargo, en muchas ocasiones para controlar unas condiciones apropiadas de calidad del agua, no es suficiente con establecer actividades de monitorización que brinden acciones correctivas ante determinado tipo de contaminación; además de esto, es necesario contar con modelos o mecanismos que permitan anticiparse a la materialización del riesgo con el suficiente rango de tiempo para prevenir los efectos negativos que afecten la calidad del recurso hídrico.

De acuerdo con lo anterior, el concepto de predicción científica se centra en una declaración precisa de lo que ocurrirá en determinadas condiciones específicas; de esta manera, su validez se mide por el éxito o acierto que tengan sus predicciones. Sin embargo, existe una multitud de campos en la ciencia donde la predicción se convierte en una tarea compleja, bien sea por el gran número de variables involucradas o por la misma dinámica desconocida de los fenómenos involucrados en una situación problemática [5].

En este sentido, la predicción de la calidad del agua desempeña un papel importante para muchos sectores socio-económicos que dependen del uso del preciado líquido. Es así como en los últimos años, el campo de la inteligencia artificial (IA) [6] ha introducido algoritmos y técnicas de predicción que cuentan con la capacidad de estimar las condiciones futuras de un cuerpo de agua con base en el análisis de los datos que han sido recolectados en el pasado. Entre las principales técnicas utilizadas están

las redes neuronales artificiales (RNA), las máquinas de vector de soporte (MVS), la programación evolutiva (PE), la inteligencia de enjambres (IE) y los sistemas inmunes artificiales (SIA) [7].

Con base en lo anterior, es importante destacar los avances y aportes de estos trabajos dentro del campo de la predicción de la calidad del agua; sin embargo, ninguno de ellos considera la aplicación de su modelo de predicción sobre diferentes usos del recurso hídrico. Los trabajos encontrados se limitan a realizar el proceso de predicción tomando como referencia los datos pertenecientes a un determinado uso del agua, en ocasiones, sesgando los resultados hacia ese escenario en particular, sin tener en cuenta la aplicabilidad que el mismo modelo de predicción pueda tener para otros usos, lo cual puede representar beneficios para diversos sectores que utilizan el agua en sus procesos productivos.

Para abordar la temática expuesta, este artículo se estructura de la siguiente forma: la sección 1 describe el método de investigación empleado para construir el mapeo sistemático; la sección 2 presenta los trabajos relacionados con los ejes temáticos principales que se abordan alrededor de la predicción de la calidad del agua; los resultados son presentados en la sección 3, y discutidos en la sección 4; finalmente, la sección 5 expone las conclusiones del mapeo sistemático de la literatura.

## 1. MATERIALES Y MÉTODOS

Para establecer un punto de partida alrededor del conocimiento generado en el campo de la predicción de la calidad del agua haciendo uso de Inteligencia Computacional (IC), en este artículo se presenta un mapeo sistemático de la literatura referente a este eje temático. Tomando como referencia la propuesta realizada por Kitchenham [8] para la construcción de una revisión sistemática de la literatura orientada al área de la ingeniería del software, este artículo adopta como guía esta metodología aplicando la adaptación que se propone en el trabajo de Petersen [9], el cual está basado en dicho estudio con la diferencia de que está planteado para el desarrollo de un mapeo sistemático de la literatura. Este trabajo establece un procedimiento conformado por cinco etapas: A) definir preguntas de investigación, B) realizar la búsqueda literaria, C) seleccionar estudios, D) clasificar artículos y E) extraer y realizar la agregación de datos. A continuación, se describe cada uno de los pasos para la realización del mapeo sistemático.

A. **Preguntas de investigación.** El objetivo principal de este estudio se centra en conocer las técnicas de IC más utilizadas dentro del campo de aplicación de la predicción de la calidad del agua y cuál es la tendencia en los nuevos trabajos de investigación. Para obtener un conocimiento más detallado y una visión integral del tema, se plantean las siguientes preguntas de investigación:

RQ.1. ¿Qué temas interesan a la comunidad científica respecto a la predicción de la calidad del agua aplicando técnicas de IC dentro de un marco temporal reciente?

RQ.2. ¿Qué técnicas de IC se han utilizado para brindar capacidad adaptativa a los mecanismos de predicción?

RQ.3. ¿Cuáles son las técnicas de IC que aún no han sido exploradas en su totalidad y que pueden ser una importante alternativa dentro de la predicción adaptativa de la calidad del agua?

B. **Fuente de datos y estrategia de búsqueda.** Para implementar la estrategia de búsqueda fue tenida en cuenta la terminología referente a la predicción de la calidad del agua principalmente en el idioma inglés, en el cual se encuentran escritos la mayor parte de los trabajos. Las cadenas de búsqueda utilizadas son presentadas en la tabla 1. Por otro lado, las bases de datos utilizadas para realizar la búsqueda fueron establecidas a partir de la importancia de cada una dentro de la comunidad científica al nivel internacional; estas bases de datos son: Elsevier, IEEE, ACM y Springer.

Tabla 1. Precisión de las cadenas de búsqueda

| <i>Cadena de Búsqueda</i>  | <i>N.º Trabajos</i> | <i>Motivo Exclusión</i> | <i>N.º Trabajos Seleccionados</i> | <i>Precisión</i> |
|--|---------------------|-------------------------|-----------------------------------|------------------|
| water quality prediction   | ~1000               | HR, LP                  |                                   | ~0 %             |
| water quality forecast   | ~1000               | HR, LP                  |                                   | ~0 %             |
| <b>water quality prediction AND computational intelligence</b>             | <b>20</b>           |                         | <b>10</b>                         | <b>50 %</b>      |
| <b>water quality prediction AND artificial intelligence</b>                | <b>25</b>           |                         | <b>18</b>                         | <b>72 %</b>      |
| water quality parameters calibration                                       | 347                 | LP                      |                                   | ~0 %             |
| water quality parameters optimization                                      | 170                 | LP                      |                                   | ~0 %             |
| <b>water quality parameters calibration AND computational intelligence</b> | <b>17</b>           |                         | <b>6</b>                          | <b>35.3 %</b>    |
| water quality parameters calibration AND artificial intelligence           | 13                  |                         | 5                                 | 38.5 %           |
| Total  |                     |                         | 39                                |                  |

En negrita las cadenas de búsqueda tenidas en cuenta para el mapeo sistemático

Fuente: elaboración propia.

- C. **Selección de estudios.** Dentro del proceso de selección de estudios se establecen dos criterios para determinar la relevancia de los mismos:

*Inclusión:* aquellos trabajos que se centran en predicción de la calidad del agua haciendo uso de técnicas de IC y también los que utilizan técnicas de inteligencia artificial en general.

*Exclusión:* los trabajos que no contengan el término “predicción”, “calidad del agua”, “inteligencia computacional”, “inteligencia artificial” o alguna técnica de estos campos especificada directamente (ej. redes neuronales, algoritmos genéticos, entre otras).

El criterio de exclusión permite descartar algunas de las cadenas de búsqueda debido a que su utilización presenta problemas como el high recall (HR), en el cual se obtiene la información relevante, pero, además, una cantidad enorme de información que no es de utilidad para el estudio; y, por otro lado, low precision (LP), al obtenerse muy pocos trabajos relevantes en comparación con el total de trabajos encontrados. La tabla 1 muestra las cadenas de búsqueda utilizadas para seleccionar los trabajos relevantes dentro de la temática del presente estudio.

La cadena con la cual se obtuvieron los mejores resultados en cuanto a precisión fue “water quality prediction AND artificial intelligence” con un 72 % en la precisión. Finalmente, en esta etapa fueron seleccionados 39 trabajos relevantes para el presente estudio.

- D. **Clasificación de artículos.** El proceso de clasificación de trabajos está basado en la agrupación planteada por Engelbrecht [7] en la cual se establecen varios paradigmas de IC y en cada uno de ellos se han distribuido los trabajos seleccionados en el paso C de acuerdo con su relación con un paradigma determinado (ver tabla 2).

Tabla 2. Clasificación de trabajos según paradigma de IC

| <i>Paradigma de IC</i>                | <i>N.º Trabajos</i> | <i>Porcentaje</i> |
|---------------------------------------|---------------------|-------------------|
| Redes Neuronales Artificiales (RNA)   | 15                  | 38.5 %            |
| Máquinas de Vectores de Soporte (MVS) | 10                  | 25.7 %            |
| Computación Evolutiva (CE)            | 5                   | 12.8 %            |
| Inteligencia de Enjambres (IE)        | 7                   | 17.9 %            |
| Sistemas Inmunes Artificiales (SIA)   | 2                   | 5.1 %             |

Fuente: elaboración propia

Como se puede apreciar en la tabla anterior, el paradigma que agrupa mayor número de trabajos es el de RNA. En un menor porcentaje, pero muy significativo, se encuentran los trabajos concernientes a la utilización de las MVS, en los cuales se realiza una comparación con las RNA obteniendo resultados más precisos en cuanto a la predicción.

Finalmente, una menor cantidad de trabajos se encuentra agrupada en los paradigmas de CE, IE y SIA.

- E. **Extracción de datos y síntesis.** Para obtener información detallada y concisa de cada trabajo que ha sido seleccionado y clasificado previamente, se elabora una planilla de extracción de datos por cada estudio; en esta planilla se consigna su información principal como el título, el resumen, las palabras clave y lo más importante, los aportes y ventajas, así como las falencias que derivan al final en las brechas relacionadas con el problema de investigación.

## 2. TRABAJOS RELACIONADOS

La predicción de la calidad del agua ha cobrado una gran importancia en los últimos años, y en muchos campos de estudio se reporta una significativa cantidad de trabajos que han abordado este problema a partir de diferentes enfoques. Uno de los campos de la ciencia más prometedores es la inteligencia artificial, dentro de la cual encontramos dos ramas: la inteligencia artificial convencional (IAC) [6], la cual está basada en el análisis formal y estadístico del comportamiento humano ante diferentes problemas, y la inteligencia computacional, que implica desarrollo o aprendizaje interactivo, y dicho aprendizaje se realiza con base en datos empíricos.

Teniendo en cuenta lo anterior, la IC presenta características importantes para el proceso de predicción, las cuales no son tenidas en cuenta en la IAC. Dichas características están enfocadas en mejorar la inteligencia humana aprendiendo y descubriendo nuevos patrones, relaciones y estructuras complejas en ambientes dinámicos para resolver problemas prácticos [10]. De este modo, la mayor parte de los trabajos que se presentan a continuación se centran en la predicción de la calidad del agua utilizando diferentes técnicas de IC. De igual manera, se consideran los trabajos cuyo objetivo es la calibración de parámetros para incrementar la precisión de los mecanismos de predicción a través de la IC.

- A. **Predicción de la calidad del agua mediante IC.** En la actualidad empiezan a cobrar importancia los métodos inspirados en los procesos biológicos de la naturaleza, incluido el ser humano. Estas nuevas técnicas surgen como una alternativa prometedora en muchos campos de aplicación, gracias a su capacidad

de adaptarse muy bien a los cambios en su entorno, imitando la forma en que lo hacen los diversos sistemas naturales. Estos mecanismos incluyen los paradigmas de la IA que muestran capacidad de aprender o adaptarse a nuevas situaciones, generalizar, abstraer, descubrir y asociar. Las técnicas individuales de estos paradigmas de IC se han aplicado con éxito para resolver problemas del mundo real; sin embargo, la tendencia actual es el desarrollo de paradigmas híbridos, ya que ningún paradigma es superior a los demás en todas las situaciones. Al hacerlo, se aprovechan los puntos fuertes de los componentes del sistema de IC híbrido, y se eliminan las debilidades de los componentes individuales [7].

En el área de la IC se destacan las investigaciones realizadas utilizando las técnicas como las redes neuronales, máquinas de vectores de soporte, computación evolutiva, entre otros. Dentro de las técnicas de aprendizaje automático se ha trabajado en la predicción de parámetros biológicos a partir de parámetros físico-químicos [11] y viceversa [12]. En estos dos trabajos son presentados diferentes modelos predictivos por cada parámetro; estos modelos están basados en árboles de regresión [13] y son comparados con técnicas como “vecino más cercano” y regresión lineal; los resultados son muy similares en términos de precisión; sin embargo, los árboles de regresión cuentan con la ventaja de ser más fáciles de interpretar que las dos técnicas mencionadas. Džeroski [12] propone un árbol de regresión por cada variable, mientras que Blockeel [14] analiza estos árboles y plantea disponer de un solo árbol para predecir más de una variable a la vez sin tener pérdida en la precisión. Finalmente cabe mencionar que un inconveniente al utilizar los árboles de regresión es el elevado tiempo de procesamiento en la ejecución.

Continuando con las técnicas de aprendizaje automático aplicadas a la predicción de la calidad del agua, están los métodos de ensamble [15], específicamente el enfoque de los algoritmos voraces de selección de ensamble expuesto en [16], mediante los cuales se agrupan varios conjuntos de datos y por cada uno de ellos se realiza una predicción del valor de una variable fisicoquímica (temperatura, pH, conductividad y salinidad). A su vez, Tan propone en [17], un método de mínimos cuadrados apoyado de MVS para la predicción de parámetros de calidad del agua. Este trabajo enfoca la predicción en el total de fósforo del agua a través de un método que plantea ventajas tales como la capacidad de predecir el valor real, la optimización global y una alta velocidad de operación.

Por otra parte, el problema de predicción de parámetros de calidad del agua se ha abordado ampliamente dentro de las técnicas de redes neuronales artificiales (RNA) [18], computación evolutiva (CE) [19] y sistemas de lógica difusa (SLD) [20]. El campo de las RNA es uno de los que más se ha explorado con miras de



abordar el problema mencionado anteriormente. Dentro de los trabajos realizados en este campo, se puede observar el propuesto por Romero y Shan [21], en el cual desarrollan una herramienta de software basada en redes neuronales para la predicción de la temperatura del agua de descarga del canal en una planta de energía de carbón. Las variables consideradas en este sistema incluyen los parámetros de funcionamiento de la planta y las condiciones climáticas locales, además de información correspondiente a mareas.

Por otro lado, Hatzikos [22] utiliza las redes neuronales con neuronas de activación como herramienta de modelado para la predicción de la calidad del agua del mar. El enfoque propuesto predice si el valor de cada variable aumentará o decaerá en el día siguiente. Los experimentos se centran en cuatro indicadores de calidad, a saber, la temperatura del agua, el pH, la cantidad de oxígeno disuelto y la turbidez. Por su parte, Palani [23] utiliza las RNA para predecir y pronosticar las características cuantitativas de los cuerpos de agua en una zona costera de Singapur, de tal forma que facilite la rápida evaluación y pronóstico de determinadas variables de calidad del agua como salinidad, temperatura, oxígeno disuelto, y clorofila-a.

Por otra parte, el objetivo del trabajo de Aguilera [24] es evaluar la red neuronal de Kohonen [18] como una posible herramienta en el proceso de toma de decisiones para predecir el estado trófico de las aguas costeras; esta herramienta se basa en una serie de mapas de activación compuestos por un conjunto de neuronas, las cuales se activan formando diferentes patrones que determinan la calidad del agua. He [25], propone una red neuronal con una conexión *feed-forward* y un entrenamiento con propagación hacia atrás (en aprendizaje supervisado), para predecir parámetros biológicos como el Indicador de bacteria fecal, coliformes totales, coliformes fecales y enterococo, a partir de parámetros físico-químicos como el pH, la temperatura, la conductividad, la turbidez y el oxígeno disuelto. El dominio de aplicación de este enfoque son las aguas para uso recreacional en las zonas costeras de San Diego, USA.

En RNA, finalmente se destacan los trabajos de Gutiérrez [22], García [23] y Saint-Gerons [24], los cuales se desarrollan sobre calidad del agua en ríos colombianos, mexicanos y españoles, respectivamente. En el primero los autores implementan una metodología de RNA como herramienta para la estimación de la calidad del agua en la cuenca alta y media del río Bogotá. En el segundo trabajo se desarrolla un modelo de RNA para predecir el transporte de contaminantes en aguas subterráneas; para esto los investigadores hacen uso de la retro-propagación en una estructura de 3 capas con datos de entrenamiento, los cuales son evaluados a partir de la ecuación propuesta por Ogata y Banks [25]. Por último, en el tercer trabajo, se monitoriza el oxígeno disuelto mediante el desarrollo de un sensor

software que hace uso de RNA; este utiliza como entradas tres variables medibles: temperatura, turbidez y potencial rédox.

- B. ***Calibración de técnicas de predicción de calidad del agua.*** Los trabajos encontrados en este núcleo temático están enfocados principalmente en la optimización o calibración de modelos de monitorización de parámetros de calidad del agua; dichos trabajos hacen uso de algoritmos bio-inspirados convirtiéndolos en un amplio espacio de investigación en este eje temático.

QUAL2Kw [26] es una herramienta para la simulación de la calidad del agua en arroyos y ríos; incluye un algoritmo genético para facilitar la calibración del modelo en su aplicación a cuerpos de agua particulares. El algoritmo genético es utilizado para encontrar la combinación entre parámetros de velocidad cinética y constantes, lo que se traduce en un mejor ajuste para un modelo de aplicación en comparación con los datos observados. En el mismo sentido, Liu [27] propone un algoritmo genético para calibrar un modelo de contaminación difuso, y es implementada una herramienta de indicadores de fósforo (Phosphorus Indicators Tool, PIT), en la cual fueron utilizados 78 parámetros para predecir la pérdida anual de fósforo. Una de las principales dificultades de este trabajo es la gran cantidad de parámetros utilizados en la predicción; a pesar de esto, los tiempos de cálculo fueron razonables.

Continuando con la aplicación de los algoritmos genéticos, el estudio de Huang [28] tiene como objetivo el acoplamiento de un algoritmo genético híbrido (AGH) y una RNA para la calibración multi-objetivo de modelos de calidad de agua superficial. Este enfoque tiene la ventaja de contar con una mayor eficiencia al realizar la evaluación de la función objetivo; sin embargo, a pesar de que las RNA cuentan con una alta capacidad de generalización, no siempre se garantiza hallar los valores óptimos de los parámetros.

Chau [29] presenta la aplicación de un modelo de optimización por nubes de partículas (PSO) por fracción de paso, para entrenar perceptrones con el fin de predecir la dinámica en tiempo real de proliferación de algas en Tolo Harbour, Hong Kong. Los resultados indican que, en comparación con el algoritmo de propagación hacia atrás de RNA, el algoritmo propuesto alcanza una mayor precisión (en 3 escenarios propuestos, los coeficientes de correlación son superiores respecto a los valores reales) en un tiempo mucho más corto (el tiempo de entrenamiento fue inferior en aproximadamente un 50 %).

En el trabajo de Baltar [30], se aplica un algoritmo multi-objetivo de optimización por nubes de partículas (MOPSO, por su sigla en inglés) para encontrar soluciones que permitan minimizar las desviaciones de variables objetivo tales como

temperatura, oxígeno disuelto, sólidos totales disueltos y pH. Además de lo anterior, fue desarrollado un método gráfico interactivo que permite al usuario tomador de decisiones identificar la mejor solución para ser aplicada en su modelo o, bien, soluciones similares que pueden ser de gran importancia en el proceso de toma de decisiones.

Afshar [31] implementa un modelo de calibración de la calidad del agua determinando parámetros pertenecientes a modelos de hidrodinámica y de calidad del agua. El autor utiliza la técnica de nubes de partículas como una herramienta de optimización. Finalmente, Zhangzan [32] propone un método que combina la técnica de la percepción visual (PV) con la selección negativa (SN) en la teoría de los sistemas inmunes artificiales. La técnica de PV se utiliza para adquirir el índice de calidad del agua (ICA), mientras que la técnica de la SN es usada para evaluarlo.

### 3. RESULTADOS

Aplicando los criterios de selección establecidos anteriormente (inclusión y exclusión), se obtuvo un conjunto de 39 estudios que se encuentran directamente relacionados con el tema de investigación. Como resultado del mapeo sistemático se establecen las respuestas a las preguntas de investigación que se plantearon previamente.

Teniendo en cuenta la pregunta de investigación RQ.1 referente a los temas de interés para la comunidad científica dentro de una ventana de tiempo reciente (últimos 10 años), la predicción de la calidad del agua representa un amplio campo de investigación en el cual los autores enfocan sus trabajos alrededor de diferentes subtemas específicos. Por consiguiente, se ha establecido una clasificación con los tópicos hacia los cuales se orientan con mayor frecuencia los trabajos seleccionados y, además, los temas de investigación planteados como trabajo a futuro. En la tabla 3 se presenta la clasificación mencionada.

Tabla 3. Tópicos de interés científico

| <i>Tópico</i>  | <i>N.º Trabajos</i> | <i>Porcentaje</i> |
|--|---------------------|-------------------|
| Combinación de técnicas para predicción (modelos híbridos) | 18                  | 46.1 %            |
| Calibración de parámetros de entrada                       | 12                  | 30.8 %            |
| Evaluación de incertidumbre en modelos de predicción       | 3                   | 7.7 %             |
| Comparación de modelos de predicción                       | 6                   | 15.4 %            |

Tópicos de mayor importancia para la comunidad científica en la predicción de la calidad del agua

Fuente: elaboración propia

La clasificación anterior permite establecer una tendencia hacia la utilización de modelos híbridos para la predicción de la calidad del agua, es decir, la combinación de diferentes técnicas que, aplicadas conjuntamente, permiten obtener mejores resultados respecto a las técnicas que son utilizadas de forma independiente. Por otro lado, la calibración de parámetros es un tema que también cobra gran importancia en la actualidad y que persigue el mismo objetivo de implementar modelos de predicción más precisos.

Para responder a la pregunta RQ.2 acerca de las técnicas de IC que se han utilizado para dotar con capacidad adaptativa a los mecanismos de predicción, varios trabajos proponen principalmente la adaptación aplicando estos mecanismos sobre diferentes escalas de tiempo como en el trabajo presentado en [15], el cual utiliza un algoritmo de proyección y un algoritmo de mínimos cuadrados para adaptar las predicciones del día actual a las predicciones del día siguiente. Otros estudios se centran en obtener valores óptimos en los parámetros de un modelo para conseguir que los valores predichos se ajusten de una mejor forma a los valores reales como en [27–30] y [33]. Sin embargo, ninguno de estos trabajos considera la predicción adaptativa de la calidad del agua desde el punto de vista de utilizar un mecanismo de predicción en diferentes usos del agua.

Finalmente para dar respuesta a la pregunta RQ.3 acerca de cuáles son las técnicas de IC que aún no han sido exploradas en su totalidad y que pueden ser una importante alternativa dentro de la predicción adaptativa de la calidad del agua, existen estudios que destacan la importancia de técnicas como la optimización por nubes de partículas (PSO) [30, 31, 34], la cual es utilizada ampliamente en problemas donde las soluciones cambian dinámicamente; esta podría ser una característica importante para aplicar un mecanismo de predicción de calidad del agua a diferentes usos, tomando en cuenta que las características de determinado cuerpo de agua varían de acuerdo con su uso. Por otro lado, la computación evolutiva también proporciona la capacidad de adaptación permitiendo a la comunidad sobrevivir ante un cambio dinámico del entorno. De este modo los trabajos desarrollados en [26–28, 33] muestran interesantes propuestas de investigación aplicando la técnica de algoritmos genéticos.

#### 4. DISCUSIÓN

A partir del mapeo sistemático desarrollado en el presente estudio, a continuación son establecidas algunas observaciones respecto a las ventajas y a las falencias de los diferentes trabajos seleccionados.

Como un importante aporte que brindan la mayor parte de los trabajos consultados, está el uso de modelos híbridos para abordar el tema de predicción planteado. Varios estudios proponen la combinación de las RNA con los modelos ARIMA como en [35] y [36], debido a que en muchas ocasiones los datos de entrada constan de

complejos patrones lineales y no lineales teniendo en cuenta que los modelos ARIMA no muestran una alta precisión en las predicciones realizadas; esta es una ventaja que poseen las RNA, las cuales tienen la capacidad de reconocer patrones no lineales que incluyen eventos extremos (alteraciones en la calidad del agua sin un determinado patrón de ocurrencia).

Por otro lado, las MVS muestran un importante aporte respecto a las RNA dados sus buenos resultados en la precisión de los valores predichos; cabe resaltar que las MVS presentan un buen comportamiento incluso cuando no está disponible una suficiente cantidad de datos de entrada. En el mismo sentido las MVS tienen mayor tolerancia al ruido que las RNA; es decir, la carencia de algunos datos o la baja calidad de los mismos no afecta de manera drástica su comportamiento [10, 33].

No obstante, los trabajos consultados presentan algunas falencias que permiten establecer una serie de brechas en torno al tema de investigación. A continuación, se abordan los principales vacíos y falencias de los trabajos, de tal forma que se pueda establecer un punto de referencia en este estudio para la generación de nuevo conocimiento alrededor del campo de la predicción de la calidad del agua.

Principalmente dentro de los trabajos que se enfocan en la predicción de la calidad del agua, la precisión de las predicciones depende del conjunto de datos de entrenamiento; con una menor cantidad de datos, la precisión tiende a disminuir. Por otro lado, la tasa de error tiende a aumentar si las predicciones se realizan sobre intervalos de tiempo prolongados (por ejemplo, si los datos de prueba son registros históricos diarios en un mes, el intervalo de predicción abarcaría algunos días posteriores sin que la precisión disminuya).

En este mismo orden de ideas, con respecto a los trabajos que presentan un enfoque hacia la calibración de técnicas de predicción de calidad del agua, su capacidad predictiva está ligada a las interrelaciones que presenten algunas variables de calidad del agua en un uso determinado. Del mismo modo el objetivo de estos trabajos es incrementar la precisión de las predicciones; sin embargo, su potencial no es aprovechado totalmente teniendo en cuenta la capacidad adaptativa de las técnicas propuestas, permitiendo beneficios como la posibilidad de aplicar el mecanismo de predicción a diferentes usos del agua.

## 5. CONCLUSIONES

Este estudio tiene como principal objetivo desarrollar un mapeo sistemático de la literatura relacionada con la predicción de la calidad del agua haciendo uso de técnicas de inteligencia computacional. Es importante resaltar que esta temática es bastante

amplia y el análisis documental se aborda desde el planteamiento de 3 preguntas de investigación que permiten explorar de forma sistemática el tema de investigación.

La pregunta que hace referencia a los temas de mayor interés para la comunidad científica en torno a la predicción de la calidad del agua fue resuelta mediante la búsqueda en títulos y resúmenes, lo cual no fue suficiente para responderla, y de esta forma fue requerido un análisis mucho más profundo de los trabajos. A partir de lo anterior se estableció que el tema de mayor importancia es la construcción de modelos de predicción híbridos, con un 46.1 % de los trabajos consultados enfocados en este tópico, principalmente para cubrir las desventajas de algunas técnicas y aportar mejoras con la incorporación de otras técnicas como refuerzo.

Las técnicas de IC que han sido utilizadas como parte del componente adaptativo de un mecanismo de predicción han sido el algoritmo de proyección y el algoritmo de mínimos cuadrados; sin embargo, estas técnicas no son utilizadas en la adaptación de dicho mecanismo a diferentes usos del agua. No obstante, existen técnicas que, a pesar de no haber sido exploradas a fondo dentro de un componente adaptativo, poseen características importantes como la capacidad de adaptarse a entornos cambiantes; tal es el caso de PSO y la computación evolutiva en especial los algoritmos genéticos.

Para permitir que un mecanismo de predicción se pueda utilizar en diferentes usos del agua, en este estudio se presentó de manera general una propuesta de investigación orientada hacia el uso de técnicas de IC. Como trabajo futuro se plantea la implementación y evaluación de un mecanismo de predicción adaptativo basado en esta propuesta con el fin de determinar su eficacia en diferentes usos del agua.

## AGRADECIMIENTOS

A la Red Interinstitucional de Cambio Climático y Seguridad Alimentaria (RICCLISA), a los Grupos de Estudios Ambientales y de Ingeniería Telemática de la Universidad del Cauca.

## REFERENCIAS

- [1] P. J. and J. S. Claudia Pahl-Wostl y C., Paul Jeffrey, and Jan Sendzimir. Pahl-Wostl, Adaptive and integrated management of water resources. publisherNameCambridge University Press, 2011.
- [2] Consejo Económico y Social de Castilla-Mancha, “Estudio La gestión del Agua en Castilla-La Mancha”. 2004.
- [3] Comunidad Autónoma de Extremadura, Agentes Forestales de Extremadura. Legislacion Básica Ebook. MAD-Eduforma, 2003.

- [4] IDEAM, “Calidad del Agua Superficial en Colombia”, en Estudio Nacional del Agua, 2010, pp. 231-277.
- [5] N. Rescher, *Predicting the Future: An Introduction to the Theory of Forecasting*. SUNY Press, 1998.
- [6] E. Kumar, *Artificial Intelligence*. I.K. International Publishing House Pvt. Limited, 2008.
- [7] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. Wiley, 2007.
- [8] B. Kitchenham y S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering”, Keele University and Durham University Joint Report, UK, EBSE 2007-001, 2007.
- [9] K. Petersen, R. Feldt, S. Mujtaba, y M. Mattsson, “Systematic Mapping Studies in Software Engineering”, en *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, Swinton, UK, UK, 2008, pp. 68–77.
- [10] J. M. Pérez, *Inteligencia computacional inspirada en la vida*. Servicio de Publicaciones e Intercambio Científico de la Universidad de Málaga, 2010.
- [11] J. Grbović y S. Džeroski, “Knowledge discovery in a water quality database”, *Proc 1st Intl Conf Knowl. Discov. Data Min. KDD95 AAAI Press Menlo Park CA* 1995.
- [12] S. Džeroski, D. Demšar, y J. Grbović, “Predicting Chemical Parameters of River Water Quality from Bioindicator Data”, *Appl. Intell.*, vol. 13, n.º 1, pp. 7–17, jul. 2000.
- [13] L. Breiman, *Classification and regression trees*. Chapman & Hall, 1984.
- [14] H. Blockeel, S. Džeroski, y J. Grbovic, *Simultaneous prediction of multiple chemical parameters of river water quality with TILDE*. 1999.
- [15] T. G. Dietterich, “Ensemble Methods in Machine Learning”, en *Multiple Classifier Systems*, Springer Berlin Heidelberg, 2000, pp. 1-15.
- [16] I. Partalas, G. Tsoumakas, E. V. Hatzikos, y I. Vlahavas, “Greedy regression ensemble selection: Theory and an application to water quality prediction”, *Inf. Sci.*, vol. 178, n.º 20, pp. 3867-3879, oct. 2008.
- [17] G. Tan, J. Yan, C. Gao, y S. Yang, “Prediction of water quality time series data based on least squares support vector machine”, *Procedia Eng.*, vol. 31, pp. 1194-1199, 2012.
- [18] K. Gurney, *An Introduction to Neural Networks*. Taylor & Francis, 2003.
- [19] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [20] E. Cox, *The Fuzzy Systems Handbook: A Practitioner’s Guide to Building, Using, and Maintaining Fuzzy Systems*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.

- [21] C. E. Romero y J. Shan, "Development of an Artificial Neural Network-based Software for Prediction of Power Plant Canal Water Discharge Temperature", *Expert Syst Appl*, vol. 29, n.º 4, pp. 831–838, nov. 2005.
- [22] J. Gutiérrez, W. Riss, y R. Ospina, "Bioindicación de la calidad del agua con macroinvertebrados acuáticos en la sabana de Bogotá, utilizando redes neuronales artificiales", *Caldasia*, vol. 26, n.º 1, pp. 151-160, 2004.
- [23] I. García, J. G. Rodríguez, F. López, y Y. M. Tenorio, "Transporte de contaminantes en aguas subterráneas mediante redes neuronales artificiales", *Inf. Tecnológica*, vol. 21, n.º 5, pp. 79-86, 2010.
- [24] A. I. Saint-Gerons y J. M. Adrados, "Desarrollo de una red neuronal para estimar el oxígeno disuelto en el agua a partir de instrumentación de EDAR".
- [25] A. Ogata y R. B. Banks, "A solution of the differential equation of longitudinal dispersion in porous media", 1961.
- [26] G. J. Pelletier, S. C. Chapra, y H. Tao, "QUAL2Kw – A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration", *Environ. Model. Softw.*, vol. 21, n.º 3, pp. 419-425, mar. 2006.
- [27] S. Liu, D. Butler, R. Brazier, L. Heathwaite, y S.-T. Khu, "Using genetic algorithms to calibrate a water quality model", *Sci. Total Environ.*, vol. 374, n.º 2-3, pp. 260-272, mar. 2007.
- [28] Y. Huang y L. Liu, "Multiobjective Water Quality Model Calibration Using a Hybrid Genetic Algorithm and Neural Network-Based Approach", *J. Environ. Eng.*, vol. 136, n.º 10, pp. 1020-1031, 2010.
- [29] K. Chau, "A Split-Step PSO Algorithm in Prediction of Water Quality Pollution", en *Advances in Neural Networks – ISNN 2005*, J. Wang, X.-F. Liao, y Z. Yi, Eds. Springer Berlin Heidelberg, 2005, pp. 1034-1039.
- [30] A. M. Baltar y D. G. Fontane, "A generalized multiobjective particle swarm optimization solver for spreadsheet models: application to water quality", *Proc. Twenty Sixth Annu. Am. Geophys. Union Hydrol. Days*, pp. 20-22, 2006.
- [31] A. Afshar, H. Kazemi, y M. Saadatpour, "Particle Swarm Optimization for Automatic Calibration of Large Scale Water Quality Model (CE-QUAL-W2): Application to Karkheh Reservoir, Iran", *Water Resour. Manag.*, vol. 25, n.º 10, pp. 2613-2632, ago. 2011.
- [32] J. Zhangzan, X. Gang, C. JiuJun, y G. Fei, "Anomaly detection of water quality based on visual perception and V-detector", *Inf. Control*, vol. 1, p. 026, 2011.
- [33] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, y Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction", *Math. Comput. Model.*, vol. 58, n.º 3-4, pp. 458-465, ago. 2013.



- [34] S. Liu, L. Xu, D. Li, Q. Li, Y. Jiang, H. Tai, y L. Zeng, “Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization”, *Comput. Electron. Agric.*, vol. 95, pp. 82-91, jul. 2013.
- [35] D. Ömer Faruk, “A hybrid neural network and ARIMA model for water quality time series prediction”, *Eng. Appl. Artif. Intell.*, vol. 23, n.º 4, pp. 586-594, jun. 2010.
- [36] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson, y J. A. Moncada-Herrera, “A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile”, *Atmos. Environ.*, vol. 42, n.º 35, pp. 8331-8340, nov. 2008.

