

Indicador de calidad del agua para acuacultura utilizando una memoria asociativa modificada

Raúl Jiménez Cruz¹, Midory Esmeralda Viguera Velazquez²,
Miguel González Mendoza¹

¹Instituto Tecnológico de Estudios Superiores,
Estado de México,
México

²Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México

{rjimenezc, mgonza}@tec.mx, midory.viguera@gmail.com

Resumen. El *Chirostoma estor estor* (conocido coloquialmente como charal blanco) es una especie muy importante que habita en el Lago de Pátzcuaro, ubicado en Michoacán, México. Este trabajo propone el uso de una memoria asociativa modificada para evaluar la calidad del agua en estanques de cultivo intensivo con sistemas de fotoperiodo. El fotoperiodo mejora las tasas de reproducción debido a que se controla la puesta de huevos de los peces. Se estudiaron la temperatura, el pH, el oxígeno disuelto, el amoníaco no ionizado, el amoníaco total, la alcalinidad total, los sólidos suspendidos totales, los fluoruros totales, la dureza total y los coliformes fecales. Una memoria hetero asociativa tiene la funcionalidad de recordar etiquetas que pertenecen a un patrón en el conjunto de datos. En otras palabras, realiza una clasificación de patrones. Los resultados experimentales muestran un buen rendimiento del modelo propuesto en comparación con los reportados en la literatura, evitando errores de medición y proporcionando una nueva herramienta para la investigación en acuacultura.

Palabras clave: Memoria asociativa, clasificador, reconocimiento de patrones.

Water Quality Indicator for Aquaculture Using a Modified Associative Memory

Abstract. The *Chirostoma estor estor* (commonly known as whitefish) is a highly important species inhabiting Pátzcuaro lake, located in Michoacán, Mexico. This paper proposes the use of a modified associative memory to evaluate water quality in intensive cultivation ponds with photoperiod systems. The photoperiod enhances reproduction rates by controlling the fish's egg laying. Temperature, pH, dissolved oxygen, un-ionized ammonia, total ammonia, total alkalinity, total suspended solids, total fluorides, total hardness, and fecal coliforms were studied. A heteroassociative memory has the functionality of remembering labels that belong to a pattern in the dataset. In other words, it performs pattern classification. Experimental results show good performance of the proposed

model compared to those reported in the literature, avoiding measurement errors and providing a new tool for research in aquaculture.

Keywords: Associative memory, classifier, pattern recognition.

1. Introducción

El Lago de Pátzcuaro, ubicado en Michoacán, México, se encuentra a 63 km al oeste de la ciudad de Morelia. Este es uno de los atractivos turísticos más importantes de la entidad, formando parte de un conjunto arqueológico, histórico, recreativo y cultural. El turismo representa la mayor fuente de ingresos para los habitantes. En este lago se encuentran 12 especies de peces diferentes: ocho son endémicas y cuatro son introducidas. Las dos especies más importantes en este lugar son el *Chirostoma estor* y *Algansea lacustris*. Ambas especies generan altos ingresos para la población debido a su alto volumen de pesca y demanda en el mercado regional.

El fotoperiodo en la acuicultura se utiliza para modificar el ciclo reproductivo, mejorar la sincronización de la maduración sexual e inducir la puesta, solucionando así problemas de baja población. En la literatura, se han realizado varios trabajos en los últimos años que han contribuido a comprender los efectos de la luz en la pesca, lo que ha generado avances tecnológicos considerables en la acuicultura.

Por ejemplo, en [2] se propuso un sistema informatizado de monitoreo y control ambiental donde un sistema supervisa y controla el fotoperiodo y la temperatura, proporcionando un registro continuo de estos parámetros en formato digital. Además, en [3] se propuso el uso de un sistema de apoyo a la toma de decisiones para la investigación y gestión en acuicultura. Sin embargo, estos trabajos carecen de metodologías de evaluación precisa y desarrollo tecnológico para un estudio más profundo de la especie *Chirostoma estor* y para un proceso de fotoperiodo exitoso. Por lo que se propone aplicar un clasificador inteligente para ayudar a la clasificación del agua para esta especie de pez.

2. Requisitos ambientales

Para poder obtener un dataset que pueda ser utilizado es necesario que se realicen análisis semanales de nitratos, que incluyen amoníaco no ionizado y amoníaco total. La demanda bioquímica de oxígeno, la demanda química de oxígeno, la temperatura y el pH se monitorearon diariamente debido a que representan los parámetros más críticos según expertos en la acuicultura de peces blancos (véase Tabla 1). Es importante destacar que otros parámetros no fueron considerados debido a que los sistemas cultivados están controlados ecológicamente y rara vez presentan niveles problemáticos. Por lo tanto, un conjunto reducido de los parámetros más importantes puede ser modelado con éxito para proporcionar un indicador preciso y viable que pueda ser medido en un sistema computacional. La importancia de los parámetros involucrados en la cría de peces blancos se detalla en la siguiente lista:

1. La Demanda Bioquímica de Oxígeno (**DBO**): Muestra la calidad del agua desde el punto de vista orgánico, ya que mide la cantidad de oxígeno consumido por

Tabla 1. Parámetros de calidad del agua utilizados para evaluar el hábitat del pez blanco.

Parámetros Críticos	Por requerimiento
Demanda Bioquímica de Oxígeno (DBO)	Amoníaco Total (TAN)
Demanda Química de Oxígeno (DQO)	Dureza Total (TOT_HAR)
Temperatura (TEMP)	Alcalinidad Total (TOT_ALC)
pH	Sólidos Suspendidos Totales (TSS)

microorganismos en la oxidación química de la materia orgánica presente en la muestra de agua [4].

2. La Demanda Química de Oxígeno (**DQO**): Es una medida de la cantidad de oxígeno disuelto consumido, bajo condiciones preestablecidas, por la oxidación química de la materia orgánica biodegradable presente en el agua [5].
3. Los parámetros de DBO y DQO: Proporcionan información diferente. Normalmente, los valores de DQO son más altos que los valores de DBO, porque el oxidante químico es capaz de reaccionar con sustancias que son difíciles de biodegradar para los microorganismos [6].
4. La temperatura (**TEMP**): Esta propiedad termodinámica afecta considerablemente las características físicas, químicas y biológicas de los cuerpos de agua. Controla el desove, la incubación (etapa más crítica en el desarrollo del pez blanco); regula actividad o suprime las tasas de crecimiento y puede ser letal en condiciones extremas. La especie resiste un amplio rango de temperatura de 15 °C a 24 °C. Las tasas de crecimiento se verán afectadas en temperaturas por debajo de los 15 °C. La temperatura óptima para el desarrollo de la especie se considera que está en 22°C [7].
5. **pH**: Este parámetro está determinado por el número de iones de hidrógeno libres (H⁺) y nos ayuda a indicar los niveles de acidez o alcalinidad en el estanque cultivado. El valor de pH ideal para el pez blanco es en un agua completamente neutra, lo que significa un valor exacto de 7. Sin embargo, el pH puede oscilar entre 7 y 8.5. Se puede observar que la especie no acepta un amplio rango en el pH, ya que el pez blanco estaría sujeto a estrés, lo que provoca altas tasas de mortalidad y problemas de amoníaco no ionizado [8].

2.1. Niveles de parámetros

La estabilidad de los parámetros de calidad del agua es importante en cualquier sistema de acuicultura. Las variaciones en sus niveles ocurren con frecuencia a lo largo del día y medirlos permite entender la dinámica del ambiente. Los parámetros diarios son extremadamente importantes porque pueden desestabilizar considerablemente el hábitat cultivado. La Tabla 2 muestra los niveles permitidos donde el hábitat del pez blanco puede considerarse óptimo para un buen proceso de cultivo.

Tabla 2. C_l y C_u corresponden a los límites inferior y superior definidos respectivamente, mientras que m es el nivel óptimo y deseable para la especie.

Parámetros	Unidades	Rango	Límites Permitidos		
			C_l	C_u	m
TEMP	Celsius	15 – 24	15	24	19.5
DO	mg/l	4.5 – 9	4.5	---	6
pH	Unidad pH	7 – 8.5	7	8.5	7.75
TAN	mg/l	0.1 – 1.0	0.1	1.0	0.55
TOT_HAR	mg/l	80-120	80	120	100
TOT_ALC	mg/l	90-110	90	110	100
TSS	mg/l	50-100	50	100	75
FEC_COL	mg/l	50-200	50	200	189.5
TOT_FLUO	mg/l	0.8-1.1	0.5	1.5	1.0

3. Modelos de aprendizaje automático

3.1. Linear Associator

El modelo de clasificación consiste en el uso del Linear Associator implementándole una técnica de ortonormalización de la matriz de entrenamiento para reducir el inconveniente principal del modelo original.

El Linear Associator, es una memoria asociativa, cuyo desarrollo se atribuye a dos científicos Kohonen en Finlandia [9] y Anderson en Estados Unidos [10] en el año de 1972. Pero este modelo tiene un gran problema, requiere vectores ortonormales para funcionar correctamente, lo cual no existe en los datasets de uso habitual.

3.2. Linear Associator con SVD

En el año 2021 se creó una nueva variante de este modelo el clasificador LA-SVD para más información revisar la referencia [11], A grandes rasgos esta nueva variante del modelo original propone dos ideas muy importantes, la primera es que al realizar la predicción de los patrones el modelo suele regresar valores con punto decimal lo cual está mal, pero si las etiquetas de clase son catalogadas en números ya sea como 1, 2, 3... y así sucesivamente.

Ahora bien, el modelo rara vez dará con la etiqueta correcta debido a que es muy importante que la matriz entrenada sea ortonormal por lo que tendera a clasificar mal los patrones. Por lo que la primera propuesta consiste en aplicar un redondeo en el momento que el clasificador obtenga la pseudo etiqueta de clase.

Ya que los valores que se obtienen al final son valores aproximados numéricamente a la etiqueta de clase. La segunda propuesta que se implementó en este modelo es el uso de la Descomposición en Valores Singulares o SVD en inglés. Esta técnica puede ayudar a obtener en la matriz de entrenamiento valores aproximadamente cercanos a los ortonormales de la matriz. Por lo que mejoraría la clasificación de los patrones al ser implementado.

El algoritmo con la SVD consiste en lo siguiente: Una vez elegido el método de validación comenzaremos tomando el conjunto de entrenamiento para entrenar la memoria asociativa. Se descompondrá el conjunto de entrenamiento usando la descomposición en valores singulares reducida o SVD economy los cual nos otorgará tres matrices, U , Σ y V . Esto debido a que la matriz U es una matriz que contiene valores aproximados a los ortonormales en una matriz.

El siguiente paso consiste en aplicar la fase de entrenamiento del linear Associator es decir se va a realizar un producto externo de la siguiente manera:

$$M = \sum_{\mu=1}^p y^{\mu} \cdot (x^{\mu})^T, \quad (1)$$

donde:

- M es la matriz entrenada.
- y^{μ} son el vector de etiquetas de clase es decir la codificación de los patrones de salida.
- x^{μ} son los patrones que se encuentran en el conjunto de entrenamiento, pero transpuestos.

Una vez obtenida la matriz M se finaliza la fase de entrenamiento del modelo. El siguiente paso es la fase de recuperación la cual consiste en tomar el conjunto de entrenamiento y transformarlo en un conjunto ortonormal de la siguiente manera:

$$x_{\text{test_new}} = x_{\text{test}} \cdot (V^+)^T \cdot S^+, \quad (2)$$

donde:

- $x_{\text{test_new}}$ es el conjunto de prueba transformado.
- x_{test} es el conjunto de entrenamiento original.
- $(V^+)^T$ es una de las matrices obtenidas en la SVD antes de entrenar se le aplico la pseudo inversa de Moore Penrose además de la transpuesta.
- S^+ es una de las matrices ob tenidas en la SVD antes de entrenar se le aplico la pseudo inversa de Moore Penrose.

Una vez transformado el conjunto de prueba nótese que jamás interactúa con el conjunto de entrenamiento se aplica la fase de recuperación del linear Associator la cual consiste en lo siguiente:

$$\text{round}(M \cdot x_{\text{test_new}}). \quad (3)$$

Al aplicar el redondeo se obtendrá la etiqueta de clase predicha.

4. Descripción del dataset

El conjunto de datos que se analizó en el presente trabajo fue proporcionado por el repositorio de datos de CONAGUA [12], que desafortunadamente contenía muchos valores faltantes pero se le aplico un preprocesamiento de imputación para que este fuese más fácil de implementar. La base de datos de CONAGUA tiene las siguientes

Tabla 3. Accuracy.

SMO	Naive Bayes	Perceptrón Multicapa	Bosques Aleatorios	LA SVD	LA SVD Polar
94.96	92.44	95.68	95.68	96.55	94.96

Tabla 4. Balance accuracy.

SMO	Naive Bayes	Perceptrón Multicapa	Bosques Aleatorios	LA SVD	LA SVD Polar
0.9520	0.9315	0.9610	0.9590	0.9619	0.9478

características: Contiene 1390 patrones y diez características, todas numéricas, que son las siguientes: TEMP, pH, DBO, DQO, TOT_ALC, TAN, TSS, TOT_FLUO, TOT_HAR y FEC_COL.

Las clases pertenecientes a esta base de datos corresponden a un semáforo que CONAGUA estableció por medio de la RENAMECA, de la siguiente manera:

- **Verde:** El agua del sitio se encuentra dentro de los rangos de calidad excelente, buena calidad y aceptable, en todos los indicadores. Esta clase tiene 564 patrones.
- **Rojo:** El agua del sitio está contaminada con Demanda Bioquímica de Oxígeno (DBO) o con Demanda Química de Oxígeno (DQO) y además el agua del sitio está contaminada con Sólidos Suspendidos Totales (TSS). Esta clase tiene 826 patrones.

Como se puede observar el dataset se encuentra balanceado ya que tiene un IR (Imbalance Ratio) [13] de 1.46, siendo la clase Verde la clase minoritaria y la clase Rojo la mayoritaria. Cabe aclarar que este dataset extraído es para todo el país, con alrededor de 5000 datos y en este caso se aplicó un filtro en los rasgos que no usaban las características necesarias para un cultivo de peces *Chirostoma*, así como un filtro en los patrones ya que el caso de estudio fue para el lago de Pátzcuaro donde se realiza acuicultura del *Chirostoma* reduciéndolo a 1390 patrones que representan zonas de acuicultura, además los datos son del año 2016 pero aplicando el mismo filtro se puede implementar con datos más actuales. La razón de usar un dataset del 2016 es porque se encuentra más completo que los actuales.

5. Resultados experimentales y discusión

Como el conjunto de datos maneja datos completamente numéricos, se utilizó un método de imputación, que consiste en obtener la media de cada característica para cada clase del conjunto de datos y así eliminar la problemática de los valores perdidos. Posteriormente, se programó el algoritmo LA-SVD y a su vez se le realizó una pequeña variante donde se implementa un cambio de coordenadas cartesianas a esféricas para n dimensiones, esto con el fin de cambiar el espacio muestral bajo la hipótesis de que cada patrón puede alejarse o acercarse de la clase a la que pertenece, todo esto se realizó utilizando las siguientes ecuaciones:

$$x_n = r \cos \alpha_{n-1} \prod_{n=1}^{n-2} \sin \alpha_n, \quad (4)$$

donde x_n es el enésimo patrón, r es el radio de una esfera n dimensional y α_n es el enésimo ángulo del plano esférico en n dimensiones. Para entender la siguiente ecuación es necesario segmentarla en partes como sigue:

$$r = \sqrt{\sum_{n=1}^n}. \quad (5)$$

La ecuación 2 se refiere al cálculo del radio n -dimensional de la esfera. Las siguientes ecuaciones son útiles para determinar el ángulo alfa (α) de los $n - 1$ componentes:

$$\alpha_1 = \cos^{-1}\left(\frac{x_1}{\sqrt{x_n^2 + x_{n-1}^2 + \dots + x_1^2}}\right), \quad (6)$$

$$\alpha_3 = \cos^{-1}\left(\frac{x_2}{\sqrt{x_n^2 + x_{n-1}^2 + \dots + x_2^2}}\right), \quad (7)$$

$$\alpha_{n-2} = \cos^{-1}\left(\frac{x_{n-2}}{\sqrt{x_n^2 + x_{n-1}^2 + x_{n-2}^2}}\right), \quad (8)$$

$$\alpha_{n-1} = \cos^{-1}\left(\frac{x_{n-1}}{\sqrt{x_n^2 + x_{n-1}^2}}\right). \quad (9)$$

Una vez realizado este cambio, se le aplicó el método de validación leave-one-out para este clasificador con la razón de que los resultados son deterministas al implementarlo por lo que no hay variabilidad en los mismos. Finalmente, se implementó el clasificador LA SVD con coordenadas esféricas, el cual mostró un porcentaje de accuracy del 81.22%, pero debido a que el data set se encuentra desbalanceado se procedió también a usar la medida de desempeño balance accuracy la cual dio un resultado de 96.55%.

El siguiente paso fue comparar el resultado del clasificador propuesto con algunos clasificadores del estado del arte utilizando la plataforma WEKA. Los resultados de precisión se muestran en la Tabla 3. Debido a que el conjunto de datos está desequilibrado, es necesario analizar la matriz de confusión y verificar que no haya sesgo en las clases con más patrones:

- **SMO**: A pesar de tener un buen accuracy, clasificó correctamente 564 patrones en la primera clase es decir todos, pero solo clasificó 771 de 826 en la segunda clase.
- **Naive Bayes**: Para la primera clase solo se clasificaron correctamente 546 patrones de 564, para la segunda clase se clasificaron correctamente 739 patrones de los 826.
- **Perceptrón multicapa**: La primera clase obtuvo 555 patrones bien clasificados de 564. La segunda clase obtuvo 7775 patrones bien clasificados.

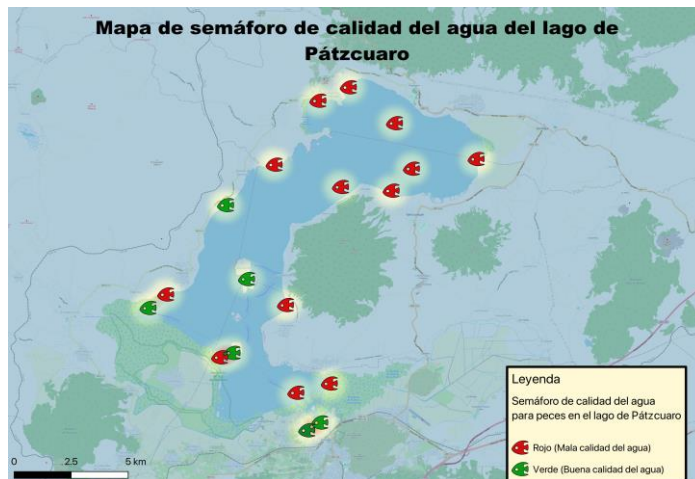


Fig. 1. Mapa que representa en un semáforo la calidad del agua en el Lago de Pátzcuaro.

- **Bosques Aleatorios:** Obtuvo 547 aciertos en la primera clase y 783 en la segunda clase.
- **LA SVD:** El algoritmo propuesto obtuvo 532 patrones bien clasificados en la primera clase, la segunda clase obtuvo 810 patrones bien clasificados.
- **LA SVD Polar:** La variante del algoritmo obtuvo 529 patrones bien clasificados para la primera clase, mientras que en la segunda clase se obtuvieron 791 patrones bien clasificados.

Debido a que se notaron algunos sesgos en la clasificación se procedió a realizar el cálculo del Balance Accuracy para verificar si existe dicho sesgo esto puede verse en la Tabla 3. Accuracy.

SMO	Naive Bayes	Perceptrón Multicapa	Bosques Aleatorios	LA SVD	LA SVD Polar
94.96	92.44	95.68	95.68	96.55	94.96

Tabla 4. Finalmente, como puede apreciarse en la tabla anterior el modelo más estable es el LA SVD en cuanto a una clasificación más balanceada de los patrones de este dataset además de ser el mejor en accuracy.

Posteriormente la clasificación puede ser representada en un mapa debido a que los datos de Conagua tienen asociados cada punto un par de coordenadas por lo que una vez clasificado se puede mostrar en un sistema de información geoespacial (véase Fig. 1). La figura anterior es un ejemplo de cómo expandir este proyecto a más que clasificar donde podemos ver lugares donde el agua tiene una mala calidad y que se puedan atender de forma inmediata en cuanto se detecte un estanque con calidad mala.

6. Conclusiones

Este trabajo se desarrolló con el fin de crear una herramienta especializada para monitorear los estanques de *Chirotostoma estor estor*, la cual es una especie endémica

mexicana en peligro de extinción. Se debe destacar que su principal ventaja radica en el análisis de los parámetros específicos cruciales para su reproducción.

El uso de un clasificador se revela como un elemento fundamental para evaluar la calidad del agua necesaria para asegurar el bienestar de este ejemplar en su hábitat, es decir, se puede tomar como un recurso esencial para la gestión y mejora del ecosistema acuático en caso de ser necesario. Como trabajo futuro, se planea fortalecer el modelo de evaluación, incorporando más parámetros para una evaluación exhaustiva del fotoperíodo, lo que potencialmente mejoraría la reproducción de la especie. Además, se considera la implementación de nuevas técnicas de preprocesamiento que minimicen el sesgo en la clasificación, asegurando una evaluación más precisa de la calidad del agua.

Referencias

1. Martínez Palacios, C.A., Toledo Cuevas, M.E.C., Blanco Michoacán. *Rev. Digit. Univ.*, vol. 10, pp. 42–44 (2005)
2. Fabbrocini, A., Di-Stasio, M., D'Adamo, R.: Computerized Sperm Motility Analysis in Toxicity Bioassays: A New Approach to Pore Water Quality Assessment. *Ecotoxicology and Environmental Safety*, vol. 73, no. 7, pp. 1588–1595 (2010). DOI: 10.1016/j.ecoenv.2010.05.003.
3. Mølmann, J.A., Steindal, A.L., Bengtsson, G.B., Seljåsen, R., Lea, P., Skaret, J., Johansen, T.J.: Effects of Temperature and Photoperiod on Sensory Quality and Contents of Glucosinolates, Flavonols and Vitamin C in Broccoli Florets. *Food Chemistry*, vol. 172, pp. 47–55 (2015) DOI: 10.1016/j.foodchem.2014.09.015.
4. Tchinda, D., Henkanatte-Gedera, S., Abeysiriwardana-Arachchige, I., Delanka-Pedige, H., Munasinghe-Arachchige, S., Zhang, Y., Nirmalakhandan, N.: Single-step Treatment of Primary Effluent by *Galdieria Sulphuraria*: Removal of Biochemical Oxygen Demand, Nutrients, and Pathogens. *Algal Research*, vol. 42, pp. 101578 (2019). DOI: 10.1016/j.algal.2019.101578.
5. Kabir, H., Zhu, H., Lopez, R., Nicholas, N.W., McIlroy, D.N., Echeverria, E., May, J., Cheng, I.F.: Electrochemical Determination of Chemical Oxygen Demand on Functionalized Pseudo-Graphite Electrode. *Journal of Electroanalytical Chemistry*, vol. 851, pp. 113448 (2019). DOI: 10.1016/j.jelechem.2019.113448.
6. Nguyen, L.A.T., Ward, A.J., Lewis, D.: Utilisation of Turbidity as an Indicator for Biochemical and Chemical Oxygen Demand. *Journal of Water Process Engineering*, vol. 4, pp. 137–142 (2014). DOI: 10.1016/j.jwpe.2014.09.009.
7. Zheng, G., Bao, A., Li, J., Zhang, G., Xie, H., Guo, H., Jiang, L., Chen, T., Chang, C., Chen, W.: Sustained Growth of High Mountain Lakes in the Headwaters of the Syr Darya River, Central Asia. *Global and Planetary Change*, vol. 176, pp. 84–99 (2019). DOI: 10.1016/j.gloplacha.2019.03.004.
8. Qin, Y., Alam, A. U., Pan, S., Howlader, M.M., Ghosh, R., Hu, N., Jin, H., Dong, S., Chen, C., Deen, M.J.: Integrated Water Quality Monitoring System with pH, Free Chlorine, and Temperature Sensors. *Sensors and Actuators B: Chemical*, vol. 255, pp. 781–790 (2018). DOI: 10.1016/j.snb.2017.07.188.
9. Kohonen, T.: Correlation Matrix Memories. *IEEE Transactions on Computers*, vol. C–21, no. 4, pp. 353–359 (1972). DOI: 10.1109/tc.1972.5008975.
10. Anderson, J.A.: A Simple Neural Network Generating an Interactive Memory. *Mathematical Biosciences*, vol. 14, no. 3–4, pp. 197–220 (1972). DOI: 10.1016/0025-5564(72)90075-2.
11. Jiménez-Cruz, R., Velázquez-Rodríguez, J., López-Yáñez, I., Villuendas-Rey, Y., Yáñez-Márquez, C.: Supervised Classification of Diseases based on an Improved Associative Algorithm. *Mathematics*, vol. 9, no. 13, pp. 1458 (2021). DOI: 10.3390/math9131458.

Raúl Jiménez Cruz, Midory Esmeralda Viguera Velazquez, et al.

12. CONAGUA: Indicadores de la calidad del agua superficial y subterránea. Red nacional de medición de la calidad del agua. <http://files.conagua.gob.mx/Ica20/Contenido/Documentos/PresentaciondeIndicadoresdeCalidaddeAgua.pdf> (2022)
13. KEEL: KEEL-dataset, data set repository. <http://sci2s.ugr.es/keel/imbanced.php> (2005)