

Using Machine Learning and Linear Regression to Forecast the Water Quality in Al-Sain Lake

Dr. Raed JAFAR*

(Received 3 / 4 / 2023. Accepted 23 / 5 / 2023)

□ ABSTRACT □

Effective management of the quantity and quality of water requires accurate assessment and determination of the pollution levels of surface and groundwater. The goal of this study is to assess the effectiveness of multiple linear regression (MLR) and 19 machine learning (ML) models, which utilize various algorithms such as regression, boosting, and decision tree. Among of these models are linear regression (Lr), least angle regression (Lar), Bayesian ridge chain (Br), ridge regression (Ridge), k-nearest neighbors regression (K-nn), extra tree regression (Et), extreme gradient boosting (XGBoost), etc. By employing these models, the study aims to accurately predict the surface water quality of Al-Sain lake in Latakia city.

To define the water quality index (WQI), data from the drinking water lake intake for the years 2021-2022 were analyzed. The effectiveness of the multiple linear regression (MLR) and machine learning (ML) models were assessed using statistical tools such as the coefficient of determination (R^2) and the root mean square error (RMSE) to gauge their accuracy.

The results indicated that the multiple linear regression model (MLR) and 3 of the machine learning (ML) models, including linear regression (Lr), least angle regression (Lar), and Bayesian ridge chain (Br), performed extremely in predicting the (WQI) index with high accuracy ($R^2 = 0.99$, RMSE = 0.15) for the (MLR) model, and high accuracy ($R^2 = 1.0$, RMSE \approx 0.0) for the three aforementioned machine learning (ML) models. The results support the use of multiple linear regression models and machine learning models in predicting the water quality index (WQI) with very high accuracy, which will contribute to improving of water quality management.

Keywords: Water quality, surface water, multiple linear regression (MLR), machine learning (ML), regression-based algorithms, water quality index (WQI), Al-Sain lake.

Copyright



:Tishreen University journal-Syria, The authors retain the copyright under a CC BY-NC-SA 04

* Associate Professor –Department of Environmental Engineering - Faculty of Civil Engineering - Tishreen University – Lattakia- Syria. r-jafar@tishreen.edu.sy

استخدام التعلم الآلي والانحدار الخطي للتنبؤ بجودة المياه في بحيرة السن

د. رائد جعفر*

(تاريخ الإيداع 3 / 4 / 2023. قُبِلَ للنشر في 23 / 5 / 2023)

□ ملخص □

تتطلب الإدارة الفعالة لكمية ونوعية المياه تقييم وتحديد مستوى تلوث المياه السطحية والجوفية الموجودة بشكل دقيق. تهدف هذه الدراسة إلى تقييم أداء نموذج الانحدار الخطي متعدد المتغيرات التقليدي (MLR) وكذلك تسع عشرة نموذجاً للتعلم الآلي (ML)، بما في ذلك الخوارزميات القائمة على الانحدار وشجرة القرار والتعزيز (الانحدار الخطي (Lr)، انحدار الزاوية الأدنى (Lar)، سلسلة الانحدار البايزية (Br)، انحدار ريدج (Ridge)، انحدار الجيران الأقرب k (Knn)، الانحدار الإضافي للأشجار (Et)، الانحدار العشوائي للغابات (Rf)، الانحدار المعزز المتدرج (Gbr)، تعزيز التدرج الشديد (Xgboost)، ... إلخ، في تقدير جودة المياه السطحية لبحيرة السن في محافظة اللاذقية. تم استخدام بيانات جودة المياه لمأخذ مياه شرب مدينة اللاذقية في البحيرة للعامين (2021-2022) من أجل حساب مؤشر جودة المياه (WQI). تم تقييم أداء التنبؤ لنماذج الانحدار الخطي (MLR) وتعلم الآلة (ML) باستخدام عدة طرق احصائية لتقدير كفاءتهم مثل (معامل التحديد R^2 وجذر متوسط الخطأ التربيعي RMSE). أشارت النتائج إلى أن نموذج الانحدار الخطي متعدد المتغيرات (MLR) وثلاثة من نماذج تعلم الآلة (ML) وهي الانحدار الخطي (Lr)، الانحدار الأدنى للزاوية (Lar) و سلسلة التلال البايزية (Br) تتمتع بأداء عالٍ جداً في التنبؤ بمؤشر (WQI) وبدقة ($R^2 = 0.999$, $RMSE = 0.149$) بالنسبة لنموذج (MLR)، ودقة ($RMSE \sim 0.0$, $R^2 = 1$) بالنسبة لنماذج تعلم الآلة الثلاثة المذكوره أعلاه. تدعم النتائج استخدام نماذج الانحدار الخطي متعدد المتغيرات ونماذج تعلم الآلة في التنبؤ بمؤشر جودة المياه (WQI) وبمستوى عالٍ جداً من الدقة، مما سيزيد من تحسين إدارة جودة المياه.

الكلمات المفتاحية: جودة المياه، المياه السطحية، الانحدار الخطي متعدد المتغيرات، التعلم الآلي، خوارزميات الانحدار، مؤشر جودة المياه (WQI)، بحيرة السن.



حقوق النشر : مجلة جامعة تشرين- سورية، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص

CC BY-NC-SA 04

* أستاذ مساعد - قسم الهندسة البيئية - كلية الهندسة المدنية - جامعة تشرين - اللاذقية - سورية. r.jafar@tishreen.edu.sy

مقدمة:

الماء مورد طبيعي أساسي لجميع أشكال الحياة على كوكب الأرض، يجب أن يكون الماء الآمن خالياً من أي مواد كيميائية ضارة أو كائنات حية دقيقة بتركيز تسبب أضراراً صحية وفقاً لتوصيات (منظمة الصحة العالمية، 2017) [1]. مع تزايد النمو والتطور في جميع أنحاء العالم، تعرضت مصادر المياه السطحية مثل الأنهار والبحيرات لتلوث واسع النطاق من مصادر مختلفة وفقاً (لبرنامج الأمم المتحدة للبيئة ، 2016) [2]. تعتبر الأنهار والبحيرات من المصادر الرئيسية للمياه العذبة، وتمثل أحد أهم الموارد المائية للاستخدامات المختلفة، مثل الزراعة والصناعة والاستخدامات المعيشية وغيرها، وتعد بمثابة شريان الحياة للمجتمعات، وتلعب دوراً أساسياً في التنمية الاجتماعية والاقتصادية والبيئية [3].

ومع ذلك، تتعرض هذه المسطحات المائية للاستنزاف الشديد نتيجة الأنشطة البشرية المفرطة، مثل التصنيع والتحصير والنمو السكاني. بالإضافة إلى ذلك، تسبب سوء إدارة الموارد المائية وتغير المناخ في انخفاض جودة المياه خلال العقود الأخيرة، وهو ما أفضى إلى تلوث المياه السطحية [4]. إذ تعتمد جودة المياه السطحية في منطقة ما إلى حد كبير على طبيعة ومستوى الأنشطة البشرية المختلفة، مثل الصناعة والزراعة وغيرها، في الأحواض المائية الصبابة المعنية. تخضع تركيبات المياه السطحية الكيميائية والفيزيائية والبيولوجية للتأثيرات العديدة، بما في ذلك التأثيرات الطبيعية مثل هطول الأمطار، جغرافية الأحواض الصبابة للمياه، الغلاف الجوي، والجيولوجيا، بالإضافة إلى التأثيرات البشرية مثل الأنشطة الصناعية والزراعية والمنزلية [5]. ويؤدي التلوث المتزايد للمياه السطحية إلى تدهور جودتها وتهديد صحة الإنسان، ويؤثر على توازن النظام البيئي المائي ويعيق التنمية الاقتصادي والتطور الاجتماعي [6].

وفقاً لتقرير منظمة الصحة العالمية، تسبب المياه نحو 80% من جميع الأمراض التي تصيب البشر. وعندما تتلوث المياه الجوفية، فإنه لا يمكن استعادة جودتها إلا من خلال وقف تدفق الملوثات من المصدر. لذلك فمن الضروري مراقبة جودة المياه السطحية والجوفية باستمرار وتحسين الطرق والوسائل لحمايتها.

ويعد مؤشر جودة المياه (WQI) أحد أكثر الأدوات فعالية لنقل المعلومات حول جودة المياه إلى المجتمعات المحلية والمهتمين وصناع القرار، ويمكن استخدامه لتحديد الجودة العامة للمياه وتحديد التحسينات الممكنة في النظام البيئي المائي [7]. إن تطبيق الأساليب الإحصائية متعددة المتغيرات مثل الانحدار الخطي المتعدد (MLR)، التحليل العنقودي Cluster Analysis (CA)، تحليل المكونات الرئيسية (PCA) Principal Components Analysis، تحليل العوامل Factor Analysis (FA)، التحليل التمييزي (DA) Discriminant Analysis، جميعها مفيدة في تخفيض تعقيد بيانات جودة المياه الكبيرة (تقليل عدد المتغيرات) دون فقدان المعلومات الأصلية [8]، يساعد تطبيق هذه التقنيات الإحصائية في تفسير البيانات المعقدة لفهم حالة جودة المياه البيئية بشكل أفضل ومعرفة المصادر أو العوامل المحتملة التي تؤثر على أنظمة المياه، بالإضافة إلى حل سريع لمشاكل التلوث من أجل تقييم بسيط وفعال من حيث التكلفة لجودة المياه [9].

في السنوات الأخيرة، تم استخدام تقنيات التعلم الآلي (Machine Learning (ML) على نطاق كبير لتقييم جودة المياه، بما في ذلك تقدير مؤشر جودة المياه WQI [10]. أثبتت هذه التقنيات أنها أدوات قوية لنمذجة السلوك الخطي وغير الخطي المعقد في الأبحاث البيئية والموارد المائية [11]. توضح مراجعة الأدبيات التي أجريناها أن كل خوارزمية من خوارزميات التعلم الآلي (ML) لها نقاط قوتها وعيوبها. ويعتمد سلوكها على متغيرات إدخال جودة المياه في مناطق الدراسة المختلفة [12]، [13].

يستخدم العديد من الباحثين مؤشر جودة المياه (WQI) على نطاق واسع لتقييم جودة المياه الجوفية والسطحية، وذلك باستخدام طرق مختلفة لحسابه. يجمع WQI بين متغيرات المياه المختلفة الفيزيائية والكيميائية والجرثومية، ويحولها إلى قيمة واحدة تعكس حالة المياه. ويساعد هذا المؤشر صانعي القرار على اتخاذ إجراءات فعالة لإدارة الموارد المائية والحفاظ على جودتها [14].

تم دعم صياغة واستخدام مؤشرات الجودة بشدة من قبل الهيئات المسؤولة عن إمداد المياه والتحكم في تلوثها. حيث تظهر الحاجة إلى ترجمة بيانات نوعية المياه بشكل يسهل فهمها بمجرد جمعها من خلال أخذ العينات والتحليل المخبرية. وعند تطوير وتطبيق مؤشر جودة المياه (WQI)، فإنه يعمل كأداة ملائمة لدراسة ميول التلوث وتسليط الضوء على ظروف بيئية محددة، وبالتالي يمكن أن يساعد صانعي القرار في تقييم فعالية برامج حماية هذه المصادر المائية [15]. كان حوض السن محط اهتمام الباحثين على مدى السنوات الأخيرة، وذلك نظراً لأهميته الكبيرة في تغذية مدينتي اللاذقية وطرطوس بمياه الشرب إضافة إلى استخداماته الأخرى الزراعية والصناعية [16]، وقد تم حساب مؤشر جودة المياه لبحيرة السن (NSFWQI) National Sanitation Foundation Water Quality Index للسنوات (2011-2007-2004-1991)، وأظهرت نتائج التقييم أن جودة المياه تتراوح بين المجال الجيد والوسط وفقاً لنقطة الرصد المعتمدة. بالمقابل، ركزت دراسة أخرى [17] على تقييم جودة مياه البحيرة من خلال بارامتر العكارة وتطبيق نموذج الانحدار الذاتي غير الخطي للشبكات العصبونية للتنبؤ بقيمتها.

في هذه الدراسة، تم جمع مجموعة بيانات لمأخذ مياه شرب مدينة اللاذقية للعامين (2021-2022) من خلال برنامج رصد وحماية بحيرة السن (530 قياساً)، وقد تم حساب مؤشر جودة المياه بالطريقة الحسابية الموزونة Weighted arithmetic method، وتم تطبيق نموذج الانحدار الخطي متعدد المتغيرات (MLR) وتطويره بحيث يتضمن أهم المتغيرات المسؤولة عن تقييم جودة مياه المأخذ. بعد ذلك، تم تطبيق عدة نماذج للتعلم الآلي (ML) والتي تعتمد على الخوارزميات القائمة على الانحدار (Regression) وشجرة القرار (Decision tree) والتعزيز (Boosting)، وكذلك تم تطبيقها وفق السلسلة الزمنية المدروسة (Time series)، من أجل التنبؤ بقيم مؤشر جوده مياه المأخذ (WQI).

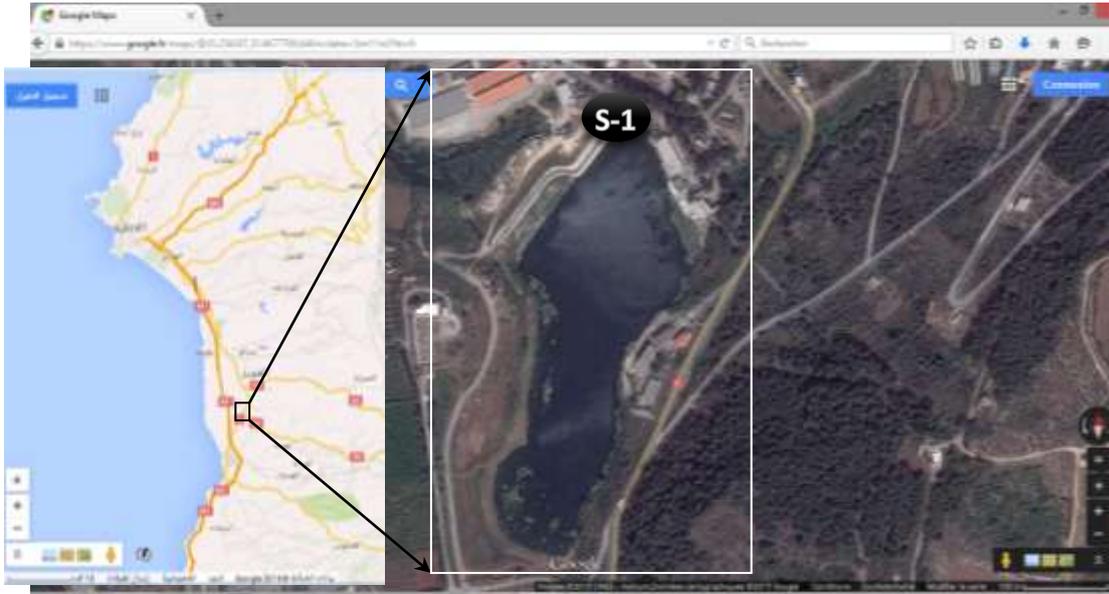
أهمية البحث وأهدافه:

تكمن أهمية البحث في تقييم جودة مصادر المياه السطحية والحد من تلوثها، وذلك بدراسة العديد من بارامترات التلوث المختلفة الفيزيائية والكيميائية على مدى عامين متتاليين (2021-2022) باستخدام مؤشر جودة المياه (WQI) بالطريقة الحسابية الموزونة، ونموذج الانحدار الخطي متعدد المتغيرات (MLR)، وكذلك مجموعة من نماذج التعلم الآلي المتعددة والتي تستند إلى خوارزميات مختلفة. تساعد هذه التقنيات في تفسير مجموعات البيانات، تقييم جودة المياه السطحية، تقليص عدد التجارب المخبرية المختلفة وتواترها من خلال استخدام هذه النماذج.

تهدف هذه الدراسة من خلال استخدام هذه التقنيات إلى تقييم نتائج التحاليل المخبرية لبارامترات تلوث بحيرة السن (والتي تم جمعها من مأخذ مياه شرب مدينة اللاذقية كنقطة رصد معتمدة) وذلك عن طريق مقارنتها مع الحدود المسموحة عالمياً (WHO)، ومن ثم حساب مؤشر جودة المياه (WQI) Water Quality Index لتحديد تصنيفها ودرجة تلوثها. كذلك مقارنة أداء نماذج الانحدار الخطي متعددة المتغيرات في التنبؤ بهذا المؤشر لجودة مياه المأخذ مع نماذج تعلم الآلة (ML) ذات الانحدار وشجرة القرار والتعزيز، وكذلك أيضاً مع نموذج السلسلة الزمنية المدروسة، واعتماد النموذج الأمثل الذي يساهم في تخفيض كلفة العمل المخبري والجهد والوقت.

طرائق البحث ومواده:**1. منطقة الدراسة Study area**

تتشكل بحيرة السن من نبع السن الذي يعتبر من المصادر المائية الهامة والرئيسية في سورية، يغذي البحيرة 14 ينبوع، وتقع البحيرة بين دائرتي عرض (35° 15' 31'' N) - (35° 15' 13'' N) وخطي طول (35° 58' 09'' E) - (35° 57' 59'' E)، يبلغ محيط البحيرة وفق (Google Earth Pro) 1267 m ومساحتها 64337 m²، تبلغ سعة البحيرة 400,000 m³، العمق الأعظمي لها 9 m، يبين الشكل (1) موقع وحدود البحيرة. تغذي مدن الساحل السوري (اللاذقية وطرطوس) وبعض قرى بمياه الشرب، والتي يشكل عدد سكانها حوالي 5 % من سكان سورية، كما تشكل مياه نبع السن حوالي 60 % من الموارد المائية المستخدمة للشرب في حوض الساحل، يبلغ معدل ضخ المياه إلى مناطق الاستهلاك حوالي 13500 m³/h. هذا إلى جانب استخداماته الأخرى الهامة للري والصناعة، ومما يزيد من أهميته أيضاً هو الحاجة المتنامية لمياهه نتيجة التوسع السكاني والزراعي والصناعي في مدن وريف الساحل السوري، وتتطلع السلطات للاستفادة من مياهه في إمداد مدن أخرى بمياه الشرب، غير أن وجود المناطق السكنية القريبة من حرم بحيرة السن والمتناثرة في حوضه، وكذلك الأراضي الزراعية المجاورة يساهم في تدهور نوعية مياه البحيرة ومياه الآبار والينابيع في المناطق الريفية القريبة والموجودة في حوض السن.



الشكل (1) موقع نقطة رصد مأخذ مياه شرب مدينة اللاذقية على بحيرة السن

2. جمع العينات وتحليلها Sample collection and analysis

تم الاعتماد في هذه الدراسة على قياسات العينات المتوفرة والمأخوذة من قبل المعنيين بمراقبة وحماية البحيرة لعامين متتاليين (2021-2022) بمعدل عينة واحدة باليوم، حيث أخذت العينات من نقطة رصد مأخذ مدينة اللاذقية بشكل يومي وفق برنامج المراقبة المعتمد كما هو موضح في الشكل (1) والجدول (1). تضمن العمل المخبري إجراء عدد من التحاليل على العينات المأخوذة من نقطة الاعتيان في مأخذ مياه شرب اللاذقية، وشملت تحديد قيم العديد من المؤشرات الدالة على نوعية المياه والتي تمثلت ب: الأس الهيدروجيني (pH)، الكبريتات (SO₄⁻), النترات (NO₃⁻)، النتريت (NO₂⁻)، الأمونيوم (NH₄⁺)، الفوسفات (PO₄⁻³)، العكارة (Tur) والناقلية الكهربائية (EC).

يوضح الجدول (2) بعض المؤشرات التحليلية الإحصائية الوصفية للبيانات المقاسة والمعتمدة في البحث. كما يبين الشكل (2) مصفوفة الارتباط بين كافة متغيرات الدراسة بما فيها مؤشر جودة المياه بعد عملية حسابه.

الجدول (1): نقطة مأخذ مياه شرب مدينة اللاذقية.

نقطة الرصد	اسم النقطة	خط طول	دائرة عرض	الارتفاع (m)
Station_S-1	مأخذ اللاذقية	Lon : 035° 58' 04'' E	Lat : 35° 15' 31'' N	19

الجدول (2): المؤشرات الإحصائية الوصفية

Descriptive Statistics					
المؤشر	N	Minimum	Maximum	Mean	Std. Deviation
WQI	530	15.969	69.499	23.6427	4.24165
pH	530	7.00	08.90	7.7704	0.17769
SO ₄ ⁻ (mg/l)	530	6.00	18.00	11.0811	2.19447
NO ₃ ⁻ (mg/l)	530	0.66	12.00	3.3711	0.89827
NO ₂ ⁻ (mg/l)	530	0.00	03.20	0.0212	0.13915
NH ₄ ⁺ (mg/l)	530	0.00	0.05	0.0023	0.00592
PO ₄ ⁻³ (mg/l)	530	0.01	0.90	0.2053	0.09166
Turbidity (NTU)	530	0.38	09.94	1.9497	1.00350
EC (µS/cm)	530	445.00	503.00	477.7830	7.92975



الشكل (2): مصفوفة الارتباط بين متغيرات الدراسة

3. مؤشر جودة المياه Water quality index WQI

من أجل حساب مؤشر جودة المياه لمأخذ مياه شرب اللاذقية في بحيرة السن، تم تحديد ثمانية بارامترات فيزيائية-كيميائية وفقاً لأهميتها وهي: الأس الهيدروجيني (pH)، الكبريتات (SO_4^-)، النتريت (NO_2^-)، النترات (NO_3^-)، الأمونيوم (NH_4^+)، الفوسفات (PO_4^{-3})، العكارة (Tur) والناقلية الكهربائية (EC)، وبالاعتماد على قاعدة بيانات القياسات اليومية لهذه البارامترات والقيم المعيارية الموصى بها من قبل المنظمات الدولية المختصة. ترتبط الخطوات التالية بتطوير أي مؤشر للجودة WQI، بالنظر إلى التطوير المستهدف، ويمكن أيضاً اتخاذ خطوات إضافية، وهذه الخطوات هي: اختيار البارامترات، تحويل البارامترات ذات الواحدات والأبعاد المختلفة إلى مقياس مشترك، تعيين الأوزان لجميع البارامترات، تجميع المؤشرات الفرعية لتحديد درجة المؤشر النهائية. حيث تم حساب مؤشر جودة المياه وفقاً للمراحل التالية:

1- إعطاء كل بارامتر من البارامترات الثمانية السابقة وزناً بقيمة تتراوح من (1-5) تبعاً لأهمية كل بارامتر ودوره في التأثير على جودة المياه بالاعتماد على آراء الخبراء في الدراسات المرجعية السابقة. القيم المعيارية والأوزان الملحقه بالبارامترات والأوزان النسبية للبارامترات الثمانية مبينة في الجدول (3):

الجدول (3): القيم المعيارية المسموحة والأوزان المخصصة والنسبية لبارامترات الدراسة.

المؤشر	Water quality stander (Si) معيار جودة المياه	Assigned weight (wi) الوزن المخصص	Relative weight (Wi) الوزن النسبي
pH	6.5-8.5	4	0.1212
SO_4^- (mg/l)	250	4	0.1212
NO_3^- (mg/l)	50	5	0.1515
NO_2^- (mg/l)	1	5	0.1515
NH_4^+ (mg/l)	0.50	3	0.0910
PO_4^{-3} (mg/l)	0.50	4	0.1212
Turbidity (NTU)	5	4	0.1212
EC ($\mu S/cm$)	1000	4	0.1212
Total		$\sum wi = 33$	$\sum Wi = 1$

تم حساب الوزن النسبي وفق المعادلة التالية:

$$Wi = \frac{wi}{\sum_{i=1}^n wi} \quad \text{Eq.1}$$

حيث:

Wi: الوزن النسبي.

wi: الوزن المخصص للبارامتر.

n: عدد البارامترات.

2- من أجل كل بارامتر، تم حساب مقياس تقدير الجودة (qi) من خلال تقسيم قيم تراكيز القياسات المخبرية على القيم المعيارية وفقاً لمنظمة الصحة العالمية، وتم ضرب الناتج بـ 100 وفق المعادلة التالية:

$$q_i = \frac{C_i}{S_i} \times 100 \quad \text{Eq.2}$$

مع ملاحظة أنه من أجل باراميتري الأس الهيدروجيني والأوكسجين المنحل فقط يتم حساب مقياس تقدير الجودة (q_i) وفق المعادلة التالية:

$$q_i \text{ Do, pH} = \frac{(C_i - V_i)}{(S_i - V_i)} \times 100 \quad \text{Eq.3}$$

حيث:

C_i : القيمة المقاسة لباراميتري جودة المياه.

V_i : القيمة المثالية للأوكسجين المنحل وهي 14.7 ، ولأس الهيدروجيني 7.

S_i : القيمة المعيارية لباراميتري جودة المياه.

3- يتم ضرب الوزن المخصص بالوزن النسبي للحصول على المؤشرات الفرعية S_{li} ، ويكون WQI هو مجموع المؤشرات الفرعية وفق المعادلة:

$$S_{li} = W_i \times q_i \quad \text{Eq.4}$$

$$WQI = \sum_{i=1}^n S_{li} \quad \text{Eq.5}$$

يتم تصنيف قيم WQI وفقاً للمقياس الذي تم اقتراحه في الدراسات المرجعية السابقة، الجدول (4).

الجدول (4) مقياس مؤشر جودة المياه. [18]

مؤشر جودة المياه WQI	0-25	26-50	51-75	76-100	أكبر من 100
جودة المياه	ممتازة	جيدة	ضعيفة	ضعيفة جداً	غير مناسبة

النتائج والمناقشة:

1. تقييم تراكيز بارامترات جودة مياه مأخذ اللاذقية: Evaluation of water quality parameter concentrations for the Latakia water intake

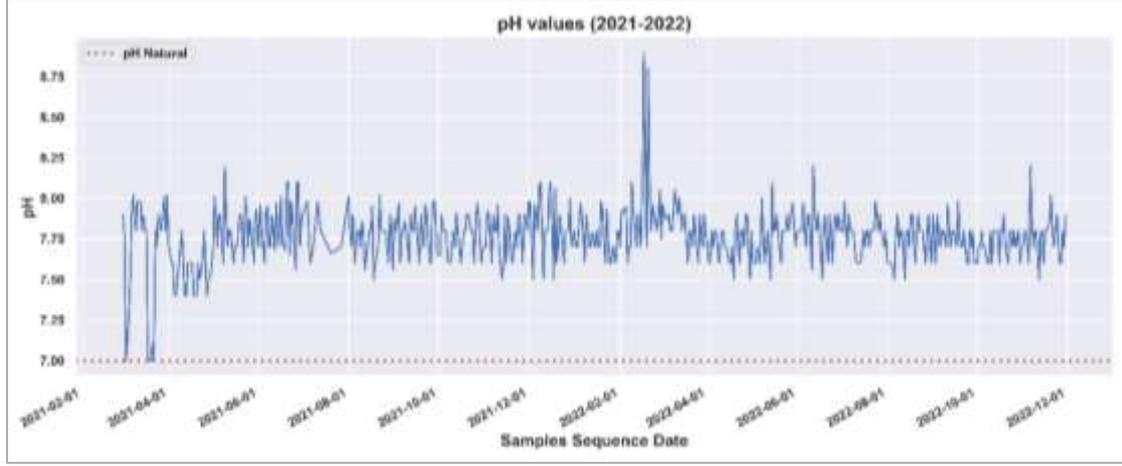
من أجل دراسة مدى تأثر نوعية المياه لمأخذ مياه اللاذقية بمصادر التلوث الموجودة في منطقة الدراسة، تم قطف عينات مائية من نقطة الرصد المذكورة بشكل يومي، على مدار عامين (2021-2022)، وذلك لقياس أهم مؤشرات التلوث فيها والمتمثلة بالأس الهيدروجيني (pH)، الكبريتات (SO_4^-)، النترات (NO_3^-)، النتريت (NO_2^-)، الأمونيوم (NH_4^+)، الفوسفات PO_4^{-3} ، العكارة (Tur) والناقلية الكهربائية (EC)، وفق الآتي:

1.1.1 الأس الهيدروجيني pH

يعد الأس الهيدروجيني (pH) مؤشراً مهماً لجودة مياه الشرب، حيث يساعد على تقييم قابلية المياه لأغراض الشرب والاستخدامات الأخرى. يجب أن يتراوح الأس الهيدروجيني من أجل ضمان جودة مياه الشرب وفق معايير منظمة الصحة العالمية بين 6.5 و 8.5. إذا كان الأس الهيدروجيني أقل من 6.5، فإن ذلك يعني أن المياه حمضية ويمكن أن تتسبب في تآكل الأنابيب والتجهيزات المستخدمة في نقل وتوزيع المياه. وإذا كان الأس الهيدروجيني أعلى من 8.5، فإن ذلك يعني أن المياه قلوية ويمكن أن تتسبب في ترسيب الكالسيوم والمغنيسيوم وتكون رواسب على الأسطح والأنابيب.

تشير النتائج المبينة في الشكل (3) الى التآرجح النسبي لقيم تركيز الأس الهيدروجيني (H^+) في المياه والتي تراوحت بين (7-8.9) وإن كافة العينات تقع ضمن المجال القلوي؛ تعود القيم القلوية الى وجود أيونات البيكربونات، بينما في

حالة الانخفاض النسبي للقيم فقد يعود الى ارتفاع تركيز الأملاح وسيادة الطور الكلوريدي والكبريتي على حساب طور البيكربونات مما يؤدي الى خفض قيمة الأس الهيدروجيني قليلا نحو الحامضية، وعموماً فإن مياه العينات المدروسة هي ضمن الحدود القلوية المناسبة للشرب ماعدا قيمتين فقط تجاوزتا القيمة المسموحة كما هو موضح في الشكل (3) .



الشكل (3): تغير قيم الأس الهيدروجيني pH في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

1.2. شاردة الكبريتات SO_4^{+2} Sulfites

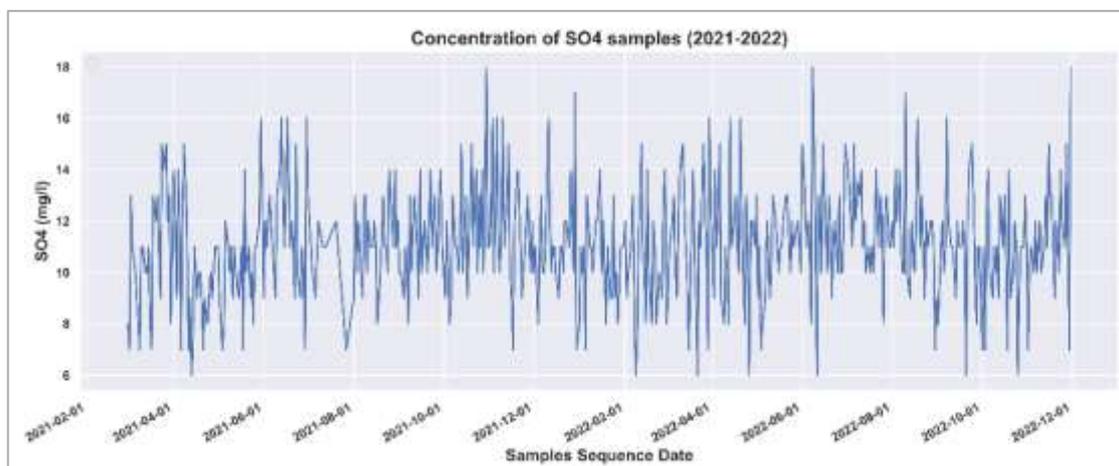
توجد الكبريتات بشكل طبيعي في بعض مصادر المياه الجوفية والسطحية وهي لا تشكل خطراً صحياً على الإنسان عندما تتواجد في مستويات منخفضة. ومع ذلك، قد يؤدي الاستخدام المفرط للمواد الكيميائية المحتوية على الكبريتات في الصناعة أو الزراعة إلى تلوث مصادر المياه وزيادة تركيز الكبريتات إلى مستويات تشكل خطراً صحياً. الكبريتات هي أملاح تحتوي على عنصر الكبريت وتتشكل بشكل رئيسي من خلال عمليات التأكسد الحيوية للمواد العضوية في المياه السطحية والجوفية، وتتشكل أيضاً من عمليات التأكسد للمعادن الموجودة في التربة والصخور المحيطة بمصدر المياه. وتتضمن المواد العضوية التي تؤدي إلى تكوين الكبريتات بشكل رئيسي المواد النباتية والحيوانية والفضلات البشرية والحيوانية والبكتيريا اللاهوائية.

عمليات التنقية التي تخضع لها المياه السطحية والجوفية قد تزيل بعضاً من الكبريتات، ولكنها غير قادرة على التخلص منها بشكل كامل، مما يعني أن معظم مياه الشرب سوف تحتوي على كميات معينة من الكبريتات.

لتقييم وجود الكبريتات في مياه الشرب، يتم مراجعة تركيزها في المياه ومقارنتها بالحدود المسموحة والمعايير الصحية المعتمدة المتعلقة بجودة المياه. على سبيل المثال، وفقاً لمنظمة الصحة العالمية، يجب أن يكون التركيز الآمن للكبريتات في مياه الشرب أقل من 500 ملليغرام / لتر.

تشير نتائج الدراسة والمبينة في الشكل (4) إن قيمة تركيز أيون الكبريتات تراوحت ضمن المجال (6-18) mg/l وهي ضمن القيم المسموحة وفق منظمة الصحة العالمية، وربما يعود هذا التركيز المنخفض جداً إلى طبيعة التكوين الجيولوجي لمنطقة حوض السن، حيث أن تراكيز أيونات الكبريتات في المياه الجوفية تعود إلى نوع الصخور والزمن التي مرت بها ويعتبر الجبس والانهيدريت (كبريتات الكالسيوم اللامائية) المصدر الرئيسي للكبريتات، بينما تعتبر الصخور الرملية ذات محتوى قليل جداً من المكونات الجيرية، بالإضافة إلى أن النشاطات البكتيريولوجية في طبقات التربة تؤدي دوراً مهماً في تفاعلات الأكسدة والإحتزال لأطوار الكبريت وغالباً ما يزداد تركيزه في المياه الجوفية بزيادة

العمق نظراً لزيادة احتمال ذوبان الصخور الملامسة لها، كما أن لعمليات التحلل اللاهوائي للمركبات الحاوية على الكبريت والجبس والانهدريت المشكل للتكوين الجيولوجي حيث يتم اختزال الكبريت بواسطة بكتريا Thiobacellus desulfuricans لتكوين مركبات الكبريت، وعند توفر الظروف الهوائية تتأكسد إلى حامض الكبريتيك.

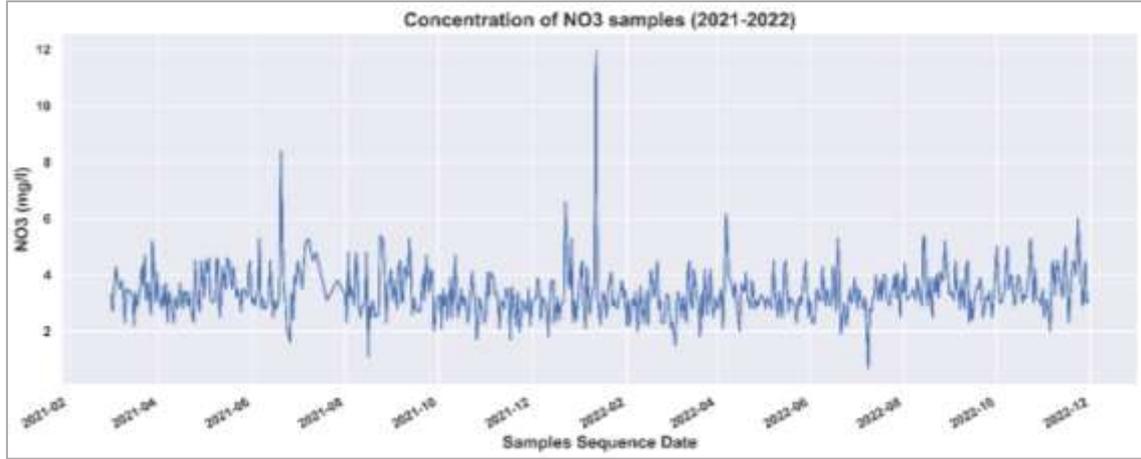


الشكل (4): تغير قيم شاردة الكبريتات $(\text{SO}_4^{+2} \text{ mg/l})$ في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

1.3. شاردة النترات NO_3^- Nitrates

تعتبر النترات من أكثر الملوثات انتشاراً في المياه الجوفية، وهي مشكلة رئيسة في بعض طبقات المياه الجوفية الضحلة، والمصادر الرئيسية للنترات في المياه الجوفية هي مياه الصرف الصحي والصناعي، البقايا النباتية والمخلفات الحيوانية، بالإضافة إلى الأسمدة الكيماوية والمبيدات الحشرية والعشبية المستعملة في الأنشطة الزراعية. يجب السيطرة على تركيز النترات في مياه الشرب، لأن الزيادة في تركيزها تسبب أمراضاً صحية فهي مادة محرضة على السرطان، كما أن الأمعاء لدى البشر يمكنها أن تقوم بكسر النترات وتحويلها إلى النتريت NO_2^- الذي يؤثر بدوره على قدرة خلايا الدم الحمراء على حمل الأكسجين، وبالتالي يسبب ما يعرف بالميثيموغلوبينيميا لدى الرضع والحوامل (مرض ازرقاق الرضع) [19]. إن الحد الأقصى المسموح للنترات في مياه الشرب حسب المواصفات القياسية السورية هو 50 mg/l .

قمنا بتمثيل نتائج التحاليل المخبرية لتغيرات تراكيز شاردة النترات المقيسة يومياً في عينات المياه المقطوفة من مأخذ مدينة اللاذقية على مدار العامين (2021-2022) وفق الشكل (5).



الشكل (5): تغير قيم شاردة النترات NO_3^- (mg/l) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

- من خلال نتائج التحاليل المخبرية والمخططات البيانية نلاحظ ما يلي:
- سجلت أقل قيمة لتركيز شاردة النترات 0.66 mg/l خلال شهر تموز 2022، في حين سُجلت أكبر قيمة لها في شهر كانون الثاني 2022، حيث بلغت 12 mg/l .
 - تتراوح قيم تراكيز شاردة النترات المقيسة في عينات المياه المدروسة ضمن المجال $(0.66 - 12) \text{ mg/l}$ ، وتقع جميعها ضمن الحد المسموح به لمياه الشرب 50 mg/l حسب المواصفة القياسية السورية.
 - من الشكل (5) نلاحظ أن القيم العليا لتراكيز شاردة النترات في عينات المياه المدروسة، قد تركزت في فترة الشتاء، ويعزى ذلك إلى دور الهطولات المطرية التي تقوم بغسل آثار الأسمدة من التربة والأراضي الزراعية وتحملها مباشرة إلى خزان المياه الجوفية.

1.4. شاردة النتريت NO_2^- Nitrite

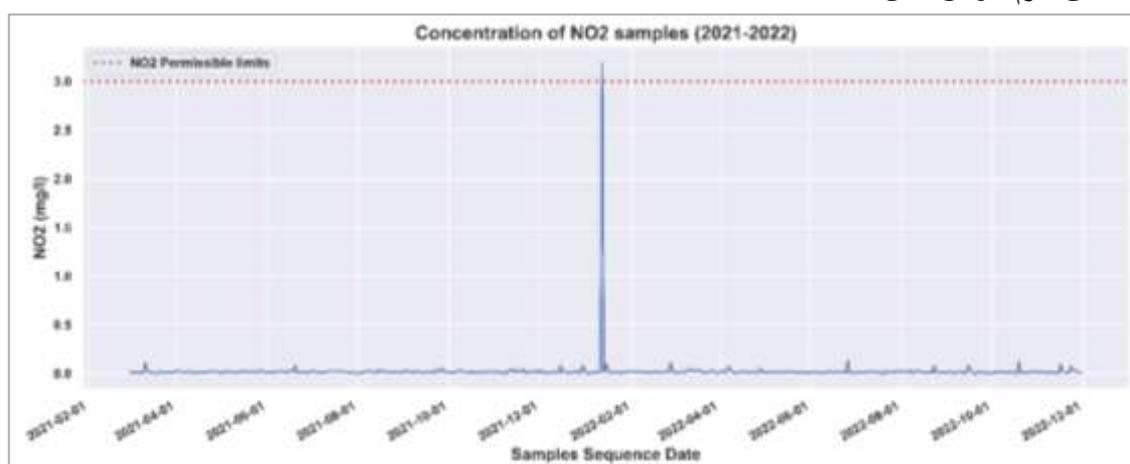
- يتم تشكل النتريت في مياه الشرب عادةً عندما تتعرض المياه للتلوث من مصادر مختلفة مثل الأسمدة الزراعية والمخلفات الحيوانية والمياه العادمة والصناعات الكيميائية. ويمكن أن يتشكل النتريت في المياه الجوفية والسطحية بشكل طبيعي أيضاً عندما تتعرض المياه للتفاعل مع النترات الموجودة في التربة.
- يمكن أن يتشكل النتريت في مياه الشرب نتيجة لعدة آليات، نذكر منها:
- التلوث الزراعي: يتم إضافة النترات والأمونيوم واليوريا إلى التربة كأسمدة نتروجينية للمحاصيل، من خلال ذلك يمكن أن تترشح المياه إلى الخزان الجوفي من التربة محتوية على نترات ونتريت. كما يمكن للنترات في الأسمدة الزراعية التحلل والتحول إلى نتريت بواسطة العمليات البكتيريولوجية.
 - التلوث الصناعي: يمكن أن تتسرب المواد الكيميائية الموجودة في المواد الخام أو المستخدمة في الصناعة إلى التربة والمياه الجوفية، ومن ثم إلى مصادر المياه السطحية. كما يمكن لبعض الصناعات المستخدمة للنترات، مثل الأسمدة والمبيدات الحشرية والمواد الكيميائية الأخرى، أن تسهم في تلوث المياه الجوفية والسطحية بالنتريت.
 - الصرف الصحي: يمكن أن تتسرب مياه الصرف الصحي والمخلفات الحيوانية من مرافق الحيوانات والدواجن إلى المياه الجوفية ومصادر المياه السطحية، مما يؤدي إلى زيادة تركيز النترات والنتريت.

- التلوث الجوي: يمكن أن تحمل الأمطار والتلوج والرذاذ الجوي الملوثات الصناعية والزراعية إلى المصادر المائية، وهذا قد يؤدي إلى زيادة تركيز النترات والنترات في المياه السطحية.

- التلوث بالأدوية: يمكن أن تتسرب بعض الأدوية إلى التربة والمياه الجوفية من مصادر مختلفة، وقد يتم العثور عليها في مصادر المياه السطحية والمياه الجوفية.

يتم قياس مستوى النترات في المياه في المخبر المتخصصة بعدد من الطرق المختلفة، منها الطريقة اللونية (والتي تعتمد على تفاعل النترات مع المواد الكيميائية لإنتاج لون يمكن قياسه)، حيث يجب ألا يتجاوز تركيز النترات في مياه الشرب الحد المسموح به بحسب معايير منظمة الصحة العالمية وهو (1 mg/l).

تشير نتائج الدراسة والمبينة في الشكل (6) إن قيمة تركيز أيون النترات تراوحت ضمن المجال (0-0.13) mg/l وهي ضمن القيم المسموحة وفق منظمة الصحة العالمية، باستثناء قيمة وحيدة في شهر كانون الثاني 2022 حيث بلغت (3.2) mg/l. وربما يعود هذا التركيز المنخفض إلى الإجراءات الصارمة فيما يخص النشاطات البشرية المسموح بها في مناطق حرم حوض السن.



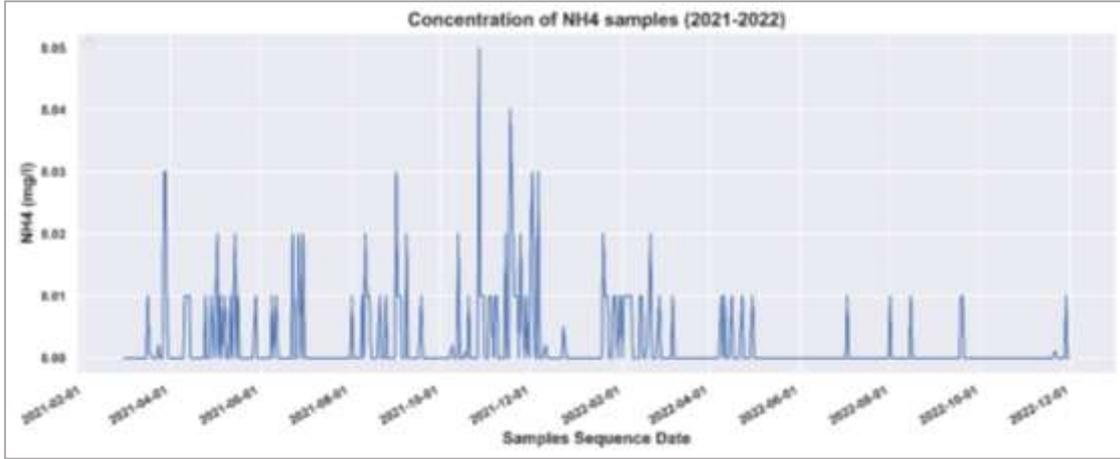
الشكل (6): تغير قيم شاردة النترات (NO_2^- (mg/l) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

1.5. شاردة الأمونيوم NH_4^+ Ammonium

تتشكل الأمونيا في مياه الشرب عادة بسبب عملية التحلل الطبيعي للمواد العضوية والنباتية والحيوانية في البيئة المائية. كما يمكن أن يحدث تشكل الأمونيا نتيجة لتسرب المخلفات الحيوانية والسماد والنفايات الصناعية إلى المياه الجوفية والسطحية. وبشكل عام، يمكن تقييم وجود الأمونيا في مياه الشرب بواسطة تحليل المياه وقياس مستوى الأمونيا فيها. ويفضل الحفاظ على مستويات منخفضة من الأمونيا في مياه الشرب، حيث أن ارتفاع مستويات الأمونيا يمكن أن يشير إلى وجود تلوث في المياه بمواد عضوية قابلة للتحلل، وهذا يمثل خطراً على صحة المستخدمين للمياه.

وبشكل عام، يمكن قياس تركيز NH_4^+ في مياه الشرب بإجراء اختبارات كيميائية، يتم ذلك عادةً بإضافة محلول من الكلور إلى عينة الماء، ومن ثم قياس مستوى الأمونيا في الماء بعد فترة زمنية محددة. يجب أن يكون مستوى NH_4^+ في الماء أقل من 0.5 جزء في المليون (ppm) وفقاً لمواصفات منظمة الصحة العالمية للمياه الصالحة للشرب.

تشير نتائج الدراسة خلال العامين (2021-2022) والمبينة في الشكل (7) إن قيمة تركيز أيون الأمونيوم تراوحت ضمن المجال (0-0.05) mg/l وهي جميعها ضمن القيم المسموحة وفق منظمة الصحة العالمية.

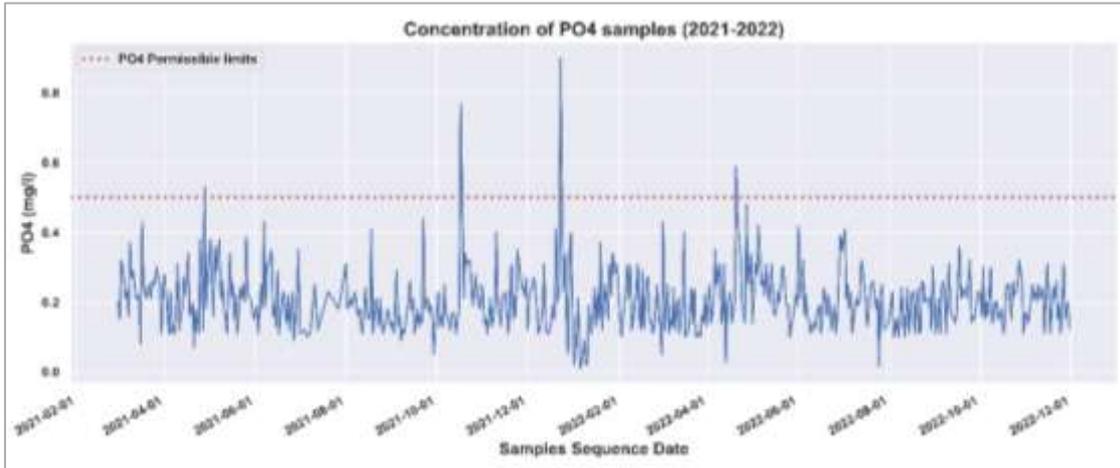


الشكل (7): تغير قيم شاردة الأمونيوم (NH_4^+ (mg/l) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

1.6. شاردة الفوسفات PO_4^{3-} Phosphates

تنتج مركبات الفوسفات من اتحاد المركبات الفوسفورية والأكسجين، ويمكن التمييز بين مركبات الفوسفات العضوية التي تنتج عن تفسخ البقايا النباتية والحيوانية، وتحلل الفضلات وبقايا الأطعمة، ومركباته اللاعضوية التي يعتبر السماد المضاف للأغراض الزراعية مصدرها الرئيسي، بالإضافة لصناعة الصابون ومياه الصرف المنزلي، وبالتالي يمكن للفوسفات أن ينتقل إلى مصادر المياه من مياه الصرف الصحي والصناعي والزراعي، لذلك يعتبر قياس تركيز شاردة الفوسفات في مياه الشرب والمياه الطبيعية مؤشراً مهماً عند دراسة جودة مصادر المياه [20]. ووفق المواصفات القياسية السورية يبلغ الحد الأقصى المسموح لتركيز شاردة الفوسفات في مياه الشرب 0.5 mg/l.

نبين في الشكل (8) نتائج التحاليل المخبرية لتغيرات قيم شاردة الفوسفات المقيسة في عينات المياه المقطوفة من مأخذ مياه شرب مدينة اللاذقية على مدار العامين (2021-2022).



الشكل (8): تغير قيم شاردة الفوسفات (PO_4^{3-} (mg/l) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

من خلال نتائج التحاليل المخبرية والمخططات البيانية نلاحظ ما يلي:

- بلغت أقل قيمة لتركيز شاردة الفوسفات 0.01 mg/l وذلك خلال شهر تموز للعام 2022، في حين سُجلت أكبر قيمة لها في شهر كانون الأول، حيث بلغت 0.9 mg/l.

- نلاحظ من الشكل (8) أن قيم تركيز شاردة الفوسفات المقيسة في عينات المياه المدروسة ضمن المجال mg/l (0.01–0.9)، وتقع ضمن المعايير المسموحة لمياه الشرب 0.5 mg/l، باستثناء خمسة نقاط تجاوزت الحد المسموح منها نقطتان في تشرين الأول للعام 2021، نقطة ثالثة في شهر كانون الأول لنفس العام، ونقطة في شهر نيسان للعامين 2021 و 2022.

- كما نلاحظ من الشكل (8) أن القيم الخمسة التي تجاوزت الحد المسموح لشاردة الفوسفات تقع في فترة الشتاء، ويُعزى ذلك إلى هطول الأمطار في هذا الفصل التي تغسل آثار الأسمدة والمبيدات الحشرية والعشبية من التربة والأراضي الزراعية، كما أن لتلك الهطولات دوراً في حمل الرشاحة الغنية بالملوثات ومنها شوارد الفوسفات، من الأراضي الزراعية في الأحواض الصبابة القريبة وإيصالها إلى خزان المياه الجوفية.

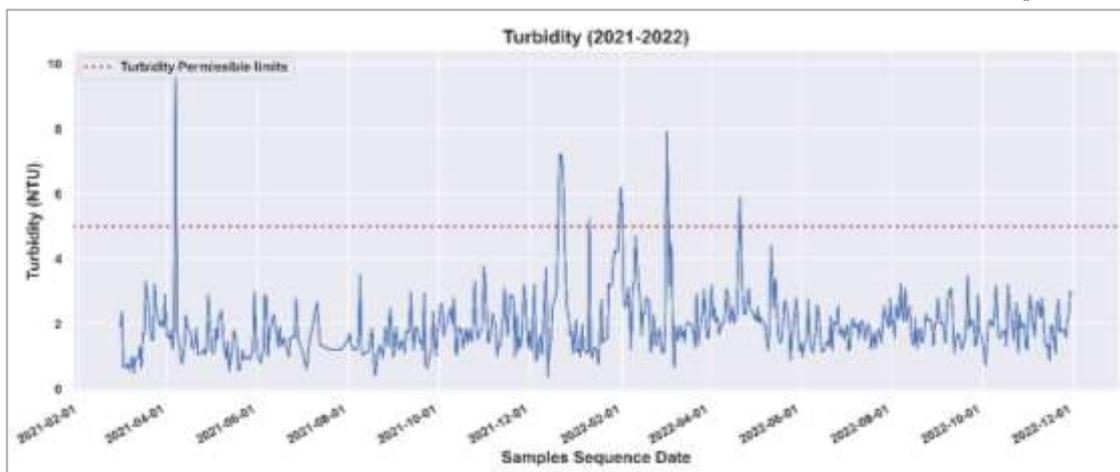
1.7. العكارة Turbidity

تتشكل العكارة في مياه الشرب عادة بسبب وجود جزيئات عالقة صغيرة جداً والمواد العضوية المنحلة وغير المنحلة في المياه، والتي يصعب رؤيتها بالعين المجردة. ويمكن أن تنشأ هذه الجزيئات والمواد العضوية من مصادر مختلفة مثل التربة والأوراق والنفايات الصناعية والمخلفات الحيوانية. تتأثر درجة العكارة بعدة عوامل، منها مصدر المياه، تدفق المياه، درجة الحرارة، مدة التخزين، طبيعة المواد العالقة في المياه، ومستوى التلوث البيولوجي والكيميائي.

تعتبر العكارة في مياه الشرب مؤشراً على وجود المواد العالقة والجراثيم في المياه، وقد تؤثر على الطعم والرائحة والمظهر العام للمياه، وقد تؤدي في بعض الحالات إلى مشاكل صحية. لذلك، يتم تحديد الحدود المسموح بها لتركيز العكارة في مياه الشرب وفقاً لمعايير الجودة العالمية والمحلية، الحد المسموح به وفق معايير منظمة الصحة العالمية هو NTU (5). حيث تعتمد درجة العكارة على حجم وعدد الجزيئات العالقة والمواد الصلبة الموجودة في المياه.

للتقليل من درجة العكارة في المياه يتم استخدام عدة طرق مثل الترشيح والتناضح العكسي وإزالة الجزيئات العالقة بواسطة إضافة مواد تساعد على تخثيرها وإزالتها.

تشير نتائج الدراسة خلال العامين (2021–2022) والمبينة في الشكل (9) إن قيمة تراكيز العكارة تراوحت ضمن المجال mg/l (0.38–9.94) وهي جميعها ضمن القيم المسموحة وفق منظمة الصحة العالمية باستثناء إحدى عشر قيمة كانت في الأشهر الماطرة.



الشكل (9): تغير قيم العكارة (Turbidity (NTU) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

1.8. الناقلية الكهربائية EC Electrical conductivity

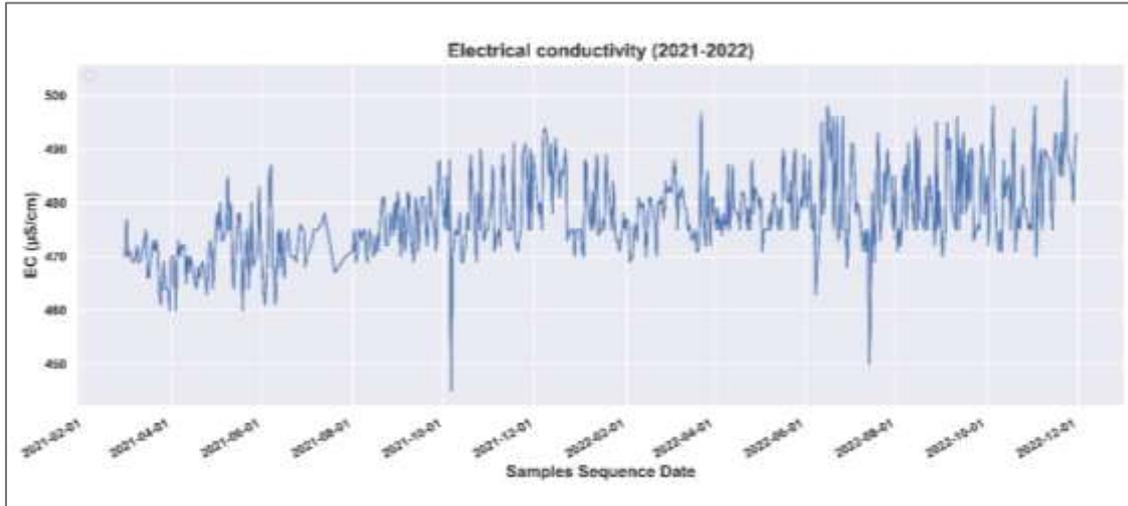
يعطي قياس الناقلية الكهربائية في مياه الشرب فكرة عن مقدار المواد المنحلة في المياه، وخاصة الأملاح المعدنية التي تذوب في المياه لتشكل أيونات، وتتأثر الناقلية الكهربائية بتركيز هذه الأيونات.

عندما تذوب المواد الكيميائية في المياه، يتم فصلها إلى جزيئات أو أيونات سالبة وكاتيونات موجبة، وتتحرك هذه الجزيئات في المياه لتحقيق التوازن الكهربائي.

وبالتالي، تزيد الناقلية الكهربائية في المياه كلما زاد تركيز الأملاح المذابة فيه. وتستخدم الناقلية الكهربائية في تقييم جودة مياه الشرب، حيث تشير القيم العالية للناقلية الكهربائية إلى وجود كميات كبيرة من الأملاح المعدنية في المياه والعكس صحيح، مما يشير إلى أن الماء غير صالح للشرب.

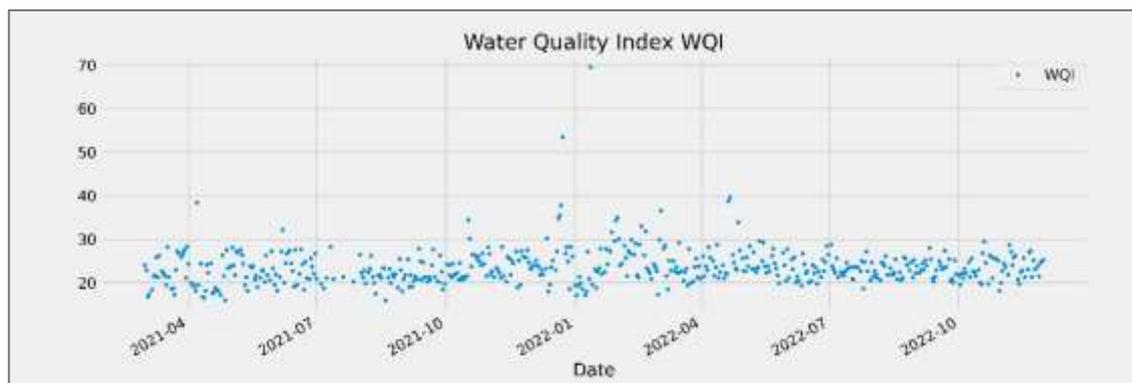
يتم قياس الناقلية الكهربائية باستخدام جهاز (Conductivity Meter)، وتعبّر القيمة الناتجة عنه عن عدد الأيونات الموجبة والسالبة في المياه. وتقاس القيمة بوحدة (Siemens)، والتي تشير إلى مقدار التيار الكهربائي المار عبر المياه بين قطبين في درجة حرارة وضغط معينين.

تشير نتائج الدراسة خلال العامين (2021-2022) والمبينة في الشكل (10) إن قيمة تراكيز الناقلية الكهربائية تراوحت ضمن المجال ($445-503$) $\mu\text{S/cm}$ وهي جميعها ضمن القيم المسموحة وفق منظمة الصحة العالمية وهو $1000 \mu\text{S/cm}$.



الشكل (10): تغير قيم شاردة الناقلية الكهربائية EC ($\mu\text{S/cm}$) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة.

يبين الشكل (11) نتائج تطبيق مؤشر جودة المياه WQI على مأخذ مياه شرب مدينة اللاذقية للفترة الزمنية (2022-2021) بشكل يومي، حيث تدل هذه النتائج على أن جودة مياه المأخذ تقع ضمن التصنيف الممتاز بنسبة (70.4%) وضمن التصنيف الجيد بنسبة (29.2%) في حين كان لدينا قيمتين فقط تقعان ضمن التصنيف الضعيف بنسبة (0.004%) .



الشكل (11): تغير قيم مؤشر جودة المياه (WQI) في مأخذ مياه شرب اللاذقية خلال فترة الدراسة (2021-2022).

2. الانحدار الخطي المتعدد (MLR) Multiple Linear Regression

من أجل معرفة العلاقة بين مؤشر جودة المياه WQI في مأخذ شرب مدينة اللاذقية والمتغيرات المفسرة المعتمدة في البحث (الأس الهيدروجيني، الكبريتات، النتريت، النتريت، الأمونيوم، الفوسفات، العكارة والناقلية الكهربائية)، تم استخدام نموذج الانحدار الخطي المتعدد المتغيرات جدول (5) والذي اعتبرت فيه المتغيرات السابقة كمتغيرات تفسيرية مستقلة ومتغير جودة المياه WQI كمتغير تابع.

الجدول (5): نتائج اختبار تحليل الانحدار الخطي المتعدد.

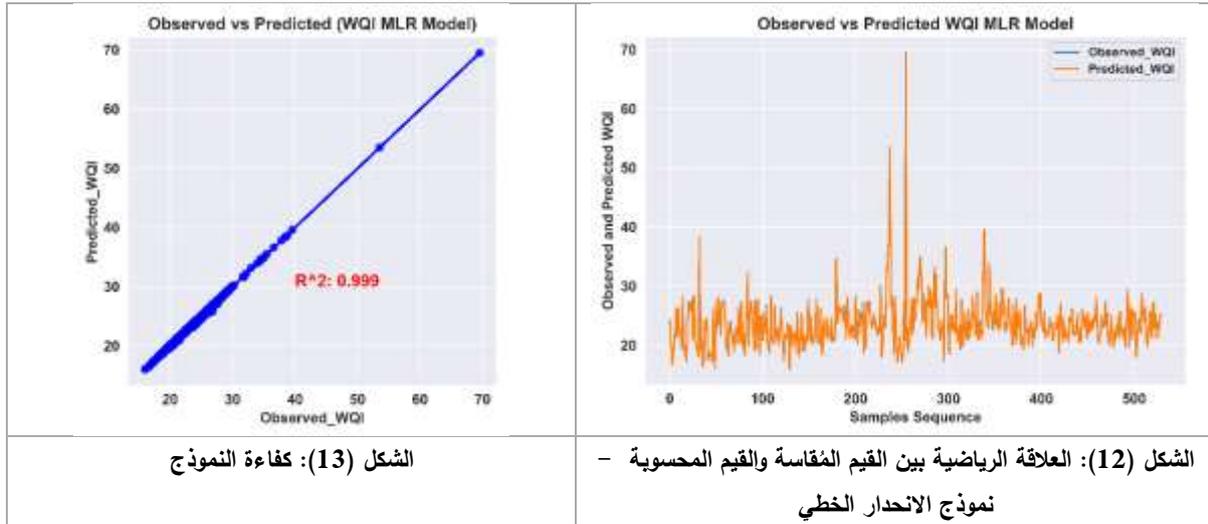
جودة النموذج		معامل تضم VIF	معامل التحديد R ²	معامل الارتباط R	قيمة Sig	قيمة t	قيمة B	نموذج الانحدار
قيمة Sig	قيمة F				0.000	-124.833	-56.751	الثابت (Constant)
0.00	69855.695	1.057	0.999	0.999	0.000	217.177	8.227	الأس الهيدروجيني pH
		1.244			0.000	37.335	0.304	النترات NO ₃ ⁻
		1.239			0.000	289.537	15.167	النتريت NO ₂ ⁻
		1.024			0.000	337.218	24.370	الفوسفات PO ₄ ⁻
		1.037			0.000	365.470	2.428	العكارة Turbidity
		1.070			0.000	13.183	0.011	الناقلية الكهربائية EC

أظهرت نتائج نموذج الانحدار في الجدول (5) أن نموذج الانحدار معنوي وذلك من خلال قيمة الاختبار F والبالغة (69855.695) بدلالة إحصائية (0.00) أصغر من مستوى المعنوية (0.01)، وهذا يدل على جودة نموذج العلاقة وصحة الاعتماد على نتائج النموذج بدون أخطاء، كما وتدل النتائج على أن المتغيرات المفسرة المستقلة تفسر 99 % من التباين الحاصل في مؤشر جودة المياه WQI وذلك بالنظر إلى معامل التحديد (R²). كما جاءت قيم بيتا B التي توضح العلاقة بين WQI والمتغيرات التفسيرية جميعها ذات دلالة إحصائية حيث يمكن استنتاج ذلك من خلال قيم t والدوال المرتبطة بها، فمثلاً كانت قيمة B التي توضح العلاقة بين WQI والعكارة بمقدار (2.43) وهي ذات دلالة إحصائية (0.00)، ويعني ذلك أنه كلما تحسنت قيمة العكارة بمقدار وحدة فإنها تحسن مستوى مؤشر جودة المياه WQI بمقدار (2.43) وحدة، وهكذا بالنسبة لبقية المتغيرات. كما يوضح الجدول (5) نتائج اختبار التعددية الخطية

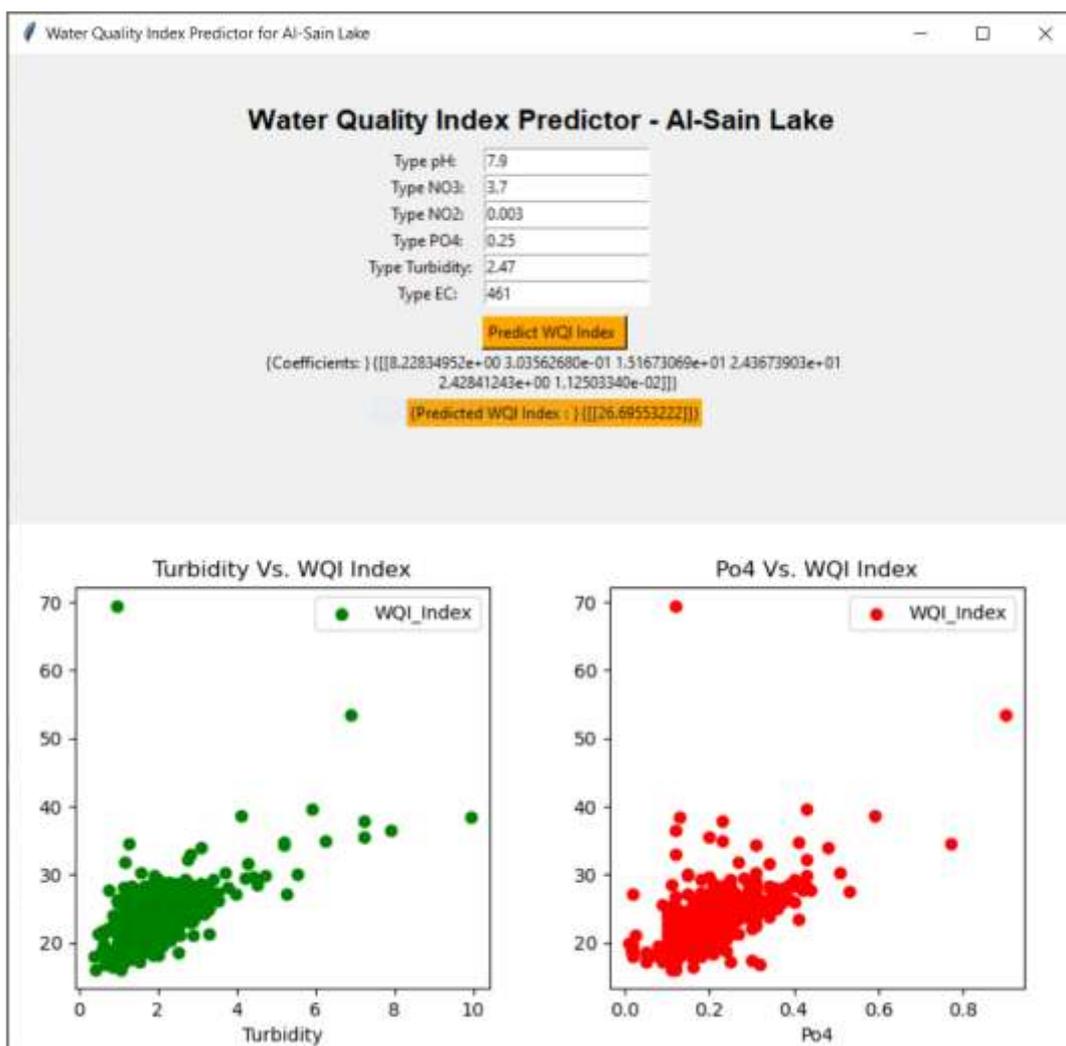
حيث بينت النتيجة أن قيم جميع عوامل تضخم التباين للنموذج كانت أصغر من ثلاثة ($VIF < 3$) مما يشير إلى عدم وجود مشكلة تعددية خطية بين متغيرات النموذج، بالتالي يمكننا صياغة معادلة الانحدار التي تتكون من قيم بيتا (B) لكل من المتغيرات المستقلة السابقة وقيمة الثابت Constant على النحو التالي:

$$WQI = -56.751 + 8.227 (pH) + 0.304 (NO_3) + 15.167 (NO_2) + 24.370 (PO_4) + 2.428 (Turbidity) + 0.011 (EC) \quad Eq.6$$

يبين الشكل (12) العلاقة الرياضية بين القيم المقاسة والقيم المُقدَّرة من خلال تطبيق النموذج على مصفوفة المعطيات. كما يبين الشكل (13) قوة أداء النموذج من خلال قيمة R^2 معامل التحديد.



بالاعتماد على نموذج الانحدار الخطي متعدد المتغيرات (MLR) ذو الأداء القوي جداً قمنا ببرمجة وتصميم أداة بواجهة بيانية (GUI) لحساب مؤشر جودة المياه من أجل استخدامها من قبل صانعي القرار والمعنيين بمراقبة والتحكم بجودة المياه بشكل بسيط وعملي، حيث يتم ادخال قيم متغيرات النموذج الناتجة عن التحاليل المخبرية ومباشرةً يتم تحديد مؤشر جودة مياه الشرب للمأخذ المدروس في بحيرة السن الشكل (14).



الشكل (14) واجهة برنامج حساب مؤشر جودة مياه شرب مأخذ اللاذقية في بحيرة السن

3. نماذج تعلم الآلة (ML) Machine Learning Models

تعلم الآلة هو فرع من فروع الذكاء الاصطناعي (AI) يهدف إلى إنشاء نماذج وتطبيقات تستطيع الآلة من خلالها التعلم والتكيف مع البيانات بدلاً من أن يتم برمجتها بشكل صريح. وبمعنى آخر، يتعلم الحاسوب من البيانات دون الحاجة إلى وضع قواعد واضحة للتعرف عليها. تشمل تقنيات تعلم الآلة العديد من النماذج المختلفة، مثل الشبكات العصبونية الصناعية (ANN)، والتحليل العائلي الرئيسي (PCA)، والغابات العشوائية (RF)، وغيرها الكثير. تتباين هذه النماذج في الطريقة التي تتعلم بها الآلة وكيفية تطبيق التعلم الذاتي على البيانات. يتم استخدام تعلم الآلة في العديد من المجالات، بما في ذلك تحليل البيانات، والتعرف على الصوت والصورة، والتنبؤ والتصنيف، والتحكم في الروبوتات، وتحسين الأداء في الألعاب، وغيرها. ويتم استخدامها في مجال إدارة جودة المياه أيضاً، حيث يتم استخدام تقنيات تعلم الآلة لتحليل البيانات في الوقت الحقيقي وتحسين جودة المياه والتحكم فيها.

تتطلب تقنيات تعلم الآلة المهارات الرياضية والبرمجية والإحصائية، بالإضافة إلى فهم عميق للمجال الذي يتم تطبيقها فيه. لذلك، فإن استخدام تقنيات تعلم الآلة قد يتطلب فريقاً متخصصاً في مجال إدارة جودة المياه وتقنيات تعلم الآلة لتنفيذها بنجاح.

تعد نماذج تعلم الآلة واحدة من الأدوات الحديثة التي يمكن استخدامها في نمذجة جودة مياه الشرب. تستخدم هذه النماذج للتحليل والتنبؤ بجودة المياه الحالية والمستقبلية الناتجة عن عدد من العوامل المؤثرة، مثل درجة الحموضة والعكارة والأملاح والعناصر الغذائية والملوثات وغيرها.

وتعتبر نماذج تعلم الآلة مفيدة في جودة مياه الشرب لأنها تتيح التنبؤ بالمخاطر المحتملة وتوفير الوقت والجهد في التحليل اليدوي للبيانات. ويمكن استخدامها كأداة مفيدة في عمليات إدارة جودة المياه لتحسينها والحد من المخاطر الصحية. كخطوة أولى مهمة لبناء نموذج ML، يلزم اختيار عدد وافٍ من متغيرات الإدخال، والتي تحتوي على معلومات أساسية كافية للتنبؤ بـ WQI. بالإضافة إلى ذلك، يمكن أن يؤدي هذا الاختيار إلى تحسين دقة النموذج من خلال تجنب التأثير غير المرغوب فيه على الأداء التنبؤي. في الدراسة الحالية، تم تحديد ثمانية متغيرات لجودة المياه كمداخلات محتملة. هناك العديد من الطرق الحالية لتقييم مجموعات المدخلات، بما في ذلك الارتباط الذاتي، والارتباط الذاتي الجزئي، والارتباط المتبادل، ومعامل الارتباط. من هذه الطرق، تم اختيار معامل الارتباط للدراسة الحالية لما له من الدقة والكفاءة [4]. توضح مصفوفة الارتباط الشكل (2) أن متغير WQI يرتبط وفق أعلى قيمة لـ R بالعكارة (Turbidity)، متبوعاً بـ PO_4^{-3} ، NO_2^- ، pH، NO_3^- ، EC على التوالي. حيث يمكننا إهمال متغيري NH_4^+ و SO_4^{-2} وذلك لارتباطها الضعيف مع WQI.

بعد اختيار متغيرات WQI المدخلة، تم بناء ومقارنة 19 نموذجاً من نماذج ML للتنبؤ بمؤشر جودة المياه والموضحة في الجدول رقم (6)، حيث تم تقسيم قاعدة البيانات (70%) من أجل عملية التدريب و (30%) من أجل الاختبار، وتم تقييم أداء هذه النماذج أثناء عملية الاختبار [21]. تم في هذه الدراسة استخدام حزمتي Scikit-Learn و PyCaret، وهما حزمتان مبنيتان على بيئة البرمجة باستخدام Python، لتطوير تسعة عشر نموذجاً لتعلم الآلة ML من أجل التنبؤ بمؤشر WQI. وتم استخدام جذر متوسط الخطأ التريبيعي (RMSE) ومعامل التحديد (R^2) كإحصاءات كفاءة نموذجية، لتقييم أداء نماذج ML في هذه الدراسة وقياس جودة التوافق بين القيم المتنبأة والقيم المقاسة. يقيس RMSE الانحراف بين القيم المرصودة والمتوقعة، بينما يقيس R^2 درجة الارتباط بين البيانات المرصودة والمتوقعة.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad \text{Eq.7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \text{Eq.8}$$

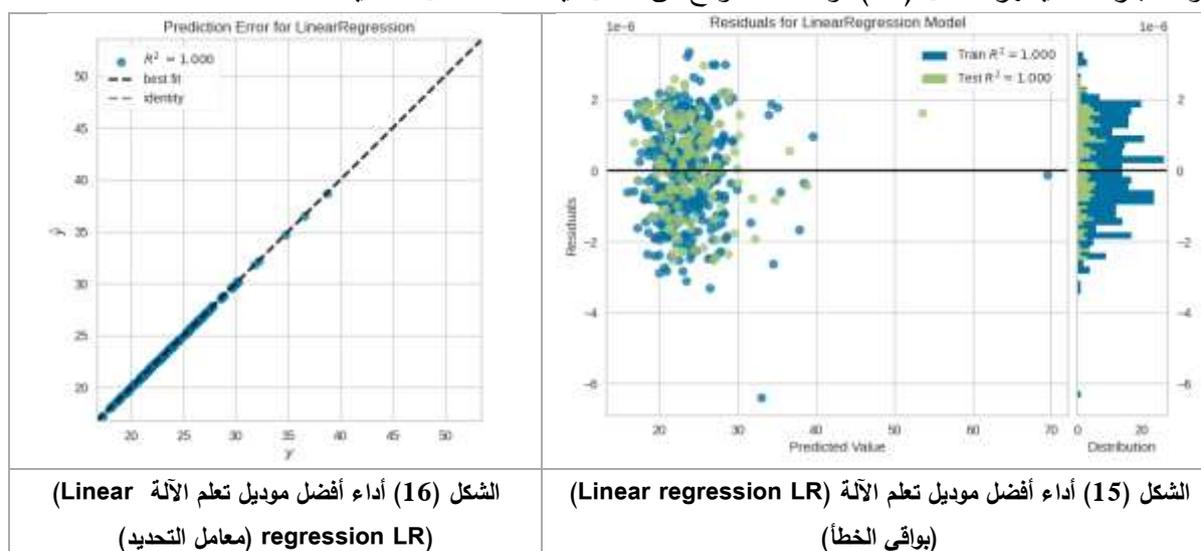
حيث n هو العدد الإجمالي للقيم المتوقعة، O_i هي القيمة المرصودة، \bar{O} هي متوسط القيم المرصودة، و P_i هي القيمة المتوقعة.

أظهرت نتائج نماذج التعلم الآلي من أجل التنبؤ بجودة مياه شرب مأخذ اللاذقية الموضحة في الجدول (6) وذلك بناء على متغيرات التلوث المدروسة خلال الفترة (2021-2022)، أن نماذج ML الثلاثة: (Linear regression LR, Least angle regression LAR, Bayesian ridge BR) حققت أفضل أداء بقيمة ارتباط بين القيم المقاسة والمتوقعة (100%).

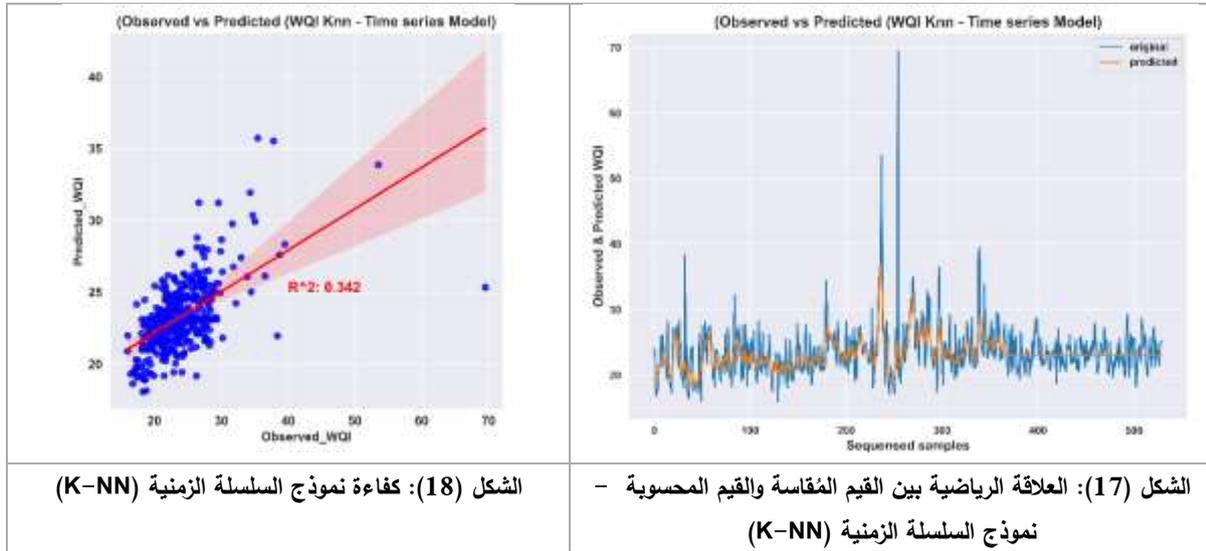
الجدول (6) مقارنة أداء نماذج تعلم الآلة ML في حساب مؤشر جودة مياه شرب مأخذ اللاذقية

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	
lr	Linear Regression	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
lar	Least Angle Regression	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
br	Bayesian Ridge	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
ridge	Ridge Regression	0.6266	5.7816	1.3336	0.8822	0.0406	0.0235
et	Extra Trees Regressor	0.6691	7.7039	1.6521	0.8227	0.0514	0.0240
gbr	Gradient Boosting Regressor	0.6713	7.7261	1.6444	0.8216	0.0505	0.0240
xgboost	Extreme Gradient Boosting	0.6906	7.8954	1.7136	0.8092	0.0531	0.0249
rf	Random Forest Regressor	0.8350	8.2094	1.8274	0.7900	0.0589	0.0312
huber	Huber Regressor	1.2459	10.5125	2.1805	0.6997	0.0772	0.0492
dt	Decision Tree Regressor	1.2773	9.9978	2.3195	0.6650	0.0782	0.0501
lightgbm	Light Gradient Boosting Machine	1.0778	9.8756	2.3649	0.6560	0.0762	0.0403
ada	AdaBoost Regressor	1.4226	10.1218	2.4847	0.6166	0.0871	0.0579
lasso	Lasso Regression	2.2606	13.7407	3.3067	0.2884	0.1212	0.0941
llar	Lasso Least Angle Regression	2.2606	13.7407	3.3067	0.2884	0.1212	0.0941
en	Elastic Net	2.3091	13.9869	3.3621	0.2580	0.1231	0.0961
knn	K Neighbors Regressor	2.4627	14.7339	3.5339	0.1645	0.1295	0.1011
omp	Orthogonal Matching Pursuit	2.7662	17.7386	3.9547	-0.0604	0.1456	0.1154
dummy	Dummy Regressor	2.7546	17.9070	3.9646	-0.0620	0.1462	0.1152
par	Passive Aggressive Regressor	3.6783	28.3909	5.0279	-0.9922	0.1892	0.1473

يبين الشكل (15) أداء نموذج تعلم الآلة الأمثل في حالة الدراسة (Linear regression LR) من خلال بواقي الخطأ وتوزيعه بين القيم المقاسة والقيم المقدرة من خلال تطبيق هذا النموذج على مصفوفة المعطيات الخاصة بمرحلة التدريب والاختبار. كما يظهر الشكل (16) قوة أداء النموذج من خلال قيمة R^2 معامل التحديد.



تستطيع نماذج تعلم الآلة (ML) التعامل مع تحليل البيانات المتسلسلة (Time Serie) بشكل فعال، وبالتالي تشكل خياراً جيداً لتحليل هذه البيانات. يعود هذا إلى قدرتها على التعامل مع البيانات المرتبطة بالترتيب الزمني وتحليلها بدقة وفعالية. يتم استخدامها في تحليل البيانات الزمنية عادة لتوقع القيم المستقبلية أو لتحليل السلوك الزمني للبيانات. ويمكن استخدامها لتحليل الأنماط الزمنية في الإنتاجية أو المبيعات أو البيئة أو أي نوع آخر من البيانات الزمنية. يتم تدريب نموذج (XGBoost) مثلاً باستخدام البيانات الزمنية المتاحة والمحافظة في ترتيب زمني. وبعد التدريب، يمكن استخدام النموذج لتوقع القيم المستقبلية بناءً على البيانات التاريخية المتاحة. يمكن استخدام (K-Neighbors regressor) أيضاً في تحليل البيانات الزمنية بشكل فعال، ولكن يتطلب الأمر فهماً جيداً للبيانات الزمنية وأساليب التحليل الإحصائي وتقنيات تعلم الآلة. ومن المهم أيضاً التحقق من جودة البيانات والتأكد من وجود بيانات كافية للتدريب والتحليل. تندرج هذه الخوارزمية ضمن خوارزميات التعلم الآلي على أساس تقنية التعلم تحت الاشراف (Supervised learning technique) وتستخدم لحل مسائل الانحدار وكذلك التصنيف، وهي خوارزمية غير حدودية (غير بارامترية) (non-parametrics algorithm) مما يعني أنها لا تجعل أي افتراض على البيانات الأساسية وتسعى إلى ملاءمة بيانات التدريب بشكل أفضل في بناء الدالة، مع الحفاظ على بعض القدرة على التعميم على البيانات غير المرئية، فهي قادرة على ملاءمة عدد كبير من الأشكال الوظيفية (الدوال). لذلك قمنا بتطبيق نماذج التعلم الآلي التسع عشرة على السلسلة الزمنية المتوفرة لدينا خلال فترة الدراسة (2022-2021) وتم التحقق من قدرة هذه النماذج (ML) في التنبؤ بجودة مياه شرب مأخذ اللادقية، وبغض النظر عن قيم متغيرات التلوث (لم تدخل في بناء النماذج)، حقق النموذج (K Neighbors regressor K-NN) أفضل أداء بقيمة ارتباط بين القيم المقاسة والمتوقعة لمؤشر جودة المياه (% 34)، وقيمة خطأ (RMSE) مقدارها (4.62) الشكل (17). كما يبين الشكل (18) قوة أداء النموذج من خلال قيمة R^2 معامل التحديد.



الاستنتاجات والتوصيات:**الاستنتاجات:**

من خلال حساب مؤشر جودة المياه WQI لمأخذ مياه شرب مدينة اللاذقية في بحيرة السن لفترة الدراسة (2022-2021) وتقييم بارامترات التلوث المعتمدة في هذه الدراسة: العكارة (Turbidity)، الفوسفات PO_4^{-3} ، النترت NO_2^- ، الأس الهيدروجيني pH، النترات NO_3^- ، الناقلية الكهربائية EC، الأمونيوم NH_4^+ والكبريتات SO_4^{-2} على جودة مياه البحيرة، يتضح ما يلي:

1- مقارنة القيم المقاسة لهذه البارامترات مخبرياً مع القيم المسموحة وفق معايير منظمة الصحة العالمية (WHO) والمواصفات القياسية السورية فقد كانت جميعها ضمن الحدود المسموحة باستثناء بعض القيم القليلة جداً والتي ليس لها تأثير مباشر على مؤشر الجودة.

2- تم تصنيف جودة مياه شرب مأخذ مدينة اللاذقية بعد حساب مؤشر الجودة بالطريقة الحسابية للوزن المرجح وكانت ضمن التصنيف الممتاز بنسبة (70%) والجيد بنسبة (30%) تقريباً.

3- تم تطبيق نموذج الانحدار الخطي متعدد البارامترات من أجل التنبؤ بمؤشر جودة مياه المأخذ (WQI). وحقق علاقة ارتباط قوية جداً، وفسرت عوامل التلوث المستقلة الداخلة في النموذج التغيرات الحاصلة في مؤشر جودة مياه المأخذ في البحيرة بنسبة 99%.

4- بالاعتماد على نموذج الانحدار الخطي متعدد المتغيرات (MLR) وعلى الأداة البرمجية ذات الواجهة البيانية (GUI) التي قمنا ببرمجتها وتصميمها، يمكن لصانعي القرار والقائمين على المراقبة والتحكم بجودة المياه استخدامها بشكل بسيط وعملي، حيث يتم ادخال قيم متغيرات التلوث المعتمدة وبشكل مباشر نصل إلى تحديد مؤشر جودة مياه الشرب للمأخذ.

5- تم التحقق من قدرة 19 نموذج من نماذج التعلم الآلي (ML) من أجل التنبؤ بجودة مياه شرب مأخذ اللاذقية وذلك بناء على متغيرات التلوث المدروسة خلال الفترة (2021-2022)، حيث حققت النماذج الثلاثة (Linear regression LR, Least angle regression LAR, Bayesian ridge BR) أفضل أداء بقيمة ارتباط بين القيم المقاسة والمتوقعة (100%).

5- تم التحقق من قدرة 19 نموذج من نماذج التعلم الآلي (ML) من أجل التنبؤ بجودة مياه شرب مأخذ اللاذقية وذلك بناء على السلسلة الزمنية المدروسة (2021-2022)، وبغض النظر عن قيم بارامترات التلوث، حيث حقق النموذج (K-Neighbors regressor K-NN) أفضل أداء بقيمة ارتباط بين القيم المقاسة والمتوقعة وفق معامل التحديد (34%)، وخطأ (RMSE) مقداره (4.62).

التوصيات:

- إمكانية استخدام نماذج الانحدار الخطي وتعلم الآلة (ML) المختلفة في تحليل جودة المياه في البلدان النامية، ولكن يتطلب الأمر التغلب على التحديات المحتملة وتوفير الموارد والبيانات الكافية والموثوقة اللازمة لتحليلها بدقة وفعالية.
- ضرورة إنشاء قواعد بيانات خاصة بتسجيل نتائج التجارب المخبرية لبارامترات التلوث المختلفة وفق البرنامج المعتمد من قبل المعنيين بمراقبة وحماية البحيرة، من أجل توظيفها في الأبحاث المستقبلية.
- ضرورة متابعة برنامج المراقبة الدورية وعدم الانقطاع في قياس بارامترات التلوث، والعمل على زيادة عدد هذه البارامترات نظراً للأهمية البالغة لهذا المصدر المائي (بحيرة السن).

- إمكانية تطبيق خوارزميات تعلم آلي أخرى مثل: Support Vector Machines (SVM) وخوارزميات الشبكات العصبونية الصناعية (ANN), حيث يتم استخدامهما لمعالجة القضايا التصنيفية والانحدارية، وكذلك في التقييم والتنبؤ بجودة المياه والتعامل مع العلاقات المعقدة بين المتغيرات.

References:

- 1 Organization, W.H., *Guidelines for drinking-water quality: first addendum to the fourth edition*. 2017.
- 2 UNEP, A., *A snapshot of the world's water quality: towards a global assessment*. Nairobi, United Nations Environment Programme, 2016.
- 3 Nouraki, A et al., *Prediction of water quality parameters using machine learning models: a case study of the Karun River, Iran*. Environmental Science and Pollution Research, 2021. **28**(40): p. 57060-57072.
- 4 Asadollah, S.B.H.S., et al., *River water quality index prediction and uncertainty analysis: A comparative study of machine learning models*. Journal of environmental chemical engineering, 2021. **9**(1): p. 104599.
- 5 Mishra, B.K., et al., *Assessment of Bagmati river pollution in Kathmandu Valley: Scenario-based modeling and analysis for sustainable urban development*. Sustainability of Water Quality and Ecology, 2017. **9**: p. 67-77.
- 6 Ewaid, S.H. and S.A. Abed, *Water quality index for Al-Gharraf river, southern Iraq*. The Egyptian Journal of Aquatic Research, 2017. **4**:(2)3p. 117-122.
- 7 Ramakrishnaiah, C., C. Sadashivaiah, and G. Ranganna, *Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India*. E-Journal of chemistry, 2009. **6**(2): p. 523-530.
- 8 Jafar, R., *Assessment of surface water quality by using multivariate statistical techniques*. Tishreen University Journal -Engineering Sciences Series, 2022. **44**(4): p. 11-31.
- 9 Abbasi, T. and S.A. Abbasi, *Water quality indices*. 2012: Elsevier.
- 10 Tung, T.M. and Z.M. Yaseen, *A survey on river water quality modelling using artificial intelligence models: 2000–2020*. Journal of Hydrology, 2020. **585**: p. 124670.
- 11 Nearing, G.S., et al., *What role does hydrological science play in the age of machine learning?* Water Resources Research, 2021. **5**:(3)7p. e2020WR028091.
- 12 Bedi, S., et al., *Comparative evaluation of machine learning models for groundwater quality assessment*. Environmental Monitoring and Assessment, 2020. **192**: p. 1-23.
- 13 Ahmed, M., R. Mumtaz, and S.M. Hassan Zaidi, *Analysis of water quality indices and machine learning techniques for rating water pollution: A case study of Rawal Dam, Pakistan*. Water Supply, 2021. **21**(6): p. 3225-3250.
- 14 Ewaid, S.H. and S.A. Abed, *Water quality assessment of Al-Gharraf River, South of Iraq using multivariate statistical techniques*. Al-Nahrain Journal of Science, 2017. **20**(2): p. 114-122.
- 15 Sharma, L.K., *Estimation of Ground Water Quality Index for Drinking Purposes in Ambah, India*. International Journal of Prevention and Control of Industrial Pollution, 2021. **7**(1): p. 42-48.
- 16 Jafar, R., *Application of the Water Quality Index (NSFWQI) on the Al-Sain Lake*. Tishreen University Journal-Engineering Sciences Series, 2016. **38**(4)
- 17 Awad, A., R. Jafar, and A. Osekrieh, *Application of Non-Linear Autoregressive Neural Networks for Predicting the Drinking Water Turbidity of Al-Sin Water Purification Plant*. Tishreen University Journal -Engineering Sciences Series, 2018. **39**(6)

- .18 Yadav, A.K., P. Khan, and S.K. Sharma, *Water Quality Index Assessment of Groundwater in Todaraisingh Tehsil of Rajasthan State, India-A Greener Approach*. E-journal of Chemistry, 2010. **7**(S1): p. S428-S432.
- .19 Pooja, D., et al., *Sensors in water pollutants monitoring: role of material*. 2020: Springer.
- .20 Singh, A.L., et al. *Nitrate and phosphate contamination in ground water of Varanasi, Uttar Pradesh, India*. Journal of Industrial Research & Technology, 2012. **2**(1): p. 26-32.
- .21 Bui, D.T., et al., *Improving prediction of water quality indices using novel hybrid machine-learning algorithms*. Science of the Total Environment, 2020. **721**: p. 137612.