

Research Paper

Assessment and Prediction of Water Quality of Talar River Using Deep Learning Models and Statistical Analyses

Mina Ahmadi Sarkhooni¹,Fatemeh Shokrian^{2*},Karim Solaimani³,Seyyed
Banihashemi⁴Masoumeh
Ataollah Kavian⁵

¹ Ph.D. Student, Department of Sciences and Watershed Management Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. minaahmadi5462@gmail.com

² *Associate Professor, Department of Science and Watershed Management Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. *(Corresponding Author: shokrian.f@gmail.com f.shokrian@sanru.ac.ir)

³ Professor, Environmental Remote Sensing & GIS Research Institute, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. solaimani2001@yahoo.co.uk ka.solaimani@sanru.ac.ir

⁴ Ph.D Caspian Sea National Research Centre, Water Research Institute, Tehran, Iran, mbanihashemi@hotmail.com

⁵ Professor, Department of Sciences and Watershed Management Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. ataollah.kavian@gmail.com



10.22125/iwe.2025.510360.1868

Received:

March 9, 2025

Accepted:

May 31, 2025

Available online:

August 4, 2025

Keywords: Taylor plot, Principal Component Analysis (PCA), Pearson correlation coefficient, Heavy metals

Abstract

This study was conducted to assess the water quality of the Talar River during both low-flow and high-flow periods. Through field survey and consideration of natural and anthropogenic factors, 15 sampling stations were established across various land uses. Water sampling was conducted during both the low-flow period (September) and the high-flow period (May). For sampling, half-litre polyethylene bottles were used, and samples were collected from a depth of 15 to 30 cm below the water surface. Various statistical methods were employed to analyse the data. Pearson's correlation coefficient was used to identify relationships between physical and chemical parameters, and Principal Component Analysis (PCA) was applied to reduce data dimensionality. Subsequently, calcium concentration was predicted using deep learning models, namely LSTM and CNN, with inputs such as river discharge, maximum temperature, precipitation, evapotranspiration, and runoff at the Shirgah station. The analysis results revealed that electrical conductivity (EC) and total dissolved solids (TDS) exhibited strong positive correlations with dissolved ions such as calcium (Ca), magnesium (Mg), and sodium (Na). Additionally, pH showed a negative correlation with certain heavy metals like aluminium (Al) and chromium (Cr), which is linked to changes in the solubility of these metals under acidic conditions. PCA indicated that during the low-flow period, two principal components explained over 90% of the data variance. The first component was primarily influenced by dissolved ions such as sodium, chloride, and sulfur, while the second component was associated with parameters like electrical conductivity and turbidity. During the high-flow period, the first component, accounting for 95.69% of the variance, played a dominant role in data variability and was mainly determined by TDS and EC. The evaluation results of calcium concentration modelling using deep learning models, based on the Taylor diagram, showed that at the Shirgah station, the CNN model had a better correlation (close to 0.8) with observational data compared to the LSTM model (approximately 0.6). These

* **Corresponding Author: Fatemeh Shokrian**

Address: Department of Sciences and Watershed Management Engineering, Sari Agricultural Sciences and Natural Resources University, Sari, Iran.

Email: f.shokrian@sanru.ac.ir

Tel: 09113513923

findings demonstrated that geological and hydrological factors have varying effects on water quality across different periods.

1. Introduction

Rivers are vital ecosystems that play crucial roles in maintaining biodiversity and providing water for human use. However, they are increasingly exposed to pollution from natural and anthropogenic sources such as land-use changes, industrial activities, and urbanisation. These changes negatively affect water quality, aquatic ecosystems, and public health, highlighting the need for effective water resource management. Rapid population growth and technological advancements have intensified land use, leading to severe water pollution that threatens both ecosystems and human development. Heavy metals and chemical pollutants from industrial and agricultural activities exacerbate water quality issues, posing risks to aquatic life and human health. Advanced statistical methods, such as the Pearson correlation and Principal Component Analysis (PCA), are essential tools for analysing water quality parameters, identifying pollution sources, and predicting changes in response to climatic and human activities. In addition, the use of advanced deep learning models, such as LSTM and CNN, in predicting water chemical parameters is important because of their ability to process temporal data and identify complex patterns. LSTM can identify long-term relationships between variables and predict changes in the calcium concentration with greater accuracy. In contrast, CNN, with its ability to extract features from data, identifies local and complex patterns. The use of these two models provides a significant improvement in the prediction accuracy and understanding of the factors affecting water quality and changes in water chemical parameters. This study focuses on the Talar River, using these methods to assess water quality, understand the dynamics of heavy metals, and provide insights for sustainable water management and ecosystem preservation.

2. Materials and Methods

In this study, to determine the sampling stations, topographic maps and access routes to the river were first examined. Following this, a preliminary field visit was conducted in the area, and suitable points for sampling were identified by considering land use patterns, natural features, and access methods to the river and its tributaries. These points were identified using land use maps and Google Earth software during the field visit, and the river water sampling process was carried out by standard sampling protocols. After mapping the sampling points, water samples were collected from 15 locations with different land uses, with two sampling rounds during the low-flow and high-flow periods. Sampling was performed using 500 ml polyethylene bottles at depths of 15 to 30 cm below the water surface. Two water samples were collected from each location: one for measuring phosphate and nitrate concentrations, and the other for analysing heavy metal concentrations. Additionally, on-site measurements of pH, electrical conductivity, and temperature were taken using portable devices, and the samples were stored at 4°C until transfer to the laboratory. The low-flow (September) and high-flow (May) periods of the Talar River were determined based on previous studies and the river's long-term discharge data. Sampling containers were initially washed with 30% nitric acid and then rinsed with distilled water. After sampling, to prevent air infiltration, the container lids were sealed with adhesive tape. Finally, samples for phosphate and nitrate analysis were sent to the central laboratory of Sari University of Agricultural Sciences and Natural Resources, while samples for heavy metal analysis were sent to the Zar Azma Laboratory in Tehran. In this study, statistical methods such as Pearson correlation coefficient and Principal Component Analysis (PCA) were used to enhance data analysis and investigate the effects of natural and anthropogenic factors on the water quality of the Talar River.

3. Results

The Pearson correlation analysis of the physical and chemical parameters of the Talar River water during the low and high flow periods revealed significant patterns in their relationships. During the low-flow period, strong positive correlations were observed between calcium (Ca) and magnesium (Mg),

sodium (Na) and chloride (Cl), and electrical conductivity (EC) and total dissolved solids (TDS), highlighting their roles in water hardness, salinity, and ion concentration. Aluminium (Al) showed weak correlations, while barium (Ba) and potassium (K) exhibited strong relationships with other elements, reflecting shared chemical origins and agricultural influences. The pH was positively correlated with nitrate but negatively correlated with many elements, indicating better solubility under acidic conditions. In the high-flow period, similar patterns were observed, with Ca and Mg maintaining strong positive correlations and Na and Cl showing significant relationships, suggesting the presence of sodium chloride compounds. Turbidity was weakly correlated with heavy metals. Heatmap visualisations further confirmed these relationships, with EC and TDS consistently showing strong positive correlations, underscoring their importance as key indicators of water quality. Overall, the analysis highlighted the significant influence of geological processes and human activities on river water quality, with EC, TDS, and turbidity emerging as critical parameters for monitoring and managing the water resources of the Talar River. PCA revealed that the first two principal components (Dim1 and Dim2) explained over 90% of the variance in the low-flow period, with sulfur (S), chloride (Cl), and sodium (Na) being the most influential variables. In the high-flow period, PCA showed that the first principal component (PC1) explained 95.69% of the variance, dominated by EC and TDS, while the second component (PC2) explained 3.66%, primarily associated with turbidity. Overall, the study identified EC, TDS, and turbidity as critical parameters for assessing the water quality of the PCA and effectively captured the dominant factors influencing its characteristics during different hydrological periods. The results of calcium prediction with deep learning models showed that CNN has a better correlation (close to 0.8) than LSTM (approximately 0.6). RMS error is the radial distance of each point from the centre of the graph, indicating the amount of RMS error. CNN was closer to the centre and had a lower RMS error than the LSTM. The LSTM model performed worse than the CNN model in predicting the calcium parameter because its standard deviation was further away from the real data, and its RMS error was larger.

4. Discussion and Conclusion

Pearson correlation analysis of the water quality parameters of the Talar River revealed strong positive correlations between calcium (Ca) and magnesium (Mg), sodium (Na), chloride (Cl), electrical conductivity (EC), and total dissolved solids (TDS), emphasising their roles in water hardness, salinity, and ion concentration. PCA results showed that the first two principal components explained over 90% of the variance, with EC, TDS, and turbidity emerging as key indicators of water quality. At Shirgah station, the CNN model outperformed the LSTM model in estimating the calcium concentration. This superiority is due to the specific characteristics of the data and environmental conditions, where the CNN model is well able to identify complex patterns and trends. These findings underscore the significant impact of geological processes and human activities on the river's water quality, providing a foundation for effective water resource management and conservation strategies.

5. Six important references

- 1) Al-Sulttani, A. O., M. Al-Mukhtar, A. B. Roomi, A. A. Farooque, K. M. Khedher and Z. M. Yaseen. 2021. Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access*, 9: 108527-108541 .
- 2) Arslan, O. 2013. Spatially weighted principal component analysis (PCA) method for water quality analysis. *Water Resources*, 40: 315-324 .
- 3) Bu, H., W. Meng, Y. Zhang and J. Wan. 2014. Relationships between land use patterns and water quality in the Taizi River basin, China. *Ecological Indicators*, 41: 187-197 .
- 4) Chung, S. Y., S. Venkatramanan, N. Park, T. Ramkumar, S. B. Sujitha and M. P. Jonathan. 2016. Evaluation of physico-chemical parameters in water and total heavy metals in sediments at Nakdong River Basin, Korea. *Environmental Earth Sciences*, 75: 1-12 .
- 5) Elsayed, S., H. Hussein, F. S. Moghanm, K. M. Khedher, E. M. Eid and M. Gad. 2020. Application of irrigation water quality indices and multivariate statistical techniques for surface water quality assessments in the Northern Nile Delta, Egypt. *Water*, 12(12): 3300 .

- 6) Ibrahim, T. N. B. T., F. Othman, N. Z. Mahmood and T. Abunama. 2021. Seasonal effects on spatial variations of surface water quality in a tropical river receiving anthropogenic influences. *Sains Malaysiana*, 50(3): 571-593 .

Conflict of Interest

The authors declared no conflict of interest.

Acknowledgement: The present research is extracted from a Ph.D. thesis; so, the authors express their appreciation and gratitude for the financial support of Sari Agricultural Sciences and Natural Resources University, Iran.

ارزیابی و پیش‌بینی کیفیت آب رودخانه تالار با استفاده از مدل‌های یادگیری عمیق و تحلیل‌های آماری

مینا احمدی سرخونی^۱، فاطمه شکریان^{۲*}، کریم سلیمانی^۳، سیده معصومه بنی‌هاشمی^۴، عطااله کاویان^۵

تاریخ ارسال: ۱۴۰۳/۱۲/۱۹

تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰

مقاله از رساله دکترا

چکیده

مدیریت پایدار منابع آب سطحی نیازمند پایش مستمر و تحلیل دقیق کیفیت آب در بازه‌های زمانی مختلف است. این پژوهش با هدف ارزیابی کیفیت آب رودخانه تالار در دوره‌های کم‌آبی (شهریور) و پرآبی (اردیبهشت) انجام شد. طی بازدیدهای میدانی، ۱۵ نقطه نمونه‌برداری در کاربری‌های مختلف تعیین گردید. نمونه‌ها از عمق ۱۵ تا ۳۰ سانتی‌متری زیر سطح آب، در دو نوبت برداشت شدند. نتایج نشان داد در دوره پرآبی، هدایت الکتریکی (EC) و کل جامدات محلول (TDS) با یون‌های محلول مانند کلسیم (Ca)، منیزیم (Mg) و سدیم (Na) همبستگی مثبت قوی دارند ($r > 0.94$). همچنین، pH با فلزات سنگین نظیر آلومینیوم ($r = -0.35$) و سیلیسیم ($r = -0.28$) همبستگی منفی داشت که به تغییرات حلالیت این عناصر در شرایط اسیدی مربوط است. تحلیل مؤلفه‌های اصلی (PCA) نشان داد در دوره کم‌آبی، دو مؤلفه اصلی بیش از ۹۰ درصد واریانس داده‌ها را توضیح می‌دهند. مؤلفه اول تحت تأثیر یون‌های محلول مانند سدیم، کلرید و گوگرد بود، در حالی که مؤلفه دوم با پارامترهایی چون EC و کدورت مرتبط بود. در دوره پرآبی، مؤلفه اول با ۹۵.۶۹ درصد واریانس، عمدتاً تحت تأثیر TDS و EC قرار داشت. مدل‌سازی غلظت کلسیم با استفاده از مدل‌های یادگیری عمیق نشان داد که مدل CNN عملکرد بهتری نسبت به LSTM دارد. در ایستگاه شیرگاه، مدل CNN با ضریب همبستگی ۰.۷۵ تطابق بیشتری با داده‌های واقعی داشت، در حالی که LSTM ضریب ۰.۶ نشان داد.

واژه‌های کلیدی: تحلیل مؤلفه‌های اصلی (PCA)، دیاگرام تیلور، ضریب همبستگی پیرسون، فلزات سنگین

^۱ دانشجوی دکتری علوم و مهندسی آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران. minaaahmadi5462@gmail.com

^۲ دانشیار، گروه علوم و مهندسی آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران.

* نویسنده مسئول: f.shokrian@sanru.ac.ir - shokrian.f@gmail.com

^۳ استاد، پژوهشکده سنجش از دور و GIS محیطی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران. ka.solaimani@sanru.ac.ir solaimani2001@yahoo.co.uk

^۴ دکتری، رئیس مرکز ملی مطالعات و تحقیقات دریای خزر mbanihashemi@hotmail.com

^۵ استاد، گروه علوم و مهندسی آبخیزداری، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران. ataollah.kavian@gmail.com



مقدمه

طبیعی متنوع هستند. حدود ۳۰٪ از آب‌های شیرین تجدیدپذیر قابل دسترس در سطح جهانی توسط صنایع و شهرها استفاده می‌شود که به‌طور مشترک مقدار زیادی فاضلاب حاوی تعداد زیادی ماده شیمیایی در غلظت‌های مختلف تولید می‌کنند (Cosgrove و Rijsberman، ۲۰۱۴). آلودگی به فلزات سنگین و آلاینده‌های مختلف در آب رودخانه‌ها به دلیل تخلیه‌های صنعتی و پساب‌های فاضلاب یک مشکل زیست محیطی در کشورهای در حال توسعه است (ناصری و همکاران، ۱۴۰۳؛ ابراهیمی و همکاران، ۱۴۰۳؛ Shah و همکاران، ۲۰۲۴؛ Saravanan و همکاران، ۲۰۲۴؛ Darma و همکاران، ۲۰۲۴؛ Wen و همکاران، ۲۰۲۵). صنعتی شدن سریع و شهرنشینی به آلودگی آب و تشدید کمبود آب برای بیش از ۸۰ درصد از جمعیت جهان کمک می‌کند (Islam و همکاران، ۲۰۱۸). فلزات سنگین (نظیر سرب، جیوه و کادمیوم) و متالوئیدها^۱ (مانند آرسنیک)، ناشی از فعالیت‌های انسانی، یکی از منابع مهم آلودگی آب می‌باشند (Rahman و Singh، ۲۰۱۹). انتشار مواد شیمیایی، فلزات سنگین و آلودگی‌های دیگر به آب‌های سطحی می‌تواند موجب آسیب به اکوسیستم‌های آبی، کاهش تنوع زیستی و تهدید سلامت انسان منجر گردد (Chung و همکاران، ۲۰۱۶). از این‌رو بررسی غلظت فلزات سنگین در آب‌های سطحی از اهمیت بالایی برخوردار است تا بتوان تأثیرات مضر آن‌ها بر اکوسیستم‌های آبی و پیامدهای ناشی از آلودگی برای سلامت عمومی را شناسایی و ارزیابی کرد (Kumar و همکاران، ۲۰۲۰). در این میان، بررسی پارامترهای فیزیکی و شیمیایی آب رودخانه‌ها به منظور ارزیابی کیفیت آب و شناسایی منابع آلودگی از اهمیت بالایی برخوردار است. مطالعات متعددی در زمینه ارزیابی کیفیت آب رودخانه‌ها با استفاده از روش‌های آماری پیشرفته و مدل‌های یادگیری ماشین انجام شده است. قشلاقی و همکاران (۱۳۹۴) در بررسی ویژگی‌های هیدروژئوشیمیایی آب رودخانه کاکارضا، از تحلیل‌های آماری برای شناسایی تیپ هیدروژئوشیمیایی آب و فرآیندهای مؤثر بر ترکیب شیمیایی آن استفاده کردند. نتایج این مطالعه نشان داد که انحلال سنگ‌ها و تبادل یونی از

رودخانه‌ها به‌عنوان اکوسیستم‌های کلیدی با نقش‌های حیاتی در حفظ تنوع زیستی و تأمین آب برای مصارف انسانی، به دلایل مختلفی در معرض آلودگی‌های طبیعی و مصنوعی قرار دارند (Bhuyan و همکاران، ۲۰۱۷؛ Bawra و Pawar، ۲۰۲۳؛ Grizzetti و Poikane، ۲۰۲۴). تغییرات در کاربری اراضی، همچون ساخت و ساز، کشاورزی و توسعه‌های صنعتی، می‌تواند به آلودگی‌های طبیعی و مصنوعی رونق بخشد و در نتیجه تأثیرات منفی بر کیفیت آب و اکوسیستم‌های آبی ایجاد کند (Cheng و Hu، ۲۰۱۳؛ Romani و همکاران، ۲۰۲۴؛ Bouriqi و همکاران، ۲۰۲۴؛ Nishan و همکاران، ۲۰۲۴). این تأثیرات نامطلوب نه تنها بر اکوسیستم‌های آبی و تنوع زیستی آن‌ها تأثیر می‌گذارد، بلکه می‌تواند کیفیت آب را کاهش داده و سلامت عمومی ساکنان اطراف رودخانه را به خطر اندازد، در نتیجه ضرورت مدیریت و حفاظت از این منابع آبی را افزایش می‌دهد (Choudhury و همکاران، ۲۰۲۲). تغییرات کاربری زمین به‌طور مستقیم بر محیط‌های آبی شهری تأثیر می‌گذارد و این تغییرات می‌تواند ساختار اکوسیستم و جریان انرژی در سیستم‌های آبی را دستخوش تغییر نماید (Vadrevu و Ohara، ۲۰۲۰). فعالیت‌های انسانی به‌ویژه در حوزه بهره‌برداری و دگرگونی منابع طبیعی، به‌طور مداوم محیط اطراف را تحت تأثیر قرار می‌دهد و این تأثیرات به‌ویژه در زمینه تغییر کاربری زمین نمایان می‌شود (Pranoto و همکاران، ۲۰۱۹). در دهه‌های اخیر، به دلیل رشد سریع جمعیت و نوآوری‌های تکنولوژی، شدت استفاده از زمین به‌طور چشمگیری افزایش یافته است (Zhang و همکاران، ۲۰۲۵). نتیجه چنین فشارهایی می‌تواند آلودگی شدید آب‌های سطحی باشد، که این خود به‌عنوان یک تهدید جدی برای بقا و توسعه انسان شناخته می‌شود (Ju و همکاران، ۲۰۲۰). این تغییرات نه تنها بر کیفیت آب تأثیر می‌گذارد بلکه می‌تواند به‌طور وسیع‌تری بر سلامت اکوسیستم‌های آبی و تداوم خدمات اکوسیستمی که انسان به آن‌ها وابسته است، اثر بگذارد. منابع آلاینده‌ها در آب‌های

¹ Metalloids

می‌باشد. محجوبی و همکاران (۱۳۹۹) نیز در مطالعه‌ای بر روی رودخانه‌های اطراف معدن مس سونگون، تأثیر فعالیت‌های معدنی بر کیفیت آب و غلظت فلزات سنگین را بررسی کردند. نتایج این تحقیق نشان داد که افزایش pH آب به دلیل واکنش با سنگ‌های آهکی بستر رودخانه، منجر به بهبود توان خودپالایی رودخانه می‌شود. این مطالعه همچنین تأکید کرد که غلظت فلزات سنگین در آب به طور مستقیم تحت تأثیر تغییرات pH و همبستگی با رسوبات منطقه قرار دارد. هدایت‌زاده و همکاران (۱۳۹۹) در بررسی آلودگی فلزات سنگین در آب و رسوبات رودخانه کارون، از روش‌های آماری مانند ANOVA و همبستگی پیرسون برای تحلیل داده‌ها استفاده کردند. نتایج این مطالعه نشان داد که پارامترهای فیزیکوشیمیایی مانند pH و مواد آلی تأثیر قابل توجهی بر توزیع و تجمع فلزات سنگین در آب و رسوبات دارند. این یافته‌ها اهمیت استفاده از روش‌های آماری پیشرفته را در شناسایی عوامل مؤثر بر کیفیت آب برجسته می‌کند. مدل‌های یادگیری ماشین ترکیبی^۲ دقت بهبودیافته‌ای در پیش‌بینی پارامترهای کلیدی مانند اکسیژن محلول، فسفر و نیتروژن نشان داده‌اند (Zheng و همکاران، ۲۰۲۴). تکنیک‌های یادگیری بدون نظارت^۳، مانند شبکه‌های باور عمیق^۴، در مقایسه با روش‌های یادگیری نظارت‌شده، توانایی بیشتری در مدیریت داده‌های متغیر از خود نشان داده‌اند (Solanki و همکاران، ۲۰۱۵؛ Khosravi و همکاران، ۲۰۲۵). شبکه‌های عصبی بازگشتی (RNN) و مدل‌های حافظه بلند-کوتاهمدت (LSTM) نیز در پیش‌بینی سطح اکسیژن محلول مؤثر بوده‌اند (Wang و همکاران، ۲۰۲۵) و عملکرد بهتری نسبت به روش‌های سنتی یادگیری ماشین داشته‌اند. شبکه‌های عصبی عمیق (DNNs) نیز با دقت بالا در پیش‌بینی شاخص‌های مختلف کیفیت آب موفق عمل کرده‌اند و امکان ارزیابی لحظه‌ای را فراهم کرده‌اند که وابستگی به تحلیل‌های آزمایشگاهی را کاهش می‌دهد (Toumi و همکاران، ۲۰۲۴). مدل ترکیبی BILSTM-SVM (Jamshidzadeh و همکاران، ۲۰۲۳) و مدل‌های LSTM (Venkata Vara Prasad و همکاران، ۲۰۲۰) با دقت بالا شاخص کیفیت آب (WQI) را پیش‌بینی کرده‌اند. همچنین توسعه مدل LSTM (Malarkodi و

اصلی‌ترین فرآیندهای مؤثر بر ترکیب شیمیایی آب هستند. همچنین، افزایش غلظت فلزات سنگین مانند آرسنیک و کادمیوم در طول رودخانه، ناشی از فعالیت‌های انسانی به ویژه کشاورزی بود. در مطالعه دیگری، سیفی و ریاحی (۱۳۹۸) با استفاده از تحلیل مؤلفه‌های اصلی^۱ (PCA) و مدل رگرسیون خطی چندگانه (MLR)، متغیرهای اصلی تأثیرگذار بر شاخص‌های آلودگی فلزات سنگین را شناسایی کردند. نتایج این تحقیق نشان داد که ۷۰ درصد از نمونه‌های آب در محدوده بحرانی از نظر آلودگی به فلزات سنگین قرار داشتند و فعالیت‌های معدن‌کاری به عنوان یکی از اصلی‌ترین منابع آلودگی شناسایی شدند. این مطالعه تأکید کرد که ترکیب روش‌های آماری پیشرفته مانند PCA و MLR می‌تواند به طور مؤثری برای پیش‌بینی و مدیریت کیفیت آب مورد استفاده قرار گیرد. Ibrahim و همکاران (۲۰۲۱) در مطالعه‌ای بر روی یک رودخانه استوایی، از تحلیل چندمتغیره برای بررسی تغییرات مکانی و زمانی پارامترهای کیفیت آب و تعیین منابع آلودگی استفاده کردند. نتایج این مطالعه نشان داد که منابع انسانی مانند محل دفن زباله، فعالیت‌های صنعتی و تخلیه فاضلاب، به عنوان اصلی‌ترین منابع آلودگی در مناطق با سطح آلودگی متوسط و زیاد شناسایی شدند. همچنین، رواناب کشاورزی به عنوان منبع اصلی آلودگی در مناطق با سطح آلودگی کم معرفی گردید. این مطالعه تأکید کرد که روش‌های آماری مانند تحلیل خوشه‌ای و تحلیل تمایز (DA) می‌توانند به طور مؤثری برای طبقه‌بندی کیفیت آب و شناسایی پارامترهای کلیدی مورد استفاده قرار گیرند. باب‌الحکمی و غلامی سفیدکوهی (۱۳۹۷) در مطالعه‌ای نشان دادند که کیفیت آب رودخانه تالار تحت تأثیر فعالیت‌های انسانی به ویژه پساب‌های صنعتی، خانگی، بیمارستانی و کشاورزی قرار دارد. با استفاده از روش‌های آماری چندمتغیره، ایستگاه‌های نمونه‌برداری به سه گروه متمایز تقسیم شدند و مشخص گردید که ۸۰ درصد تغییرات کیفی آب ناشی از سه عامل اصلی است: عامل اول شامل پارامترهای TDS، EC، Cl، SO₄، Ca، Mg، Na، K، SAR و TH (شاخص‌های اصلی شیمیایی و معدنی)، عامل دوم شامل آلاینده‌های نیتراژ و فسفات و عامل سوم شامل بیکربنات

³ Unsupervised⁴ Deep Belief Networks² Principal Component Analysis² Ensemble



منابع به‌ویژه در مواجهه با تغییرات اقلیمی و نوسانات فصلی، امری ضروری است. رودخانه تالار به‌عنوان یکی از رودخانه‌های مهم شمال کشور، تحت تأثیر عوامل مختلف طبیعی و انسانی قرار دارد که شناخت دقیق روند تغییرات کیفی آن می‌تواند راهگشای تصمیم‌گیری‌های مدیریتی باشد. نوآوری این پژوهش در تلفیق تحلیل‌های آماری مانند PCA با رویکردهای نوین مدل‌سازی مبتنی بر یادگیری عمیق است که امکان تحلیل هم‌زمان روابط بین پارامترهای کیفی و پیش‌بینی دقیق رفتار آن‌ها را فراهم می‌سازد.

مواد و روش‌ها

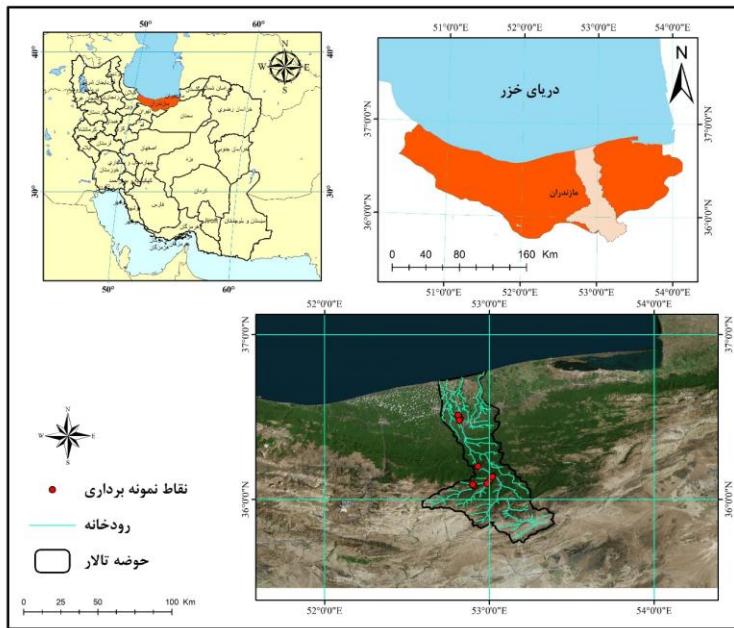
منطقه مورد مطالعه

حوزه آبخیز تالار با مساحت ۲۱۰۰۸۸ هکتار یکی از مناطق کوهستانی شمال ایران در البرز مرکزی است که در دو سمت جاده تهران-قائم‌شهر قرار دارد. این منطقه در مختصات جغرافیایی $22^{\circ} 35' 52''$ تا $23^{\circ} 23' 34''$ طول شرقی و $23^{\circ} 23' 44''$ تا $2^{\circ} 19' 36''$ عرض شمالی واقع شده است. رودخانه تالار که نام آن از این حوضه گرفته شده، از سمت جنوب به شمال جریان دارد و نقش اصلی در زهکشی این منطقه ایفا می‌کند. این حوضه دارای ارتفاعاتی متفاوت است که بالاترین نقطه آن به ارتفاع ۳۹۱۰ متر و پایین‌ترین نقطه به ۲۱ متر در ایستگاه عرب خیل واقع شده است. حوضه آبخیز تالار در زون تکتونیکی البرز قرار دارد و از نظر زمین‌ریخت‌شناسی و زمین‌ساخت، محدوده مورد مطالعه از شمال به جنوب پهنه‌هایی از البرز مرکزی را در بر می‌گیرد که خود شامل واحدها و بخش‌های ساختاری مختلفی است. در این حوضه، سنگ‌های رسوبی با منشأ دریایی و قاره‌ای، همراه با سنگ‌های تبخیری و آذرین، بخش عمده‌ی ترکیب زمین‌شناسی را تشکیل می‌دهند. همچنین، وجود رسوبات ریزدانه و نرم در منطقه، از عوامل مؤثر در افزایش میزان فرسایش به‌شمار می‌رود (سلیمانی و همکاران، ۱۴۰۰). در بررسی ویژگی‌های مورفوزنتیکی حوضه آبخیز تالار، این منطقه در سه ناحیه اقلیمی شامل نیمه‌خشک، ساوان و معتدل قرار دارد. بیشترین بخش از گستره حوضه متعلق به ناحیه ساوان با حدود ۵۰ درصد از

همکاران، ۲۰۲۱) راهکارهای نوینی در پایش و پیش‌بینی کیفیت آب ارائه نموده است. این یافته‌ها گویای اهمیت بالای یادگیری عمیق در مدیریت پارامترهای کیفیت آب می‌باشد. این روش‌های مبتنی بر هوش مصنوعی، راه‌حل‌های کارآمد و انعطاف‌پذیری برای نظارت و مدیریت کیفیت آب ارائه می‌دهند. با توجه به اهمیت رودخانه تالار به‌عنوان یکی از منابع مهم آب در منطقه، این مطالعه با هدف بررسی پارامترهای فیزیکی و شیمیایی آب این رودخانه و شناسایی روابط بین این پارامترها با استفاده از روش‌های آماری مانند ضریب همبستگی پیرسون و تحلیل مؤلفه‌های اصلی (PCA) انجام شد. هم‌چنین، بررسی ارتباط میان فلزات سنگین و دوره‌های هیدرولوژیک، اطلاعات ارزشمندی درباره دینامیک‌های آبی و تأثیرات آن‌ها بر کیفیت آب و زیستگاه‌های آبی فراهم می‌آورد. این تحلیل قادر است الگوهای نوسان در کیفیت آب را شناسایی کرده و تغییرات آن را در واکنش به تحولات اقلیمی و فعالیت‌های انسانی پیش‌بینی نماید. به‌خصوص در شرایط بحرانی کم‌آبی، چنین پیش‌بینی‌هایی می‌تواند مدیریت منابع آب را بهبود بخشد و استفاده بهینه از آن‌ها را ممکن سازد، ضمن اینکه از تشدید چالش‌های زیست‌محیطی و اقتصادی جلوگیری می‌کند. در این تحقیق، به توسعه یک سیستم پیش‌بینی کارآمد برای تخمین غلظت کلسیم در آب پرداخته شد که بتواند به مدیریت بهتر منابع طبیعی، کاهش هزینه‌های اندازه‌گیری و بهبود تصمیم‌گیری‌ها کمک کند. استفاده از مدل‌های پیشرفته یادگیری عمیق مانند LSTM و CNN برای پیش‌بینی پارامترهای کیفی آب یا خاک، یک گام نوآورانه محسوب می‌شود. از جنبه محیط زیستی، پیش‌بینی غلظت کلسیم به مدیریت کیفیت آب و حفاظت از اکوسیستم‌ها کمک می‌کند. این مدل‌ها با کاهش نیاز به نمونه‌برداری و آزمایش‌های مکرر، باعث صرفه‌جویی در زمان و هزینه می‌شوند. در نهایت، این داده‌ها نه تنها در حفظ و پایداری اکوسیستم‌های آبی نقش مؤثری ایفا می‌کنند، بلکه با افزایش آگاهی عمومی، زمینه را برای ایجاد محیطی سالم‌تر و پایدارتر برای جوامع انسانی فراهم می‌سازند. با توجه به اهمیت فزاینده منابع آب سطحی در تأمین نیازهای شرب، کشاورزی و زیست‌محیطی، ارزیابی دقیق کیفیت این

سانتی گراد ثبت شده است. الگوی آبراهه‌های این حوضه به صورت شاخه درختی و موازی دیده می‌شود و شکل کلی آن پهن و دارای امتداد شمالی-جنوبی است. زمان تمرکز حوضه به کمک روش برانسی و ویلیامز حدود ۱۰.۶ ساعت محاسبه گردیده است (محمدی، ۱۳۹۹). شکل (۱) موقعیت جغرافیایی نقاط مورد مطالعه در حوزه آبخیز تالار در استان مازندران را به تصویر می‌کشد.

کل مساحت است، در حالی که ناحیه معتدل کوچک‌ترین سهم را به خود اختصاص داده است (سیلاخوری و همکاران، ۱۴۰۲). مهم‌ترین کاربری‌های اراضی در منطقه مورد مطالعه شامل جنگل‌ها، کشاورزی دیم، زمین‌های آبی، مراتع و مناطق مسکونی می‌باشد (Kavian و همکاران، ۲۰۱۸). بارش سالانه میانگین در این منطقه ۵۵۲/۷ میلی‌متر و دماهای حداقل و حداکثر سالانه به ترتیب ۷.۷ و ۲۱.۷ درجه



شکل (۱): موقعیت جغرافیایی نقاط نمونه برداری در حوزه تالار در استان مازندران و ایران

سانتی‌متری زیر سطح آب انجام شد. فرآیند تعیین نقاط نمونه‌برداری با در نظر گرفتن عوامل محیطی، شرایط دسترسی و قابلیت اجرایی، به گونه‌ای طراحی شد که امکان تحلیل دقیق‌تر و جامع‌تر از وضعیت کیفی رودخانه در شرایط مختلف فراهم شود. از هر محل دو نمونه آب برداشت می‌گردد: یکی برای اندازه‌گیری پارامترهای فیزیکوشیمیایی و دیگری برای تحلیل غلظت فلزات سنگین. همچنین در محل نمونه‌برداری، پارامترهای اسیدیته و هدایت الکتریکی با دستگاه‌های پرتابل اندازه‌گیری شدند و نمونه‌ها نیز تا زمان انتقال به آزمایشگاه در دمای ۴ درجه سانتی‌گراد نگهداری شدند. دوره‌های کم‌آبی (شهریور) و پرآبی (اردیبهشت) رودخانه تالار با توجه به بررسی منابع پیشین و دبی بلندمدت رودخانه تعیین شدند. ظروف نمونه‌برداری ابتدا با اسید نیتریک ۳۰ درصد شسته شده و سپس با آب مقطر تمیز شدند. پس از نمونه‌برداری، برای جلوگیری از

روش تحقیق

در این پژوهش، برای تعیین ایستگاه‌های نمونه‌برداری، ابتدا نقشه‌های توپوگرافی و راه‌های دسترسی به رودخانه مورد مطالعه قرار گرفت. پس از آن، یک بازدید میدانی اولیه از ناحیه صورت گرفت و با در نظر گرفتن کاربری‌های اراضی، ویژگی‌های طبیعی و همچنین روش‌های دسترسی به رودخانه و شاخه‌های فرعی آن، نقاط مناسب برای نمونه‌برداری تعیین شد. این نقاط با بهره‌گیری از نقشه‌های کاربری اراضی و نرم‌افزار گوگل ارث در طول بازدید میدانی شناسایی شدند و فرآیند نمونه‌برداری از آب رودخانه مطابق با استانداردهای مربوط به تهیه نمونه انجام گردید. پس از تهیه نقشه نقاط نمونه‌برداری، نمونه‌برداری از آب رودخانه در ۱۵ نقطه از کاربری‌های متفاوت با دو برداشت (یک بار در شرایط کم‌آبی و یک بار در شرایط پرآبی) استفاده از بطری‌های پلی اتیلن نیم‌لیتری از عمق ۱۵ تا ۳۰



بین ۲۱ پارامتر (Cl, TDS, Turbidity, EC, PH, Sr, Si), HCO₃, SO₄, S, P, Na, Mg, Li, K, Cr, Ca, Ba, Al, NO₃, PO₄, در نرم افزار R استفاده شد. ضریب همبستگی پیرسون بر اساس رابطه (۱) محاسبه می‌شود (Hao و همکاران، ۲۰۲۲):

رابطه (۱):

$$r_{ij} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$

در این فرمول، x_i و y_i نمایندگان داده‌های دو متغیر، \bar{x} و \bar{y} میانگین این داده‌ها و Σ نمایانگر جمع‌هایی است که برای تمام داده‌ها انجام می‌شود و r همبستگی بین هر دو متغیر است. هرچه مقدار مطلق ضریب همبستگی بزرگ‌تر باشد، همبستگی بین دو متغیر بیشتر است، به این معنی که هرچه $|r|$ به ۱ نزدیک‌تر باشد، همبستگی بیشتر است. یک مقدار مثبت برای r نشان‌دهنده همبستگی مثبت است، در حالی که یک مقدار منفی برای r نشان‌دهنده همبستگی منفی می‌باشد (Edelmann و همکاران، ۲۰۲۱). فواصل ضریب همبستگی پیرسون در دامنه‌های مختلف به صورت زیر تعریف می‌شود: از صفر تا ۰/۲ همبستگی بسیار کم، از ۰/۲ تا ۰/۴ همبستگی کم، از ۰/۴ تا ۰/۶ همبستگی متوسط، از ۰/۶ تا ۰/۸ همبستگی زیاد و از ۰/۸ تا ۱ همبستگی بسیار زیاد (Hao و همکاران، ۲۰۲۲).

در این مطالعه ماتریس همبستگی به صورت یک گراف است که با استفاده از نمودار دایره‌ای یا نقشه حرارتی همبستگی^۱ است. این ابزار برای نمایش میزان همبستگی بین متغیرهای مختلف در یک مجموعه داده استفاده می‌شود (Al-Sulttani و همکاران، ۲۰۲۱) و روابط بین معیارهای مختلف (عناصر شیمیایی و ویژگی‌های آب) را نشان می‌دهد. هر خانه در این ماتریس نشان‌دهنده میزان همبستگی بین دو معیار است و مقدار همبستگی با رنگ و اندازه دایره مشخص شده است. این نمودار معمولاً در تحلیل داده‌ها برای نشان دادن قدرت و جهت رابطه بین متغیرهای عددی استفاده می‌شود. از گراف دیگری نیز استفاده شده است که ماتریس همبستگی پیرسون را نمایش می‌دهد، که ارتباط خطی بین متغیرهای مختلف را با استفاده از مقادیر

نفوذ هوا، درب ظروف با نوار چسب بسته شد و در نهایت نمونه‌های مربوط به پارامترهای فیزیکوشیمیایی (Cl, Turbidity, TDS, pH, EC, Sr, Si, S, P, Na, Mg, Li, K, Cr, Ca, Ba, Al, HCO₃, SO₄, NO₃, PO₄, Nayar، ۲۰۲۰؛ Nirmala و همکاران، ۲۰۲۲) به آزمایشگاه مرکزی دانشگاه علوم کشاورزی و منابع طبیعی ساری و نمونه‌های مربوط به فلزات سنگین به آزمایشگاه زرآما در تهران ارسال شدند. در این مطالعه، برای بهبود تحلیل داده‌ها و بررسی اثرات عوامل انسانی و طبیعی بر کیفیت آب رودخانه تالار از روش‌های آماری نظیر ضریب همبستگی پیرسون و تحلیل مؤلفه‌های اصلی استفاده شده است (Shrestha و همکاران، ۲۰۰۸). همچنین، به توسعه یک سیستم پیش‌بینی دقیق و کارآمد برای تخمین غلظت کلسیم در ایستگاه شیرگاه پرداخته شد که با استفاده از مدل‌های پیشرفته یادگیری عمیق LSTM و CNN (Nouraki و همکاران، ۲۰۲۱) محقق شد.

ضریب همبستگی پیرسون

ابزارهای آماری، از جمله ضریب همبستگی پیرسون، نقش بسزایی در تحلیل روابط میان پارامترهای مختلف کیفیت آب ایفا می‌کنند (Bu و همکاران، ۲۰۱۴). ضریب همبستگی پیرسون این امکان را می‌دهد تا روابط خطی بین دو مجموعه از داده قابل شناسایی شود. با استفاده از این ابزار، می‌توان تأثیر متغیرهای مختلف مانند PH، غلظت اکسیژن محلول، نیترات، فسفات و سایر آلاینده‌ها را بر یکدیگر مورد بررسی قرار داد. این اطلاعات می‌توانند به شناسایی الگوهای خاصی در کیفیت آب کمک کنند که با تغییرات جوی یا فعالیت‌های انسانی مرتبط باشند. همچنین، با شناسایی الگوهای تغییر کیفیت آب در طول زمان و در مکان‌های مختلف (Panda و همکاران، ۲۰۱۸)، می‌توان از این اطلاعات برای توسعه مدل‌های پیش‌بینی و ارائه استراتژی‌های بهبود بهره‌برداری کرد. ضریب همبستگی پیرسون به طور گسترده‌ای برای اندازه‌گیری درجه همبستگی بین دو متغیر استفاده می‌شود و مقادیری بین -۱ و ۱ دارد (Feng و همکاران، ۲۰۱۹). در این مطالعه، از ضریب همبستگی پیرسون برای اندازه‌گیری همبستگی

^۱ Correlation heatmap

رابطه (۲):

$$\begin{cases} Z_1 = C_{11}X_1 + C_{12}X_2 + \dots + C_{1p}X_p \\ Z_2 = C_{21}X_1 + C_{22}X_2 + \dots + C_{2p}X_p \\ Z_3 = C_{31}X_1 + C_{32}X_2 + \dots + C_{3p}X_p \\ \dots \\ Z_p = C_{p1}X_1 + C_{p2}X_2 + \dots + C_{pp}X_p \end{cases}$$

برای هر i شامل $1 = C_{i1}^2 + C_{i2}^2 + C_{i3}^2 + \dots + C_{ip}^2$

مراحل عملی تحلیل مؤلفه‌های اصلی (PCA) به صورت

زیر است:

داده‌های اصلی استانداردسازی می‌شود تا اختلاف
ابعادی بین شاخص‌های مختلف نتواند بر نتایج تأثیر بگذارد.

(۱) ماتریس ضریب همبستگی بین متغیرهای R
ایجاد می‌شود.

(۲) مقادیر ویژه λ_j و بردارهای ویژه ماتریس ضریب
همبستگی R محاسبه می‌شود.

(۳) عوامل مؤلفه اصلی به دست می‌آید و نمره کلی
محاسبه می‌شود.

نرخ مشارکت اطلاعات و نرخ مشارکت تجمعی مقدار
ویژه λ_j ($j = 1, 2, \dots, m$) به ترتیب محاسبه می‌شود. در
میان آن‌ها، $b_j = \lambda_j / \sum_{k=1}^m \lambda_k$ نرخ مشارکت اطلاعات
مؤلفه اصلی است و $a_p = \sum_{k=1}^p \lambda_k / \sum_{k=1}^m \lambda_k$ نرخ
مشارکت تجمعی هر مؤلفه اصلی می‌باشد. زمانی که a_p به
۰.۸۵ برسد، نشان می‌دهد که عوامل تأثیرگذار توانسته‌اند
متغیرهای اصلی را توضیح دهند. بنابراین، اولین متغیرهای
 p به عنوان عوامل مؤلفه اصلی انتخاب می‌شوند.

تحلیل مؤلفه‌های اصلی در نرم‌افزار R انجام شد و نتایج
مربوطه به دست آمد. این تحلیل به شناسایی الگوها و روابط
موجود در داده‌ها کمک کرد و اهمیت مؤلفه‌های اصلی را
نیز مشخص نمود. در این مطالعه بررسی روابط و توزیع
واریانس داده‌ها با نمودار تحلیل مؤلفه‌های اصلی و بای‌پلات^۱
نشان داده شده است. نمودار تحلیل مؤلفه‌های اصلی که
برای کاهش ابعاد داده‌ها و نمایش روابط بین متغیرها و
نمونه‌ها استفاده می‌شود دارای محورهای افقی و عمودی
است که به ترتیب مؤلفه اصلی اول (PC1) و مؤلفه اصلی
دوم (PC2) هستند (Zhao و همکاران، ۲۰۲۱). نمودار

رنگی نشان می‌دهد (Zhu و همکاران، ۲۰۲۲). این نمودار،
ماتریس همبستگی پیرسون برای مجموعه‌ای از پارامترهای
کیفیت آب را نشان می‌دهد. هر خانه در ماتریس
نشان‌دهنده همبستگی بین دو پارامتر است و مقادیر به
صورت طیفی از رنگ‌ها (آبی تا نارنجی) نمایش داده شده‌اند.
رنگ‌های تیره‌تر (آبی) نشان‌دهنده همبستگی منفی و
رنگ‌های روشن‌تر (نارنجی) نشان‌دهنده همبستگی مثبت
هستند.

تحلیل مؤلفه اصلی (PCA)

تحلیل مؤلفه‌های اصلی به‌عنوان یک تکنیک آماری
چندمتغیره به کار می‌رود که به شناسایی روابط میان
متغیرهای اصلی و تبدیل آن‌ها به مؤلفه‌های مستقل
می‌پردازد (Arslan، ۲۰۱۳). این روش به طور مؤثری
همبستگی‌های موجود بین شاخص‌های ارزیابی را از بین
می‌برد و به طور چشمگیری محاسبات شاخص‌ها را کاهش
می‌دهد (Loska و Wiechuła، ۲۰۰۳). در سال‌های اخیر،
کاربرد PCA در مسائل گوناگون محیط‌زیستی، از قبیل
ارزیابی جامع تغییرات زمانی و مکانی کیفیت آب‌های
سطحی و زیرزمینی، شناسایی منابع عمده آلودگی در
مناطق آلوده و بهینه‌سازی شبکه‌های نظارت بر کیفیت آب
افزایش یافته است. در این مطالعه به منظور شناسایی عوامل
مؤثر در تغییرات کیفیت آب، تحلیل مؤلفه‌های اصلی به کار
گرفته شد. از طریق اندازه‌گیری‌های میدانی، Olsen و
همکاران (۲۰۱۲) نشان دادند که PCA بر دیگر روش‌ها
برتری دارد زیرا می‌تواند شناسایی بهتری از عوامل آلاینده
مؤثر در طول رودخانه ارائه دهد. PCA اطلاعاتی درباره
مهم‌ترین پارامترهای کیفیت به دلیل تغییرات مکانی و
فصلی استخراج می‌کند (Singh و همکاران، ۲۰۰۴).

تحلیل مؤلفه‌های اصلی برای اولین بار توسط K. Pearson
در سال ۱۹۰۱ پیشنهاد شد، سپس توسط Hotelling در سال ۱۹۳۳
بهبود یافت و به بردار تصادفی گسترش داده شد (Wu و همکاران، ۲۰۲۱). مدل ریاضی
PCA به صورت رابطه (۲) قابل بیان است.

^۱ Biplot



ماهنامه می‌باشد. برای پیش‌بینی میزان کلسیم، از معیارهای موثر چون دبی، حداکثر دما، بارش، تبخیر و تعرق، و رواناب استفاده شد. این پارامترها به عنوان ورودی‌های اصلی در مدل‌سازی در نظر گرفته شدند، زیرا تأثیر مستقیمی بر کیفیت آب و غلظت مواد حل‌شده در آن دارند. از نمودار، میانگین خطای مطلق^۱ (MAE) در طول دوره‌های آموزشی (Epochs) برای دو مجموعه داده‌ی آموزش^۲ و اعتبارسنجی^۳ استفاده شد. این نمودار واریانس خطا در مراحل مختلف یادگیری و قابلیت تعمیم مدل به داده‌های جدید را ارزیابی می‌کند. در نهایت، نمودار تیلور^۴ که ابزاری کارآمد برای مقایسه عملکرد مدل‌های مختلف در شبیه‌سازی متغیرهای مشاهده‌شده است، ارائه شد. این نمودار نحوه تطابق پیش‌بینی‌های مدل‌ها با داده‌های واقعی را به تصویر می‌کشد و مقایسه بصری واضحی از دقت و تعمیم مدل‌ها ارائه دهد.

دیاگرام تیلور

دیاگرام تیلور به عنوان یک ابزار ساده و قابل فهم برای ارزیابی دقت مدل‌های پیش‌بینی و همچنین مقایسه‌ی مدل‌های مختلف بسیار مفید است. نمودار تیلور بر اساس رابطه بین ضریب همبستگی^۵ (R)، انحراف معیار و RMS (Taylor, ۲۰۰۱) بنا شده است. دیاگرام تیلور به صورت یک نمودار دو بعدی ارائه می‌شود که در آن محور افقی مقدار واقعی متغیر وابسته (داده‌های مشاهداتی) و محور عمودی مقدار پیش‌بینی شده توسط مدل را نشان می‌دهد (Simão و همکاران، ۲۰۲۰). روش ارزیابی در دیاگرام تیلور بر اساس فاصله مدل‌های پیش‌بینی از مقدار واقعی داده‌ها انجام می‌شود. هر مدل که به نقطه مرجع (داده‌های مشاهداتی) نزدیک‌تر باشد، دارای دقت بیشتری است و برآورد بهتری از پدیده‌ی مورد نظر دارد (Ghimire و همکاران، ۲۰۲۱). در واقع، فاصله بین مدل پیش‌بینی و واقعیت داده‌ها در این دیاگرام به عنوان معیاری برای اندازه‌گیری دقت مدل‌ها و روش‌های پیش‌بینی استفاده می‌شود.

بای‌پلات PCA نوعی نمودار پراکندگی است که توزیع داده‌های مختلف را بر روی دو محور اصلی (Dim-1 و Dim-2) نشان می‌دهد (Moncada و همکاران، ۲۰۲۱). این دو مؤلفه بخش عمده‌ای از واریانس داده‌ها را توضیح می‌دهند و هدف آن‌ها ساده‌سازی داده‌ها و نمایش ساختار اصلی آن‌ها در فضای دو بعدی است. در این نمودار، متغیرهای مختلف به صورت فلش‌هایی نمایش داده شده‌اند و رنگ‌ها براساس مقدار \cos^2 هستند (Meena و همکاران، ۲۰۲۳). نمودار بای‌پلات تحلیل PCA نشان می‌دهد که مربع کسینوس (\cos^2) چقدر متغیرها را در صفحه مؤلفه‌های اصلی دقیق نشان می‌دهد. در واقع مربع کسینوس معیاری برای سنجش دقت نمایندگی متغیرها در صفحه مؤلفه‌های اصلی است (Giovino و همکاران، ۲۰۲۳). مقادیر بالای \cos^2 نشان‌دهنده این است که متغیرها به خوبی در فضای PCA نمایانده شده‌اند و لذا می‌توان انتظار داشت که این متغیرها تأثیر زیادی در توزیع داده‌ها داشته باشند. برعکس، مقادیر پایین \cos^2 اشاره به این دارد که ممکن است آن متغیر در توضیح واریانس داده‌ها در این ابعاد مؤثر نباشد، بنابراین نمایندگی‌شان در نمودار غیرقابل اعتماد است. به‌طور کلی، \cos^2 می‌تواند به تحلیل‌گران کمک کند تا متغیرهای مهم‌تر را شناسایی کنند و از این رو در تفسیر و نتیجه‌گیری داده‌ها مؤثر باشد.

پیش‌بینی پارامترهای کیفیت آب با مدل‌های یادگیری عمیق در رودخانه تالار

در مطالعه پیش‌بینی پارامترهای کیفیت آب، به‌ویژه کلسیم در رودخانه تالار، از مدل‌های یادگیری عمیق شامل LSTM و CNN در ایستگاه شیرگاه و در محیط برنامه‌نویسی پایتون استفاده شده است. انتخاب کلسیم برای ارزیابی کارایی مدل‌های یادگیری عمیق به علت کامل بودن داده‌های سری زمانی این پارامتر نسبت به سایر پارامترها می‌باشد. دوره شبیه‌سازی در دوره آماری ۸ ساله (۲۰۲۰-۲۰۱۳) به‌صورت

4 Taylor Diagram
5 Correlation Coefficient

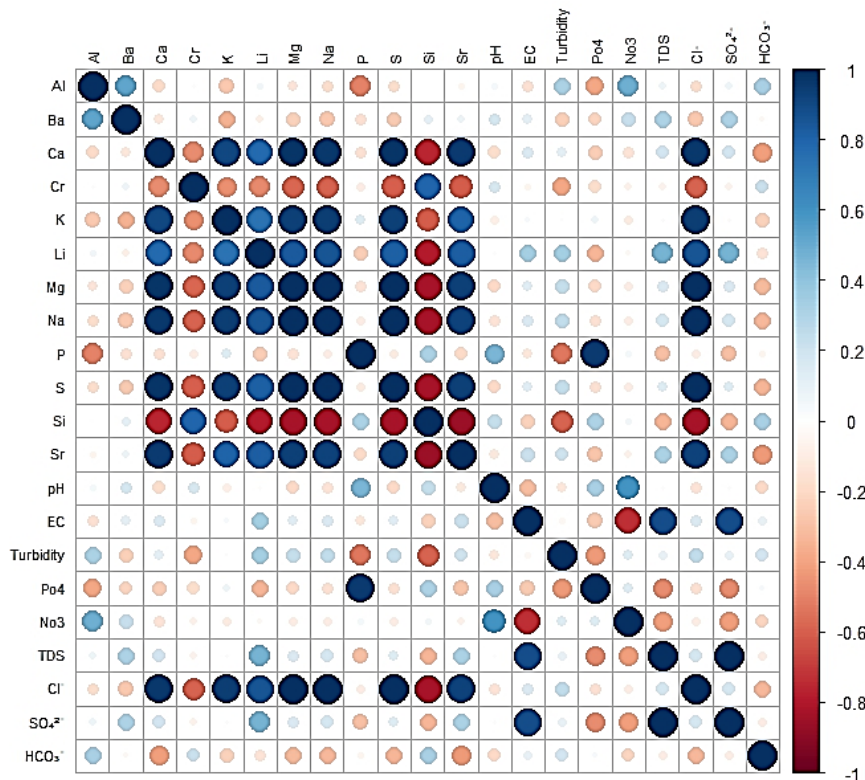
1 Mean Absolute Error (MAE)
2 Train
3 Validation

نتایج و بحث

نتایج همبستگی پارامترهای فیزیکی و شیمیایی آب رودخانه تالار با ضریب پیرسون در دوره کم آبی

برای تحلیل همبستگی میان پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره کم آبی، داده‌های مربوط به این پارامترها جمع‌آوری شد. نمونه‌برداری به صورت منظم انجام شد و داده‌ها پس از بررسی صحت و کامل بودن، برای تحلیل آماده شدند. شکل (۲) نمودار همبستگی پارامترهای شیمیایی و فیزیکی آب رودخانه تالار در دوره کم آبی را نشان می‌دهد. نتایج حاصل از بررسی همبستگی پارامترهای فیزیکی و شیمیایی آب رودخانه تالار با ضریب پیرسون در دوره کم آبی نشان داد که آلومینیوم نسبت به سایر عناصر رابطه نسبتاً ضعیفی دارد و این موضوع ممکن است به شرایط محیطی متفاوتی که در آن وجود دارد، مربوط باشد. باریوم با آلومینیوم همبستگی قوی مثبت دارد که می‌تواند ناشی از منشأ شیمیایی مشابه آن‌ها باشد. کلسیم و منیزیم نیز ارتباط مثبت قوی دارند که نشان‌دهنده مشابهت‌های زمین‌شناسی آن‌ها است. پتاسیم با سدیم، منیزیم و کلسیم همبستگی مثبت داشته و به وضوح در

مناطق کشاورزی یافت می‌شود. همچنین، لیتیوم همبستگی مثبت با پتاسیم و سدیم دارد که می‌تواند نشانه‌ای از منابع مشترک باشد، در حالی که همبستگی منفی با سیلیسیم دارد. منیزیم هم همبستگی منفی قوی با کلسیم و استرانسیوم نشان می‌دهد. همچنین، سدیم با پتاسیم و کلرید همبستگی مثبت دارد و این می‌تواند به وجود در آب‌های شور اشاره کند. همبستگی مثبت بین Na و Cl به حضور ترکیباتی مانند سدیم کلرید اشاره دارد که با نتایج Elsayed و همکاران (۲۰۲۰) مطابقت دارد. فسفر و پتانسیل آن برای همبستگی مثبت با PO_4 ، به‌ویژه به دلیل منابع مشترک (مانند کودهای شیمیایی) معنا دارد. PH نیز نسبت به نیترات همبستگی مثبت دارد، اما همبستگی منفی آن با بسیاری از عناصر نشان‌دهنده دسترسی بهتر آن‌ها در شرایط اسیدی است. هدایت الکتریکی (EC) و TDS هم به وضوح همبستگی مثبت قوی دارند که بر تأثیر کاتیون‌ها در هدایت الکتریکی آب تأکید می‌کند. در نهایت، همبستگی مثبت کلرید با بیشتر کاتیون‌ها به تأثیرات ترکیبی این عناصر بر کیفیت و ویژگی‌های آب اشاره دارد. به طور کلی، این تحلیل همبستگی بین عناصر مختلف می‌تواند به درک بهتر منشأ و شرایط محیطی مؤثر بر کیفیت آب کمک کند.



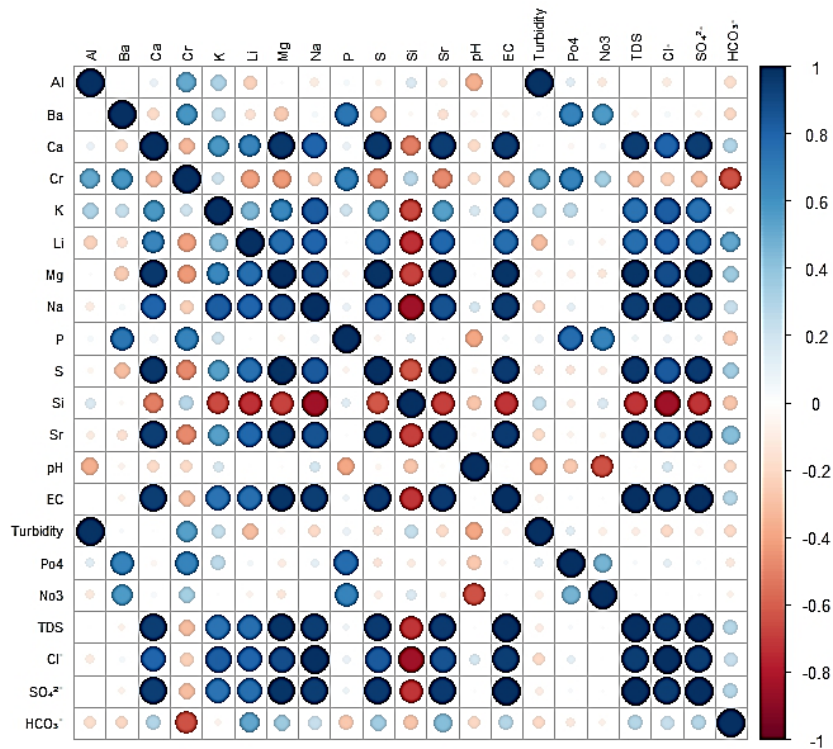


شکل (۲): نمودار همبستگی بین پارامترهای شیمیایی و فیزیکی آب رودخانه تالار در دوره کم‌آبی

قوی با Ca، با TDS و EC نیز رابطه دارد که بیانگر نقش آن‌ها در کیفیت آب است. سدیم (Na) همبستگی قوی با کلرید دارد و این وضعیت وجود ترکیبات سدیم کلرید در آب را تأیید می‌کند. فسفر (P) و کروم و باریم هم همبستگی‌هایی نشان می‌دهند. هم همبستگی مثبت با نیترات و فسفات دارد، در حالی که استرانسیوم (Sr) همبستگی مثبت با Ca و Mg دارد که نشان‌دهنده منابع مشترک طبیعی است. PH نیز تأثیر منفی بر حلالیت فلزات دارد و نشان‌دهنده وابستگی آن به شرایط محیطی است. روابط منفی pH با برخی فلزات مانند Al و Si به وابستگی حلالیت این فلزات به اسیدیته آب مرتبط است که با نتایج محجوبی و همکاران (۱۳۹۹)، Saalidong و همکاران (۲۰۲۲) مبنی بر تأثیر تغییرات PH مطابقت دارد. در نهایت، هدایت الکتریکی (EC) و کل جامدات محلول (TDS) همبستگی‌های قوی با کاتیون‌ها و دیگر عناصر دارند که به وضوح نشان‌دهنده تأثیر این یون‌ها بر کیفیت و ویژگی‌های آب است.

نتایج همبستگی پارامترهای فیزیکی و شیمیایی آب رودخانه تالار با ضریب پیرسون در دوره پرآبی

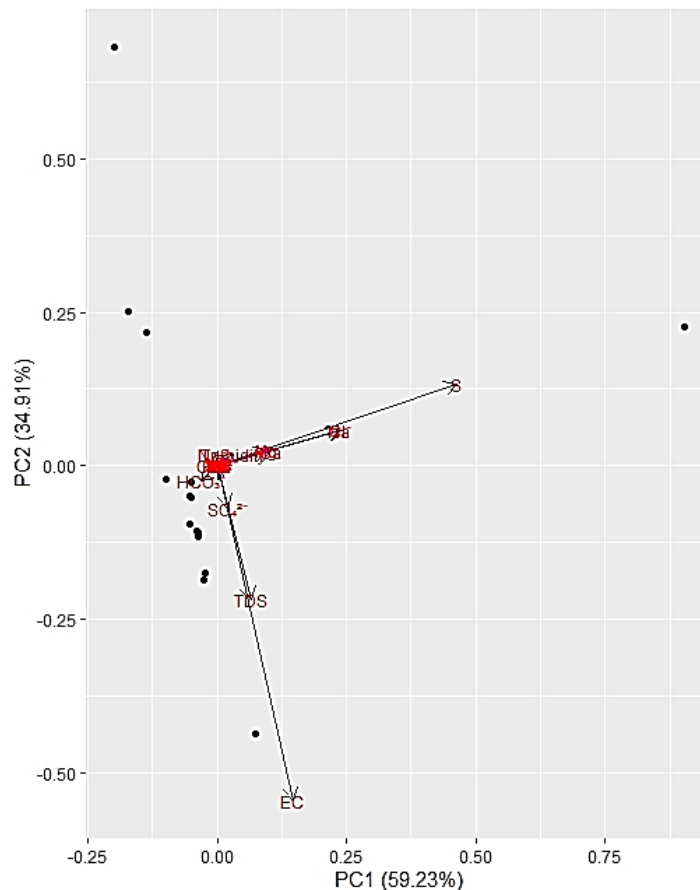
شکل (۳) یک ماتریس همبستگی است که با استفاده از نمودار دایره‌ای و نقشه حرارتی ترسیم شده است و روابط بین معیارهای مختلف (عناصر شیمیایی و ویژگی‌های آب) را نشان می‌دهد. نتایج حاصل از همبستگی پارامترهای فیزیکی و شیمیایی آب رودخانه تالار با ضریب پیرسون در دوره پرآبی نشان داد که آلومینیوم (Al) فقط همبستگی‌های قوی با کدورت و برخی پارامترها دارد و همبستگی ضعیفی با اسیدیته از خود نشان می‌دهد. باریم (Ba) هم همبستگی متوسطی با برخی یون‌ها به ویژه با TDS و EC دارد که نقش آن در هدایت الکتریکی آب را تأیید می‌کند. کلسیم (Ca) و منیزیم (Mg) همبستگی قوی و مثبت دارند که نشان‌دهنده تأثیر آن‌ها در سختی آب است. پتاسیم (K) و لیتیم (Li) نیز همبستگی‌های مثبتی نشان می‌دهند که ممکن است به دلیل حضور مشترک در محیط‌های طبیعی باشد. منیزیم علاوه بر همبستگی مثبت



شکل (۳): نمودار همبستگی بین پارامترهای شیمیایی و فیزیکی آب رودخانه تالار در دوره پرآبی

می‌شوند، معرف متغیرهای اصلی در داده‌ها هستند (متغیرهای شیمیایی یا فیزیکی مانند EC، TDS و غیره). جهت و طول بردارها نشان می‌دهد که هر متغیر چگونه با مؤلفه‌های اصلی ارتباط دارد. هرچه بردار طولانی‌تر باشد، آن متغیر بیشتر در توضیح تغییرات آن مؤلفه نقش دارد. جهت بردار نشان می‌دهد که متغیر با کدام محور ارتباط بیشتری دارد. متغیرهایی که بردارهای آن‌ها در جهت مشابهی هستند، با هم همبستگی مثبت دارند. در مقابل، بردارهایی که در جهت مخالف هستند، نشان‌دهنده همبستگی منفی بین متغیرها هستند. بردارهایی که زاویه‌ی بین آن‌ها نزدیک به ۹۰ درجه دارند، نشان می‌دهند که متغیرها نسبتاً مستقل از هم هستند. بردارهایی که در جهت مخالف یکدیگر هستند (زاویه نزدیک به ۱۸۰ درجه) نشان‌دهنده همبستگی منفی بین متغیرها هستند.

نتایج تحلیل مؤلفه‌های اصلی (PCA) در پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره کم‌آبی شکل (۴) محور افقی نشان‌دهنده اولین مؤلفه اصلی است که ۵۹/۲۳٪ از تغییرات در داده‌ها را توضیح می‌دهد. محور عمودی نشان‌دهنده دومین مؤلفه اصلی است که ۳۴/۹۱٪ از تغییرات را توضیح می‌دهد. این دو مؤلفه اصلی مجموعاً بیش از ۹۰٪ از تغییرات داده‌ها را توضیح می‌دهند. نقاط سیاه کوچک بر روی نمودار نشان‌دهنده نمونه‌های داده‌های اصلی هستند. پراکندگی این نقاط نشان می‌دهد که چگونه داده‌ها بر اساس دو مؤلفه اصلی در فضا توزیع شده‌اند. نقاط نزدیک به یکدیگر به این معناست که این نمونه‌ها شباهت زیادی دارند و در متغیرهای اولیه به هم نزدیک هستند. بردارهای قرمز که از مرکز منشأ شروع



شکل (۴): نمودار تحلیل مؤلفه اصلی (PCA) پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره کم‌آبی

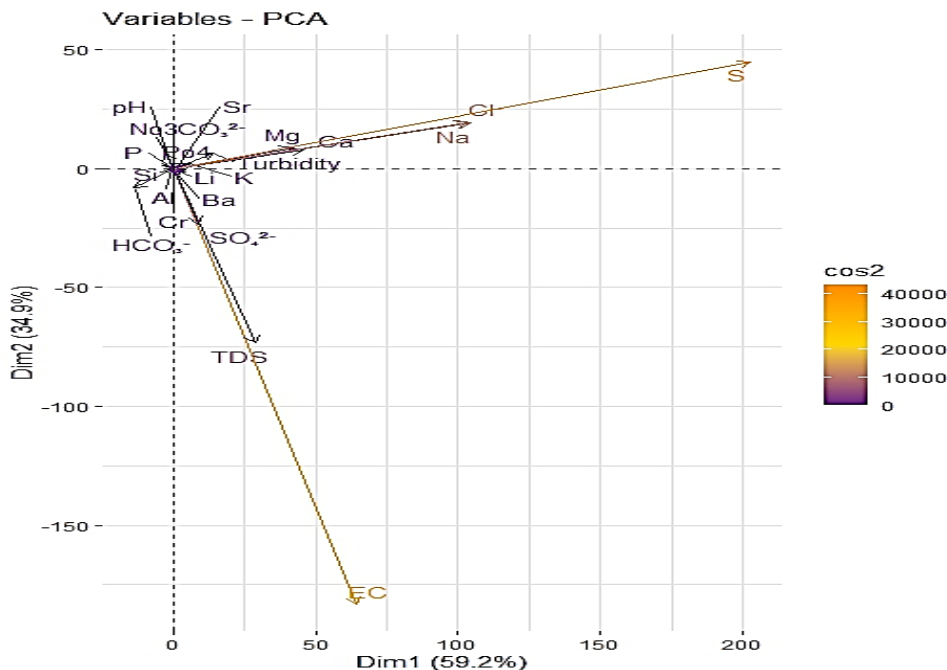
می‌دهد. محورهای این نمودار به عنوان ۱Dim و ۲Dim نمایش داده شده‌اند که به ترتیب ۵۹/۲٪ و ۳۴/۹٪ از

شکل (۵) یک نمودار تحلیل مؤلفه‌های اصلی (PCA) است که متغیرهای مختلف مرتبط با کیفیت آب را نمایش



پس از آن، متغیرهای Cl و Na نقش مهمی دارند، ولی سهم آن‌ها به مراتب کمتر از متغیر S است (حدود ۱۰-۱۵٪). مقادیر \cos^2 نمودار شامل یک مقیاس رنگی به نام \cos^2 است که مقدار کیفیت نمایش متغیرها در ابعاد را نشان می‌دهد. بردارهای زردتر مثل EC و S نشان می‌دهند که این متغیرها به خوبی در این دو بعد نمایش داده شده‌اند و می‌توانند به عنوان متغیرهای کلیدی برای تبیین واریانس کلی داده‌ها استفاده شوند. بردارهای با رنگ بنفش‌تر نشان می‌دهند که متغیرها به خوبی در این ابعاد نمایش داده نشده‌اند و ممکن است به ابعاد دیگر وابسته باشند. تفسیر متغیرهایی که نزدیک به هم قرار دارند (مثل Mg Ca Cl Na) همبستگی بالایی دارند و به احتمال زیاد با هم رابطه مشابهی در داده‌ها دارند. متغیرهایی که جهت مشابه دارند به این معنا هستند که رفتار مشابهی نسبت به این ابعاد دارند. متغیرهایی که زاویه ۹۰ درجه دارند (مانند S و TDS) نشان می‌دهد که این متغیرها همبستگی کمی با یکدیگر دارند. متغیرهایی که مخالف هم هستند (مانند S و EC) نشان‌دهنده رابطه منفی بین آن‌هاست؛ به این معنی که با افزایش یکی از آن‌ها، دیگری کاهش می‌یابد.

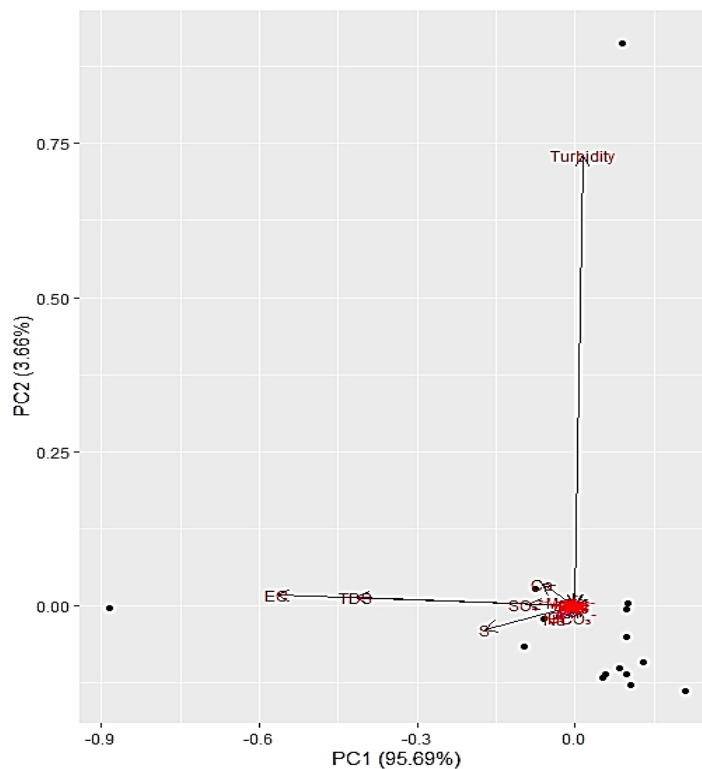
واریانس داده‌ها را توضیح می‌دهند. این بدان معنی است که بعد ۱Dim بیشترین اطلاعات (تغییرات) را در داده‌ها توضیح می‌دهد و ۲Dim به میزان کمتری این تغییرات را پوشش می‌دهد و بیشتر تحت تأثیر متغیرهای TDS (کل مواد جامد محلول) و EC (هدایت الکتریکی) است. بردارها (پیکان‌ها) نشان‌دهنده یک متغیر است و جهت و طول آن اطلاعات مهمی در مورد تأثیر آن متغیر در این تحلیل می‌دهد. طول بردار نشان‌دهنده اهمیت متغیر در تبیین واریانس در ابعاد PCA است. بردارهای بلندتر نشان‌دهنده تأثیر بیشتر در تحلیل هستند. جهت بردار نشان می‌دهد که هر متغیر چطور با ابعاد PCA همبستگی دارد. متغیر EC بیشترین سهم را در مؤلفه دوم ۲Dim دارد (بیش از ۷۰٪)، که نشان می‌دهد EC تأثیر بسیار زیادی در توضیح واریانس مربوط به ۲Dim دارد. پس از آن، TDS با سهمی حدود ۱۰-۱۵٪ در جایگاه دوم قرار دارد. سایر متغیرها مانند S و Cl سهم کمتری دارند و سایر متغیرها تقریباً نقشی در توضیح واریانس ۲Dim ندارند (مقادیر بسیار ناچیز). متغیری که بیشترین سهم را در مؤلفه اول ۱Dim دارد، متغیر S است که سهم بسیار بالایی (بیش از ۴۰٪) دارد.



شکل (۵): نمودار بای‌پلات (Biplot) مؤلفه اصلی (PCA) پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره کم‌آبی

راستای این مؤلفه است و نشان‌دهنده این است که کدورت به‌عنوان یک ویژگی مستقل از سایر متغیرها رفتار می‌کند. کدورت می‌تواند نشان‌دهنده ذرات معلق در آب باشد و معمولاً کمتر با متغیرهای شیمیایی آب مانند یون‌ها و هدایت الکتریکی همبستگی دارد. از این رو، قرارگیری آن در راستای ۲PC به معنای تأثیر کمتر آن بر ۱PC است. جهت و طول بردارها: بردارها نشان‌دهنده چگونگی تأثیرگذاری هر متغیر بر مؤلفه‌ها هستند: بردارهای بلندتر نشان می‌دهند که متغیرها نقش پررنگ‌تری در توضیح واریانس کلی داده‌ها دارند. بردارهای کوتاه‌تر یا نزدیک به مبدأ تأثیر کمتری بر مؤلفه‌ها دارند. نقاط سیاه در نمودار نشان‌دهنده نمونه‌های مختلف داده‌ها هستند و مکان آن‌ها نشان‌دهنده ترکیب خاص متغیرهاست. نمونه‌های نزدیک به هم دارای شباهت‌های بیشتری هستند و خصوصیات شیمیایی مشابهی دارند. نمونه‌های دور از مبدأ یا متغیرهای مشخص، نشان‌دهنده نمونه‌هایی با ویژگی‌های خاص‌تر هستند. نقاطی که در جهت بردارهای EC و TDS قرار دارند، احتمالاً غلظت بالاتری از مواد محلول دارند.

نتایج تحلیل مؤلفه‌های اصلی (PCA) در پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره پرآبی
شکل (۶) یک نمودار تحلیل مؤلفه‌های اصلی (PCA) است که برای کاهش ابعاد داده‌ها و نمایش روابط بین متغیرها و نمونه‌ها استفاده می‌شود. محورهای افقی و عمودی به ترتیب مؤلفه اصلی اول (۱PC) و مؤلفه اصلی دوم (۲PC) هستند. این مؤلفه‌ها به ترتیب ۹۵/۶۹٪ و ۳/۶۶٪ از واریانس داده‌ها را توضیح می‌دهند. در این نمودار، ۱PC بیشترین سهم را در توضیح واریانس داده‌ها دارد (۹۵٪/۶۹٪). این بدان معناست که متغیرهایی که در راستای ۱PC قرار دارند، بیشترین تأثیر را بر توضیح تغییرات در داده‌ها دارند. متغیرهایی مانند EC (هدایت الکتریکی) و TDS (کل جامدات محلول)، که بردارهای بلندی در راستای ۱PC دارند، نشان می‌دهند که این مؤلفه عمدتاً توسط این متغیرها تعیین می‌شود. این دو پارامتر با یکدیگر همبستگی بالایی دارند و با خصوصیات شیمیایی آب مانند میزان نمک‌ها و مواد محلول در ارتباط هستند. ۲PC (مؤلفه اصلی دوم) با اینکه تنها ۳/۶۶٪ از واریانس را توضیح می‌دهد، اهمیت خاص خود را دارد. بردار Turbidity (کدورت) در



شکل (۶): نمودار تحلیل مؤلفه اصلی (PCA) پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره پرآبی



کیفیت آب، نقش قابل توجهی دارند و تأثیر بالایی در تصمیم‌گیری‌های مرتبط با کیفیت آب خواهند داشت. در این میان، EC بیشترین سهم را دارد (حدود ۶۰٪) و TDS در مقام دوم است. متغیرهای دیگر مانند S، Ca و SO_4 نیز مشارکت دارند، اما تأثیر آن‌ها نسبت به EC و TDS کمتر است. از این رو، تمرکز بر روی EC و TDS می‌تواند به بهبود فهم کیفیت آب کمک کند. Turbidity (کدورت) به طور عمده با ۲Dim همراستا است و فلش آن به سمت بالا قرار دارد. سایر متغیرها تأثیر بسیار کمی بر ۲Dim دارند و مشارکت آن‌ها تقریباً صفر است. کدورت، که به شفافیت آب مربوط می‌شود، می‌تواند نشانگر آلودگی‌های ذره‌ای یا مواد معلق در آب باشد و نقش مکملی در تحلیل کیفیت آب دارد. متغیرهای کم‌اهمیت مانند Na، Mg و سایر عناصر شیمیایی به طور عمده در مرکز نمودار تجمع یافته‌اند و دارای فلش‌های کوتاه و رنگ‌های تیره هستند. مقدار پایین CO_2 برای این متغیرها نشان می‌دهد که آن‌ها به خوبی توسط دو مؤلفه اصلی توضیح داده نمی‌شوند. این متغیرها می‌توانند همچنان حاوی اطلاعات مفیدی باشند، اما تأثیر آن‌ها بر واریانس کل داده‌ها کمتر است.

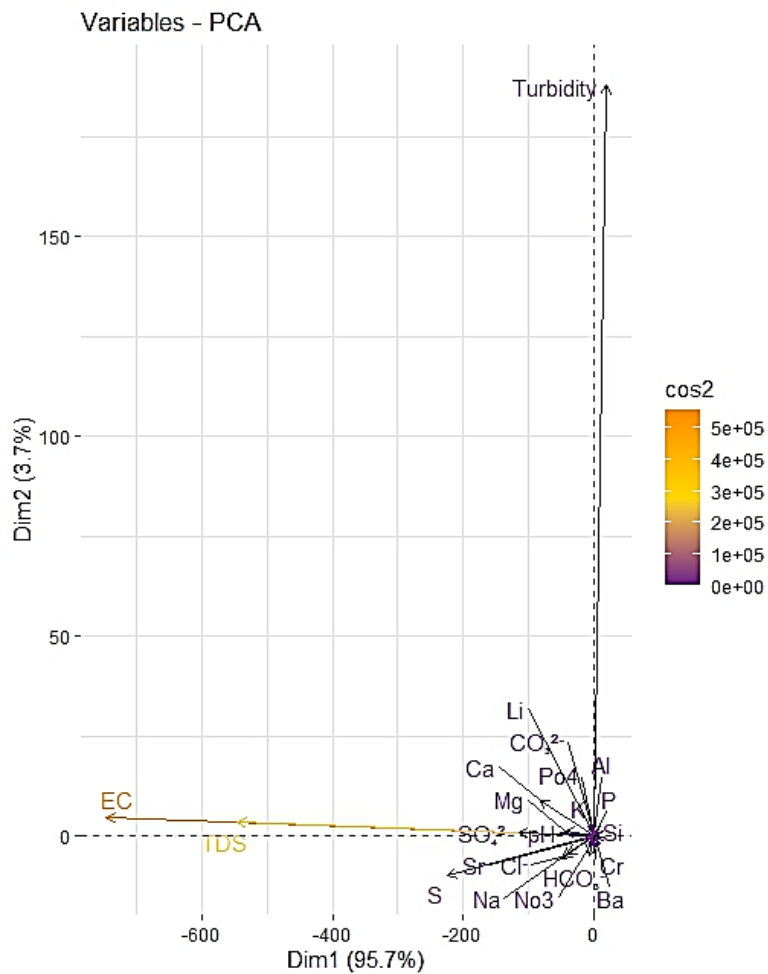
نتایج پیش‌بینی پارامترهای کیفیت آب (کلسیم) با مدل‌های یادگیری عمیق در رودخانه تالار

پیش‌بینی پارامتر کلسیم با مدل‌های یادگیری عمیق در ایستگاه هیدرومتری شیرگاه در شکل (۸) نشان داد خطای آموزش (خط آبی) به تدریج کاهش می‌یابد و به یک مقدار پایدار نزدیک می‌شود. این نشان می‌دهد که مدل به خوبی در حال یادگیری الگوهای موجود در داده‌های آموزشی است. خطاً نسبتاً پایین و پایدار در داده‌های آموزشی نشان‌دهنده توانایی مدل در یادگیری داده‌های آموزشی بدون ایجاد خطاهای بزرگ است. خطای اعتبارسنجی (خط نارنجی) نیز در ابتدا کاهش می‌یابد، اما نسبت به خطای آموزش دارای نوسانات بسیار بیشتری است. این نوسانات نشان‌دهنده تغییرات و حساسیت مدل نسبت به داده‌های اعتبارسنجی است. به نظر می‌رسد که از حدود ۲۰۰ دوره آموزشی به بعد، خطای اعتبارسنجی در مقایسه با خطای آموزش بالاتر است. در شکل (۹) خطای آموزش (خط آبی)

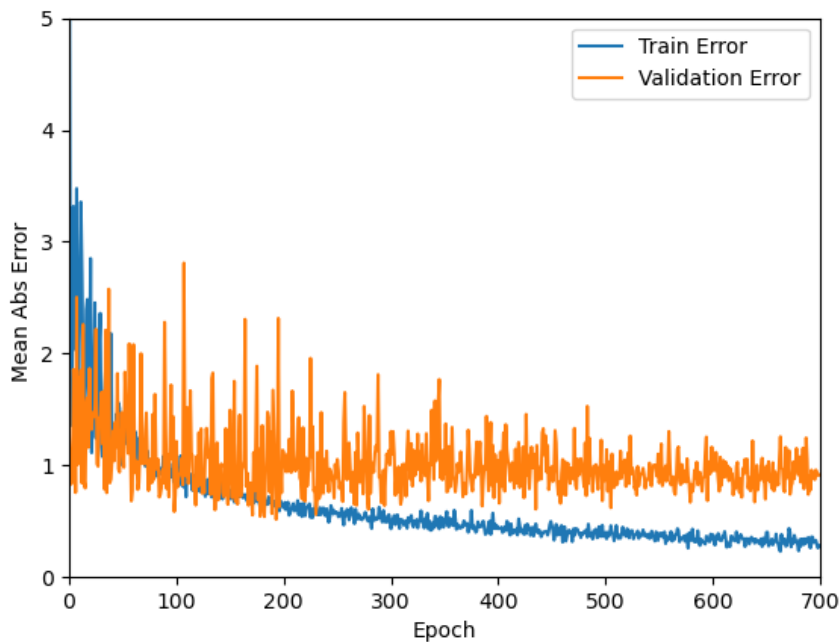
شکل (۷) نمودار بای‌پلات تحلیل مؤلفه‌های اصلی (PCA) است که وضعیت و تأثیر متغیرهای مختلف را بر روی دو مؤلفه اصلی اول (۱Dim- و ۲Dim-) نشان می‌دهد. ۱Dim- (اولین مؤلفه اصلی) به گونه‌ای محاسبه می‌شود که بیشترین مقدار از واریانس کل داده‌ها را توضیح دهد. در این نمودار، ۱Dim- ۹۵٪/۷ واریانس داده‌ها را توضیح می‌دهد که نشان می‌دهد بیشتر تغییرات موجود در داده‌ها توسط این مؤلفه توضیح داده می‌شوند. مؤلفه اول با ۹۵.۷٪ واریانس، اصلی‌ترین عامل تغییرات داده‌ها است و تحت تأثیر هدایت الکتریکی و کل جامدات محلول قرار دارد، که نقش اساسی در ارزیابی کیفیت آب دارند که با نتایج باب‌الحکمی و غلامی سفیدکوهی (۱۳۹۷) مطابقت دارد. ۲Dim (دومین مؤلفه اصلی) به گونه‌ای محاسبه می‌شود که بعد از ۱Dim، بیشترین واریانس باقی‌مانده را توضیح دهد. ۲Dim حدود ۳٪/۷ واریانس باقی‌مانده را توضیح می‌دهد. در نمودار PCA فلش‌های بلندتر نشان می‌دهند که متغیر تأثیر بیشتری بر روی مؤلفه‌ها دارد و نقش مهم‌تری در توضیح واریانس داده‌ها ایفا می‌کند. جهت فلش‌ها نشان‌دهنده رابطه و همبستگی آن‌ها با مؤلفه‌هاست. اگر فلش به سمت یک مؤلفه گرایش داشته باشد، این نشان می‌دهد که متغیر به طور عمده توسط آن مؤلفه توضیح داده می‌شود. رنگ‌بندی بر اساس CO_2 معیاری برای ارزیابی کیفیت نمایش متغیرها در نمودار PCA است: اگر مقدار CO_2 برای یک متغیر بالا باشد، نشان می‌دهد که آن متغیر به خوبی توسط مؤلفه‌های PCA توضیح داده شده است. به عبارت دیگر، درک این متغیر با استفاده از مؤلفه‌های اصلی امکان‌پذیر است. در نمودار، مقادیر بالای CO_2 با رنگ‌های روشن‌تر (زرد) نشان داده شده‌اند، و مقادیر پایین با رنگ‌های تیره‌تر (بنفش). بنابراین، متغیرهای با رنگ زرد مانند EC و TDS نقش کلیدی در توضیح داده‌ها دارند. EC و TDS این دو متغیر به صورت فلش‌های طولانی در نزدیکی محور ۱Dim قرار گرفته‌اند و دارای رنگ زرد هستند. این موضوع نشان‌دهنده تأثیر بالا و مهم این متغیرها در توضیح واریانس داده‌هاست، به طوری که بیشتر تغییرات توسط این دو متغیر توصیف می‌شود. هدایت الکتریکی (EC) و کل جامدات محلول (TDS) به عنوان شاخص‌های مهمی برای ارزیابی

نشان‌دهنده نوسانات واضحی هستند که ممکن است به دلیل رفتار فصلی یا دوره‌ای باشد. مدل‌ها توانسته‌اند این نوسانات را به طور نسبی به خوبی پیش‌بینی کنند. تغییرات سطح کلسیم در دوره زمانی مشخص به طور آشکاری قابل مشاهده است. در شکل (۱۱) نقطه‌ی LSTM (آبی) و CNN (قرمز) نشان‌دهنده عملکرد این دو مدل در پیش‌بینی پارامتر کلسیم است. CNN همبستگی بهتری (نزدیک به ۰/۸) در مقایسه با LSTM (حدود ۰/۶) دارد. خطای RMS فاصله شعاعی هر نقطه از مرکز نمودار نشان‌دهنده مقدار خطای RMS است. CNN در مقایسه با LSTM به مرکز نزدیک‌تر است و خطای RMS کمتری دارد. مدل LSTM عملکرد ضعیف‌تری نسبت به CNN در پیش‌بینی پارامتر کلسیم دارد، زیرا همبستگی کمتری با داده‌های واقعی دارد و انحراف معیار آن دورتر از داده‌های واقعی است و خطای RMS آن بیشتر است. در ایستگاه شیرگاه، مدل CNN نسبت به مدل LSTM برتری نشان داد که با مطالعات Back و همکاران (۲۰۲۰) مبنی بر عملکرد بسیار خوب مدل CNN مطابقت دارد.

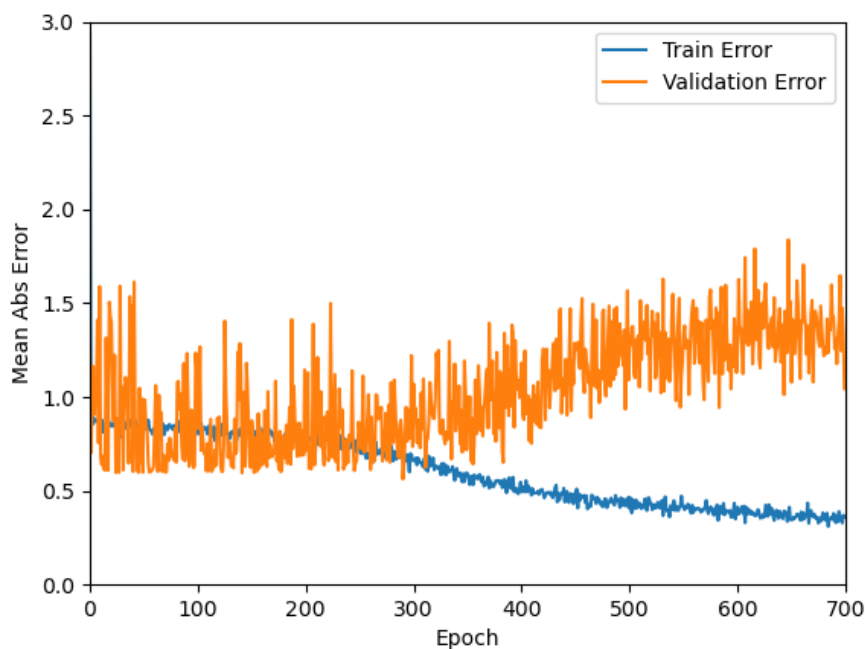
به تدریج کاهش یافته و به حدود ۰/۵ می‌رسد. این کاهش تدریجی نشان‌دهنده یادگیری پایدار مدل از داده‌های آموزشی است. این کاهش نشان می‌دهد که مدل با گذشت زمان به خوبی توانسته داده‌های آموزشی را بهینه کند. اعتبارسنجی (خط نارنجی) در ابتدا کاهش می‌یابد و با خطای آموزش تقریباً هم‌راستا می‌شود (در حدود ۵۰ تا ۱۵۰ دوره آموزشی). اما از حدود ۱۵۰ دوره به بعد، خطای اعتبارسنجی به تدریج افزایش یافته و نوسانات بیشتری را نشان می‌دهد. در این نمودار، بیش‌برازش به طور مشخص‌تری قابل مشاهده است. خطای اعتبارسنجی نه تنها افزایش یافته بلکه نوسانات قابل توجهی نیز دارد. خطای آموزش در این نمودار نسبت به نمودار قبلی کمتر است، که نشان‌دهنده یادگیری قوی‌تر روی داده‌های آموزشی است. در شکل (۱۰) هر سه خط به طور کلی الگوهای مشابهی را نشان می‌دهند و این امر نشان‌دهنده توانایی قابل توجه مدل‌های CNN و LSTM در بازسازی روند اصلی داده‌های مشاهده‌شده است. به نظر می‌رسد مدل CNN (سبز) به طور معمول هماهنگی بهتری با داده‌های واقعی دارد، به ویژه در نقاطی که تغییرات ناگهانی و جدی رخ می‌دهد. داده‌ها



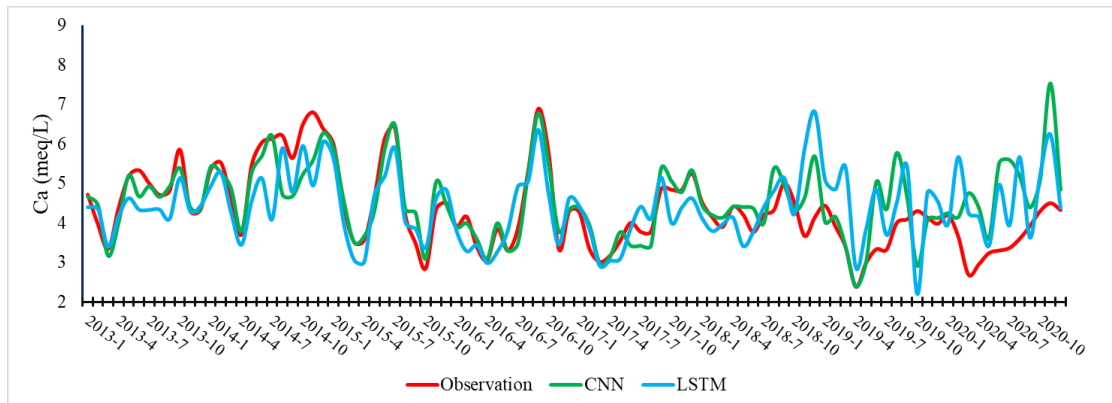
شکل (۷): نمودار بای پلات (Biplot) مؤلفه اصلی (PCA) پارامترهای فیزیکی و شیمیایی آب رودخانه تالار در دوره پرآبی



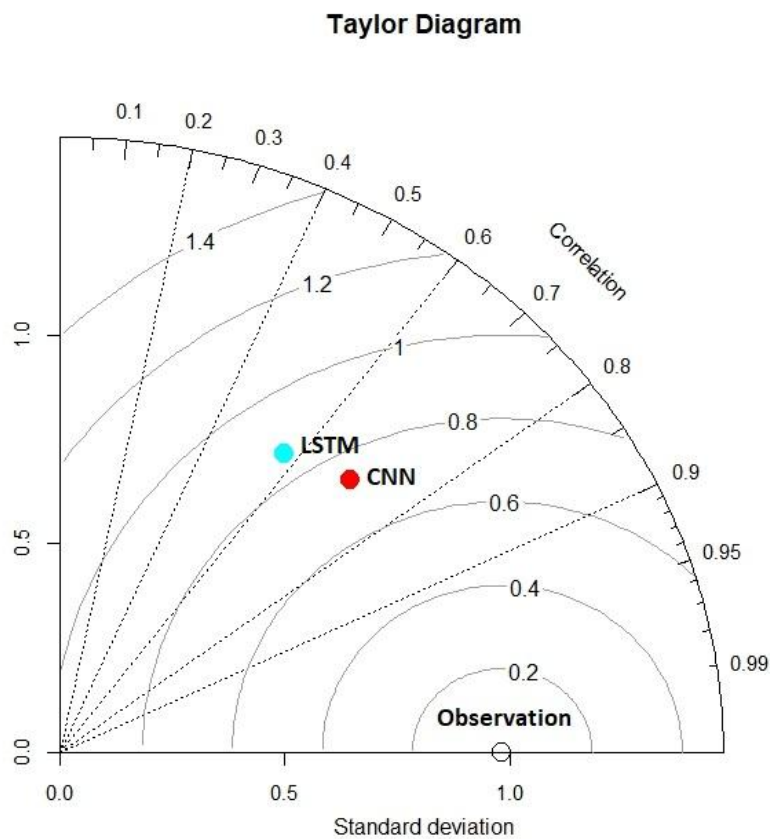
شکل (۸): نمودار تغییرات خطای میانگین مطلق (MAE) مدل CNN در ایستگاه شیرگاه



شکل (۹): نمودار تغییرات خطای میانگین مطلق (MAE) مدل LSTM در ایستگاه شیرگاه



شکل (۱۰): نمودار تغییرات پارامتر کلسیم (Ca) در طول زمان (۲۰۲۰-۲۰۱۳) با مدل‌های LSTM و CNN در ایستگاه شیرگاه



شکل (۱۱): مقایسه عملکرد مدل‌های LSTM و CNN در پیش‌بینی پارامتر کلسیم با نمودار تیلور (Taylor) در ایستگاه شیرگاه

نتیجه گیری

نتایج تحلیل کیفی آب رودخانه تالار نشان می‌دهد که بین پارامترهای فیزیکی و شیمیایی روابط معناداری وجود دارد، به طوری که همبستگی مثبت قوی بین کلسیم و منیزیم نقش مؤثر این دو عنصر در سختی آب را تأیید می‌کند و ارتباط قوی هدایت الکتریکی (EC) و TDS با کاتیون‌های اصلی مانند Ca، Mg و Na نیز نشان‌دهنده تأثیر مستقیم یون‌های محلول بر رسانایی آب است؛ همبستگی Na با Cl احتمال وجود ترکیباتی مانند سدیم کلرید را مطرح می‌سازد و روابط منفی بین pH و فلزاتی مانند Al و Cr گویای تأثیر اسیدیته بر حلالیت این فلزات است، در حالی که فسفات و گوگرد نیز با برخی یون‌ها و TDS همبستگی مثبت دارند که می‌تواند نشان‌دهنده ورود آلاینده‌های انسانی مانند کودهای شیمیایی به آب باشد؛ افزون بر این، کدورت با فلزات سنگین همبستگی ضعیفی دارد که بیانگر نقش ذرات معلق در تغییرات این پارامترهاست. تحلیل مؤلفه‌های اصلی (PCA) نیز در

دوره‌های کم‌آبی و پرآبی نشان داد که دو مؤلفه اصلی (Dim1 و Dim2) بیش از ۹۰٪ واریانس داده‌ها را توضیح می‌دهند و متغیرهایی مانند گوگرد، کلرید، سدیم، EC و TDS بیشترین سهم را در تغییرات دارند؛ به‌ویژه در دوره پرآبی که EC و TDS به‌عنوان شاخص‌های کلیدی شیمیایی معرفی شدند و کدورت نیز به‌عنوان نمادی از تغییرات فیزیکی مطرح گردید. در این میان، تحلیل عملکرد مدل‌های یادگیری عمیق برای پیش‌بینی غلظت کلسیم در ایستگاه شیرگاه نشان داد که مدل CNN نسبت به مدل LSTM عملکرد بهتری دارد که این برتری را می‌توان به ویژگی‌های خاص داده‌ها و توانایی بالای این مدل در شناسایی الگوهای پیچیده نسبت داد. در نتیجه، با توجه به یافته‌های حاصل از تحلیل‌های آماری و مدل‌سازی، پیشنهاد می‌شود در مدیریت کیفیت منابع آب رودخانه تالار از شاخص‌هایی مانند EC، TDS و کدورت به‌عنوان معیارهای اصلی استفاده شود و برای پیش‌بینی‌های دقیق‌تر، بهره‌گیری از مدل CNN در مناطق مشابه با شرایط ایستگاه شیرگاه در اولویت قرار گیرد.

قدردانی: پژوهش حاضر مستخرج از رساله دکتری می‌باشد به این وسیله نویسندگان مراتب تقدیر و تشکر را از حمایت‌های مالی و معنوی دانشگاه علوم کشاورزی و منابع طبیعی ساری اعلام می‌دارند.

منابع

- ابراهیمی، ز.، گ. فرقانی تهرانی، ع. کابلی. ۱۴۰۳. هیدروژئوشیمی زیست محیطی رودخانه زیارت در محدوده شهر گرگان، استان گلستان، مجله علوم و فنون دریایی، ۲۳(۴)، ۳۲-۵۸.
- باب‌الحکمی، ع. و م.ع. غلامی سفیدکوهی. ۱۳۹۷. تجزیه و تحلیل کیفیت آب رودخانه تالار با استفاده از تکنیک‌های چند متغیره. پژوهشنامه مدیریت حوزه آبخیز، ۹(۱۸): ۲۵۹-۲۵۰.
- سلیمانی، ک.، ف. شکریان، س. عبدلی بوژانی، و ع. صابری. ۱۴۰۰. اولویت‌بندی پتانسیل خطر سیل خیزی حوضه آبخیز تالار در محیط GIS. اکوهیدرولوژی، ۸(۳)، ۷۴۹-۷۶۲.
- سیفی، ا. و ح. ریاحی. ۱۳۹۸. ارزیابی شاخص‌های آلودگی فلزات سنگین در آب‌های سطحی معدن مس سرچشمه با روش‌های آماری چندمتغیره و GIS. مجله تحقیقات آب و خاک ایران، ۵۰(۱): ۱۷۶-۱۶۱.
- سیلاخوری، ز.، ق. وهابزاده کبریا، و ح.ر. پورقاسمی. ۱۴۰۲. تهیه نقشه حساسیت زمین‌لغزش با استفاده از مدل بیزین (مطالعه موردی: بخشی از حوضه آبخیز تالار، استان مازندران). پژوهش‌های فرسایش محیطی. ۱۳(۲): ۱۴۰-۱۲۲.



قشلاقی، ا.، گ. فرقانی تهرانی و م. حیات‌الغیب. ۱۳۹۵. هیدروژئوشیمی و غلظت فلزات سنگین در آب رودخانه کاکارضا (استان لرستان). نشریه محیط زیست طبیعی، ۶۸(۴): ۶۲۸-۶۱۹.

محبوبی، ر.، م. حسن‌پور صدقی، ع. ر. واعظی هیر و ن. مظاهری. ۱۳۹۹. غلظت عناصر سنگین و سمی در پایین‌دست باطله‌های معدنی مس سونگون و نقش بستر آهکی رودخانه در کاهش غلظت آلاینده‌ها. مجله اکو هیدرولوژی، ۷(۳): ۷۵۶-۷۴۳.

محمدی خانقاه، م. ۱۳۹۹. تغییرات زمانی منشاء و کیفیت رسوب در حوزه آبخیز تالار. رساله دوره دکتری علوم و مهندسی آبخیزداری - گرایش زمین، دانشگاه تربیت مدرس، ایران، ۸۸ صفحه.

ناصری، ح.ر.، م. عرب، ز. کی همایون، ۱۴۰۳. تاثیر رودخانه سیاهرود بر آلودگی سرب و نیترات آب های زیرزمینی دشت قائمشهر-جویبار، مجله زمین شناسی مهندسی، ۱۸(۲)، ۳۰۳-۳۱۹.

هدایت‌زاده، ف.، ح. ر. پورخباز و م. چراغی. ۱۳۹۹. تغییرات مکانی-زمانی پارامترهای فیزیکوشیمیایی و تأثیر آن‌ها بر تجمع فلزات سنگین در آب و رسوبات آبرگیر تصفیه‌خانه‌های بخشی از رودخانه کارون. نشریه اکوبیولوژی تالاب، ۱۲(۲): ۱۰۹-۹۱.

Al-Sulttani, A. O., M. Al-Mukhtar, A. B. Roomi, A. A. Farooque, K. M. Khedher and Z. M. Yaseen. 2021. Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access*, 9: 108527-108541.

Arslan, O. 2013. Spatially weighted principal component analysis (PCA) method for water quality analysis. *Water Resources*, 40: 315-324.

Baek, S. S., J. Pyo, and J. A. Chun. 2020. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water*, 12(12): 3399.

Bawra, B., & M. Pawar. (2023). Changes in Freshwater Systems Management and Governance: A Case Study from Mandla Madhya Pradesh. In *Perspectives on Global Biodiversity Scenarios and Environmental Services in the 21st Century* (pp. 1-21). IGI Global.

Bhuyan, M.S., M.A. Bakar, A. Akhtar, M.B. Hossain, M.M. Ali and M.S. Islam. 2017. Heavy metal contamination in surface water and sediment of the Meghna River, Bangladesh. *Environmental Nanotechnology, Monitoring & Management*, 8: 273-279.

Bouriqi, A., N. Ouazzani, & J. F. Deliege. (2024). Modeling the Impact of Urban and Industrial Pollution on the Quality of Surface Water in Intermittent Rivers in a Semi-Arid Mediterranean Climate. *Hydrology*, 11(9), 150.

Bu, H., W. Meng, Y. Zhang and J. Wan. 2014. Relationships between land use patterns and water quality in the Taizi River basin, China. *Ecological Indicators*, 41: 187-197.

Choudhury, M., A. Pervez, A. Sharma and J. Mehta. 2022. Human-induced stresses on the rivers beyond their assimilation and regeneration capacity. In *Ecological Significance of River Ecosystems* (pp. 281-298). Elsevier.

Chung, S. Y., S. Venkatramanan, N. Park, T. Ramkumar, S. B. Sujitha and M. P. Jonathan. 2016. Evaluation of physico-chemical parameters in water and total heavy metals in sediments at Nakdong River Basin, Korea. *Environmental Earth Sciences*, 75: 1-12.

Cosgrove, W. J. and F. R. Rijsberman. 2014. *World water vision: making water everybody's business*. Routledge.



Darma, A., S. A. Zakari, A. M. Sani, & H. S. Mamman. (2024). Soil pollution and heavy metal contamination: insight from effluent discharged from challawa industrial area, nigeria. *Fudma Journal of Sciences*, 8(6), 355–361.

Edelmann, D., T. F. Móri and G. J. Székely. 2021. On relationships between the Pearson and the distance correlation coefficients. *Statistics & Probability Letters*, 169: 108960.

Edelmann, D., T.F. Móri and G.J. Székely. 2021. On relationships between the Pearson and the distance correlation coefficients. *Statistics & Probability Letters*, 169: 108960.

Elsayed, S., H. Hussein, F. S. Moghanm, K. M. Khedher, E. M. Eid and M. Gad. 2020. Application of irrigation water quality indices and multivariate statistical techniques for surface water quality assessments in the Northern Nile Delta, Egypt. *Water*, 12(12): 3300.

Feng, W., Q. Zhu, J. Zhuang and S. Yu. 2019. An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. *Cluster Computing*, 22: 7401-7412.

Ghimire, S., Z. M. Yaseen, A. A. Farooque, R. C. Deo, J. Zhang, and X. Tao. 2021. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, 11(1): 17497.

Giovino, A., A. Marchese, F. Bonanno, G. Sala, F. P. Marra and G. Domina. 2023. Morphological and molecular characterization of Sicilian carob (*Ceratonia siliqua* L.) accessions. *Caryologia*, 76(3): 39-49.

Grizzetti, B., & S. Poikane. (2024). The Importance of Inland Waters. In *Wetzel's Limnology* (pp. 7-13). Academic Press.

Hao, N., P. Sun, W. He, L. Yang, Y. Qiu, Y. Chen and W. Zhao. 2022. Water resources allocation in the Tingjiang River Basin: construction of an interval-fuzzy two-stage chance-constraints model and its assessment through Pearson correlation. *Water*, 14(18): 2928.

Hu, Y. and H. Cheng. 2013. Water pollution during China's industrial transition. *Environmental Development*, 8: 57-73.

Ibrahim, T. N. B. T., F. Othman, N. Z. Mahmood and T. Abunama. 2021. Seasonal effects on spatial variations of surface water quality in a tropical river receiving anthropogenic influences. *Sains Malaysiana*, 50(3): 571-593.

Islam, M. S., R. Proshad and S. Ahmed. 2018. Ecological risk of heavy metals in sediment of an urban river in Bangladesh. *Human and Ecological Risk Assessment: An International Journal*, 24(3): 699-720.

Ju, X., X. Gao, W. Li, L. Han and J. Mao. 2020. Effects of land use change on surface runoff in Beijing-Tianjin-Hebei urban agglomeration. *Acta Ecol. Sin*, 40: 1413-1423.

Kavian, A., M. Mohammadi, L. Gholami and J. Rodrigo-Comino. 2018. Assessment of the spatiotemporal effects of land use changes on runoff and nitrate loads in the Talar River. *Water*, 10(4): 445.

Khosravi, K., A.A. Farooque, M. Karbasi, M. Ali, S. Heddami, A. Faghfour and S. Abolfathi. 2025. Enhanced water quality prediction model using advanced hybridized resampling alternating tree-based and deep learning algorithms. *Environmental Science and Pollution Research*, 1-20.

Kumar, V., N. Kamboj, T. Payum, J. Singh and P. Kumar (Eds.). 2020. *Advances in Environmental Pollution Management: Wastewater Impacts and Treatment Technologies*. Agro Environ Media, Publication Cell of AESA, Agriculture and Environmental Science Academy.



Loska, K. and D. Wiechuła. 2003. Application of principal component analysis for the estimation of source of heavy metal contamination in surface sediments from the Rybnik Reservoir. *Chemosphere*, 51(8): 723-733.

Meena, M. K., A. Malik, R. Singh, A. P. Singh, S. Naik, R. K. Meena and V. Kumar. 2023. Morphological and Biochemical Changes in Moth Bean during Drought Stress. *International Journal of Environment and Climate Change*, 13(11): 187-201.

Moncada, A. M., A. M. Melesse, J. Vithanage and R. M. Price. 2021. Long-term assessment of surface water quality in a highly managed estuary basin. *International Journal of Environmental Research and Public Health*, 18(17): 9417.

Nayar, R. 2020. Assessment of water quality index and monitoring of pollutants by physico-chemical analysis in water bodies: a review. *International Journal of Engineering Research and Technology*, 9.(۰۱)

Nirmala, K., P.S. Kumar, N.K. Ambujam and S. Srinivasalu. 2022. Assessment of physico-chemical parameters of surface waters of a tropical brackish water lake in South Asia. *Environmental Research*, 214: 113958.

Nishan, R. K., S. Akter, R. I. Sony, M. M. Hoque, M. J. Anee, & A. Hossain. (2024). IoT-based three-level water quality assessment system for industrial wastewater.

Nouraki, A., M. Alavi, M. Golabi and M. Albaji. 2021. Prediction of water quality parameters using machine learning models: A case study of the Karun River, Iran. *Environmental Science and Pollution Research*, 28(40): 57060-57072.

Olsen, R. L., R. W. Chappell and J. C. Loftis. 2012. Water quality sample collection, data treatment and results presentation for principal components analysis—literature review and Illinois River watershed case study. *Water Research*, 46(9): 3110-3122.

Panda, P. K., R. B. Panda and P. K. Dash. 2018. The study of water quality and Pearson's correlation coefficients among different physico-chemical parameters of River Salandi, Bhadrak, Odisha, India. *American Journal of Water Resources*, 6(4): 146-155.

Pranoto, R., M. Sopandi, R. Harumansah, S. K. Saptomo and A. Darmawan. 2019, April. Water balance prediction as impact land use change by GIS based SCS-CN and Thornthwaite-Mather method. In 2019 5th International Conference on Computing Engineering and Design (ICCED) (pp. 1-6). IEEE.

Rahman, Z. and V. P. Singh. 2019. The relative impact of toxic heavy metals (THMs)(arsenic (As), cadmium (Cd), chromium (Cr)(VI), mercury (Hg), and lead (Pb)) on the total environment: an overview. *Environmental Monitoring and Assessment*, 191: 1-21.

Romani, L. C., M. C. Raimundi, V. D. Longo, A. F. Camargo, & H. Treichel. (2024). Innovative Technologies for the Treatment of Industrial Wastewater. In *Innovative and Hybrid Technologies for Wastewater Treatment and Recycling* (pp. 261-282). CRC Press.

Saalidong, B. M., S. A. Aram, S. Otu and P. O. Lartey. 2022. Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. *PloS One*, 17(1): e0262117.

Saravanan, P., V. Saravanan, R. Rajeshkannan, G. Arnica, M. Rajasimman, B. Gurunathan, & A. Pugazhendhi. (2024). Comprehensive review on toxic heavy metals in the aquatic system: sources, identification, treatment strategies, and health risk assessment. *Environmental Research*, 119440.



Shah, S. K., M. B. Adhikari, & A. Bhattarai. (2024). Analysis of Heavy Metal Contamination in Industrial Wastewater along the Biratnagar–Duhabi Industrial Corridor: Their Environmental Impacts. *Tribhuvan University Journal*, 39(2), 27-46.

Shrestha, S., F. Kazama and T. Nakamura. 2008. Use of principal component analysis, factor analysis and discriminant analysis to evaluate spatial and temporal variations in water quality of the Mekong River. *Journal of Hydroinformatics*, 10(1): 43-56.

Simão, M. L., P. M. Videiro, P. B. A. Silva, L. P. de Freitas Assad, and L. V. S. Sagrilo. 2020. Application of Taylor diagram in the evaluation of joint environmental distributions' performances. *Marine Systems & Ocean Technology*, 15: 151-159.

Singh, K. P., A. Malik, D. Mohan and S. Sinha. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research*, 38(18): 3980-3992.

Solanki, A., H. Agrawal and K. Khare. 2015. Predictive analysis of water quality parameters using deep learning. *International Journal of Computer Applications*, 125(9): 0975-8887.

Taylor, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7): 7183-7192.

Toumi, S., S. Lekmine, N. Touzout, H. Moussa, N. Elboughdiri, R. Boudraa, ... and H. Tahraoui. 2024. Harnessing Deep Learning for Real-Time Water Quality Assessment: A Sustainable Solution. *Water*, 16(23): 3380.

Vadrevu, K. P. and T. Ohara. 2020. Focus on land use cover changes and environmental impacts in South/Southeast Asia. *Environmental Research Letters*, 15(10): 100201.

Wang, D., C. Zhang, A. Li, Y. Guo, H. Zhang and C. Tan. 2025. Spatio-temporal analysis and prediction for raw water quality of drinking water source by improved RNN algorithm. *Journal of Water Process Engineering*, 71: 107164.

Wen, T., Y. Cheng, Y. Yuan, & R. Sun. (2025). Quantitative analysis and risk assessment of heavy metal pollution in an intensive industrial and agricultural region. *Ecotoxicology and Environmental Safety*, 289, 117634.

Wu, J., Z. Wang and L. Dong. 2021. Prediction and analysis of water resources demand in Taiyuan City based on principal component analysis and BP neural network. *AQUA—Water Infrastructure, Ecosystems and Society*, 70(8): 1272-1286.

Zhang, C.F., Z.Y. Wang, Q. Wang and C.H. Yang. 2025. Interaction of population density and slope will exacerbate spatiotemporal changes in land use and landscape patterns in mountain city. *Scientific Reports*, 15(1): 3168.

Zhao, R., H. Bu, X. Song and Y. Zhang. 2021. A multivariate analysis of the spatial variations of water quality during high-flow period in the Chaobai River (Beijing, China) restored by reclaimed water. *Water Supply*, 21(6): 3168-3179.

Zheng, Y., J. Wei, W. Zhang, Y. Zhang, T. Zhang and Y. Zhou. 2024. An ensemble model for accurate prediction of key water quality parameters in river based on deep learning methods. *Journal of Environmental Management*, 366: 121932.

Zhu, X., L. Wang, X. Zhang, M. He, D. Wang, Y. Ren and H. Pan. 2022. Effects of different types of anthropogenic disturbances and natural wetlands on water quality and microbial communities in a typical black-odor river. *Ecological Indicators*, 136: 108613.