Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

PREDICTING URBAN WATER QUALITY WITH UBIQUITOUS DATA

[1]DR. B. GOHIN, [2]GUTTULA MANIKANTA

[1](Associative Professor), MCA, Swarnandhra College

[2]MCA, scholar, Swarnandhra College

ABSTRACT

Our everyday lives are greatly impacted by the quality of urban water. Urban water quality predictions aid in the prevention of water contamination and the safeguarding of human health. Predicting urban water quality is difficult because there are many variables, including weather, water consumption patterns, and land uses, that contribute to the non-linear variation in water quality in metropolitan areas. Here, we take a data-driven approach to predicting a station's water quality for the next several hours by combining information from several city-wide sources, including weather, pipe networks, road structures, and points of interest (POIs), with water quality and water hydraulic data reported by existing monitor stations. Using a battery of tests, we first isolate the components with the greatest impact on the water quality in cities. To further integrate these diverse datasets into a single learning model, we next introduce a multi-task multi-view learning approach. We test our technique on real-world datasets, and the results show that our approach is successful and that our method has benefits over other baselines.

1.INTRODUCTION

There are many facets of urban life and human health that are impacted by urban water. There has been a recent uptick in the number of requests for citywide water quality monitoring and prediction systems from city dwellers worried about the state of urban water supplies. Several chemical indices, including residual chlorine, turbidity, and pH, can be utilized as effective indicators of water quality in existing urban water distribution systems. This is important because urban

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

water quality is both "a powerful environmental determinant" and "a foundation for the prevention and control of waterborne diseases" [1].

Numerous water quality monitoring stations have been placed throughout a city's water distribution system to provide the growing demand for water quality data in real-time. Shenzhen, China's water quality monitoring stations are shown in Figure 1. Predicting urban water quality is just as important as monitoring it for many urban aquatic projects. This is because it influences policy decisions (such as when governments issue pollution alerts or carry out pollution controls), informs decision-making within waterworks (like when chlorine levels are pre-adjusted), and provides maintenance recommendations (like when certain pipelines should be replaced).

But, for the reasons listed below, predicting the quality of urban water is very difficult. To begin, water consumption patterns, land use, urban architecture, and weather are only a few of the many variables that affect the non-linear variation in urban water quality across different sites. The three stations' reported water quality indices (RC) show distinct trends, as seen in Figure 1.

Unfortunately, the current methods that rely on hydraulic models to attempt to simulate water quality from a physical and chemical standpoint are severely limited in their ability to do so. The characteristics I use in my models are also not easy to come by, which limits their applicability to different kinds of water distribution networks. Secondly, with all the stations linked together by pipelines, there are a lot of complicated aspects, including pipe network features and point-of-interest distribution, that affect the water quality at various stations. Conventional methods that rely on hydraulic models provide subpar results since they construct individual models for each station without taking into account the geographical connections between them. Therefore, it is not enough to just discover the components that have an impact; another problem is to effectively describe and include such relatedness.

Luckily, in this age of big data, previously unavailable urban data sets (such as weather, points of interest, and road networks) may provide further information to aid in the prediction of urban water quality [3] [4] [5]. One measure of water quality is its temperature; a higher reading indicates

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

higher quality water. Because more people are likely to take showers when the weather is hot, and because higher water use helps keep distribution systems' water quality from becoming worse, this might be the explanation.

In order to make use of the first-of-its-kind data in cities, this study uses a data-driven approach to forecast a station's water quality based on a number of datasets, such as those pertaining to water quality, hydraulics, meteorology, pipe networks, road networks, and points of interest (POIs). The first step in improving urban water quality is to determine which aspects are most important by conducting comprehensive experiments and data analytics involving water quality and several possible variables. Additionally, we provide a new framework called stMTMV that combines data from several domains and allows each station's local and global information to be combined into one learning model [6].

Here is a brief overview of the contributions:

A View Based on Data: We provide a new data-driven method for predicting future water quality at various locations using data from several domains. Furthermore, the method is applicable to a wide variety of urban applications, not just those involving the prediction of water quality, but also to other problems involving the co-prediction of many places.

Contributing to both our application and the overall issue of water quality prediction, we identify influential factors that are both geographically and temporally relevant, such as points of interest (POIs), pipe networks, road networks, and weather conditions.

With the goal of creating a unified learning model, we introduce stMTMV, a unique framework for integrating spatio-temporal urban data from various sources. This model offers a general approach to combining heterogeneous spatio-temporal properties for prediction and can be used for other applications based on spatio-temporal data.

Extensive tests using real-world datasets in Shenzhen, China, allow us to assess our strategy in a real-world setting. In addition to revealing intriguing findings that might improve city life for the better, the results show that our technique outperforms existing baselines including ARMA, Kalman filter, and ANN.

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

Here is the structure of the remaining portion of the paper: The structure of our approach is described in Section 2. The relationships between water quality and data collected from various urban sources are examined in Sections 3 and 4. Assessments and visualisations are provided in Section 6, whereas Section 5 provides the multi-task multi-view learning approach for predicting urban water quality. The relevant study is summarized in part 7, and the concluding part contains the conclusion.

This journal version asserts the following contributions, building upon our earlier work [6]: We began by focusing our attention on the data-driven viewpoint. Our methodology's insights and the results of a correlation study between various data sets and urban water quality were particularly highlighted. Sections 3 and 4 provide the comprehensive correlation analysis. Second, using the data correlation analysis from Section 5.4.1, we were able to improve the task connection computation in our STMTMV model by determining the optimal configuration across different pipe properties. Third, in order to confirm the accuracy of our approach, we ran more thorough trials. As an example, in Section 6.3, we updated the time series prediction baselines to include two more well-known methods (Kalman and ANN). Furthermore, in Section 6.6, we contrasted our method's performance with that of other baselines for each individual station.

2.LITERATURE SURVEY

Karlsson et al. [67] presented work that used the k-NN method to estimate rainfall and runoff in a traditional setting, and their findings were encouraging. Using historical rainfall data, Toth et al. [68] demonstrated that k-NN consistently outperformed alternative time series prediction approaches. In order to calibrate the two-dimensional surface quantity and water quality model, Ostfeld et al. [69] created a hybrid genetic k-Nearest Neighbour method. A directed graph with several layers of nodes (neurons) completely linked to each other is the building block of a natural neural network, which is the basis for Artificial Neural Networks (ANNs) [65]. Because of their versatility and effectiveness, neural networks have found extensive use in many different fields. One example is the work of Moradkhani et al. [70], who showed how an RBF network-based hourly streamflow forecasting technique outperformed previous

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

numerical prediction methods. Additionally, Kalin's [44] study used ANN to forecast watershed water quality indices. One common kind of supervised learning model that examines data for regression and classification is the Support Vector Machine (SVM) [71].It was also used to solve prediction difficulties in aquatic investigations [64]. As an example, the problem of flood forecasting was tackled by Liong et al. [72] by use of Support Vector Regression (SVR), an extension of Support Vector Machines (SVM). The water quality prediction issue in Guangzhou's Liuxi River was addressed in another study by Xiang et al. [73] using an LS-SVM model. But none of these methods have ever been used in urban settings, which is very different from what we do. Also, those current methods only handle data from one source, thus they aren't very good at combining data from other sources. As a result, their use in city settings is limited.

3. EXISTING SYSTEM

Numerous data-driven analyses of water quality issues have been conducted in the field of environmental science. These analyses have covered a wide range of topics, including physical process analysis in river basins and concurrent input and output time series [64] [65]. These researches make use of neural network models (e.g., ANN) and instance-based learning models (e.g., KNN). Typically, there are three main types of data-driven methodologies used in environmental science: Instance-based Learning (IBL), Artificial Neural Network (ANN), and Support Vector Machine (SVM).

IBLs are a class of learning algorithms that represent a decision-making issue using specific examples of training data that are considered crucial for testing the model's accuracy [66]. Due to its simplicity and very strong performance in reality, k-Nearest Neighbors (k-NN) is
commonly cited as an example of IBL.

Drawbacks:

The system is developed only for use with multi-task multi-view learning approaches. ¬ Instance-based learning models (IBL) are a class of learning algorithms that simulate a decision issue using key examples of training data.

Idea for a system:

To co-predict future water quality across stations using data from many domains, we

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

provide a unique data-driven technique. The method is also applicable to various urban applications involving multi-locations-based co-prediction, not only urban water quality prediction.

We identify features that are related to both space and time, such as points of interest (POIs), pipe networks, and road networks, as well as features that are related to both space and time, such as the time of day, weather, and water hydraulics. These features contribute to both our application and the overall problem of water quality prediction.

Unified Learning Model: With the goal of integrating various sources of spatio-temporal urban data, we introduce a new framework called stMTMV. This framework offers a general way to combine heterogeneous spatio-temporal properties for prediction, and it can be used for other applications that revolve around spatio-temporal data.

Extensive tests using real-world datasets in Shenzhen, China, assess our technique. The findings show that our technique outperforms existing baselines like ARMA, Kalman filter, and ANN, and they also unveil intriguing findings that have the potential to improve city life for the better.

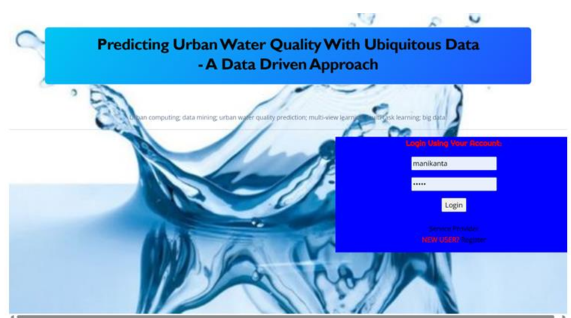
One benefit is that we get data on the city's water quality from fifteen monitoring sites in Shenzhen City every five minutes. It is made up of ph, turbidity, and residual chlorine. Since RC is the most significant and effective metric for water quality in the present urban water distribution system, it is exclusively used as an index for water quality in this work.

Facts about water: Thirteen locations measure flow and fourteen sites measure pressure; both types of hydraulic data are gathered every five minutes.

4. OUTPUT SCREENS

Remote User:

User Login: In this module the user can login to the page by giving details.



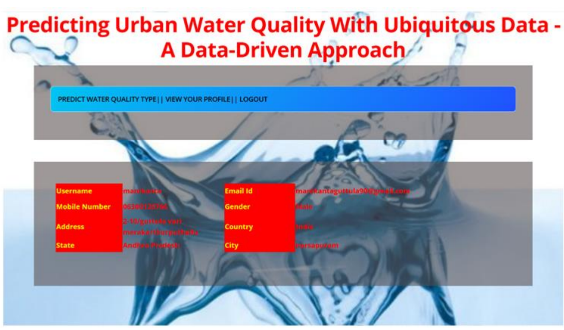
New User Registration: In this modules the new users can register with giving complete details.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

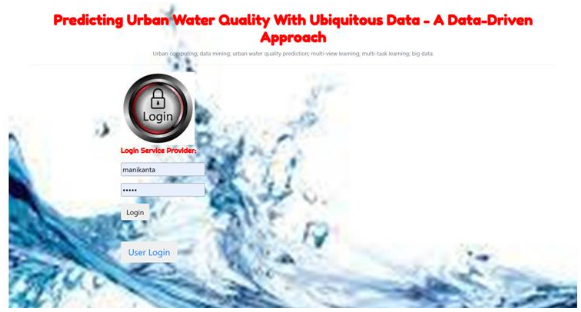
User Profile: In this module the users can see their complete profile.



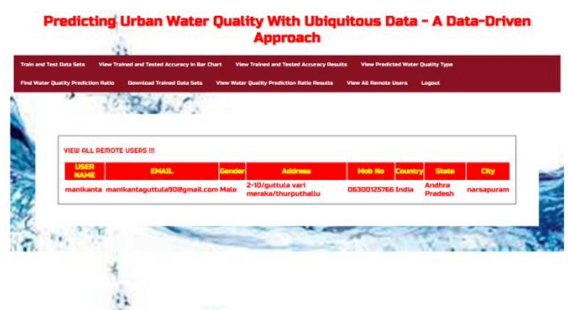
Predicting water Quality: In this module the users can check the water quality by giving the details.

Service Provider:

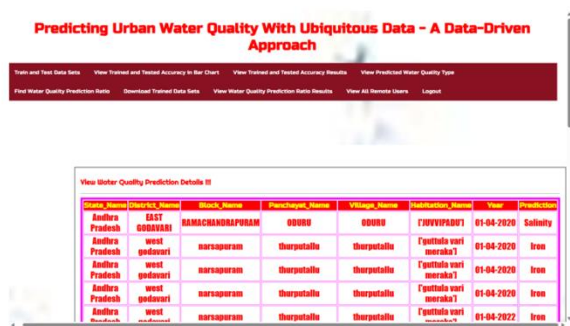
Admin Login: In this module the admin can login with details.



All Remote Users: Here the admin can see all registered users.



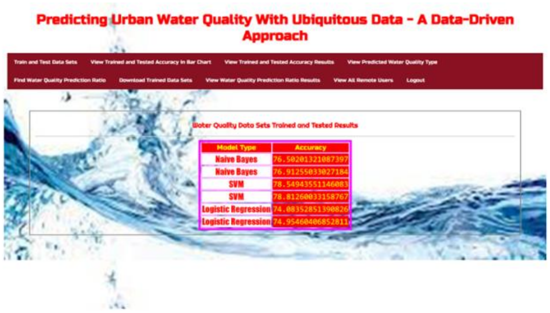
Predicted Water Quality Type: Here the admin can see the predicted water quality type data.



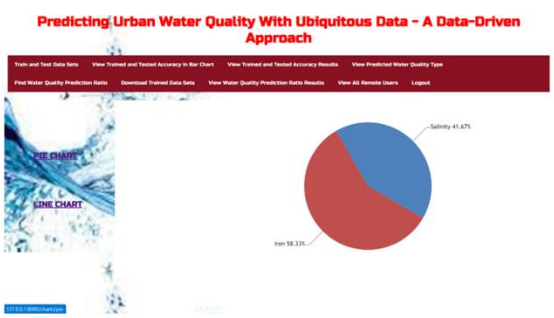
State_Name	District_Name	Block_Name	Panchayat_Name	Village_Name	Habitation_Name	Year	Prediction
Andhra Pradesh	EAST GODAVARI	RAMACHANDRAPURAM	ODURU	ODURU	['JUVVIPADU']	01-04-2020	Salinity
Andhra Pradesh	west godavari	narsapuram	thurputallu	thurputallu	['guttula vari meraka']	01-04-2020	Iron
Andhra Pradesh	west godavari	narsapuram	thurputallu	thurputallu	['guttula vari meraka']	01-04-2020	Iron
Andhra Pradesh	west godavari	narsapuram	thurputallu	thurputallu	['guttula vari meraka']	01-04-2020	Iron
Andhra	west	narsapuram	thurputallu	thurputallu	['guttula vari	01-04-2022	Iron

Train and Test data Sets: Here the admin can see the accuracy of trained and tested results.

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024



Water Prediction Ratio Results: Here the admin can see the Water Prediction Ratio results in pie chart.



5. CONCLUSION

This research introduces a new data-driven method for predicting a station's water quality utilizing a combination of several urban data sources. Using water quality data from Shenzhen and other urban statistics, we assess our method. The efficiency and efficacy of our method are shown by the experimental findings. With regard to the root-mean-squared error (RMSE) measure, our method surpasses other conventional time series prediction models (ARMA, Kalman) as well as the time-honored RC decay model [2]. Also, because our method is bipartite, we have enough data from tests and analyses to back up the claims made by each part. Sections 3 and 4 conduct in-depth tests and analyses to determine which variables have the most impact on urban water quality. This investigation forms the basis of the first component, influential factors identification. A second framework that incorporates both multi-view and multi-task learning is the spatiotemporal multi-view multi-task learning (STMTMV) framework. Results from the trials demonstrate that compared to LR (single-task technique) and t-view (single-view) and s-view (single-view), STMTMV surpasses all three with a predicted accuracy of around 85% over the following 1-4 hours. You can get the published code at: https://www.microsoft.com/en-us/research/publication/urbanwater-quality-prediction-based-multi-task-multi-view-learning-2/. We want to use a small number of water quality monitoring stations in the future to address the issues with water quality inference in urban water distribution systems.

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

6.REFERENCES

[1] W. H. Organization, Guidelines for drinking-water quality, 2004, vol. 3.

[2] L. A. Rossman, R. M. Clark, and W. M. Grayman, "Modeling chlorine residuals in drinking-water distribution systems," Journal of environmental engineering, vol. 120, no. 4, pp. 803–820, 1994.

[3] Y. Zheng, "Methodologies for cross-domain data fusion: An overview," IEEE Transactions on Big Data, vol. 1, no. 1, pp. 16–34, 2015.

[4] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," ACM Transactions on Intelligent Systems and Technology, vol. 5, no. 3, pp. 38:1–38:55, 2014.

[5] Y. Zheng, H. Zhang, and Y. Yu, "Detecting collective anomalies from multiple spatio- temporal datasets across different domains," 2015.

[6] Y. Liu, Y. Zheng, Y. Liang, S. Liu, and D. S. Rosenblum, "Urban water quality prediction based on multi-task multi-view learning," in Proceedings of the International Joint Conference on Artificial Intelligence, 2016.

[7] H. Cohen, "Free chlorine testing," http://www.cdc.gov/safewater/chlorineresid ualtesting.html, 2014, accessed on 5 August 2016.

[8] B. D. Barkdoll and H. Didigam, "Effect of user demand on water quality and hydraulics of distribution systems," in Proceedings of the World Water and Environmental Resources Congress, 2003.

[9] P. Castro and M. Neves, "Chlorine decay in water distribution systems case study–lousada network," Electronic Journal of Environmental, Agricultural and Food Chemistry, vol. 2, no. 2, pp. 261–266, 2003.

[10] L. W. Mays, Water distribution system handbook, 1999.

[11] L. A. Rossman and P. F. Boulos, "Numerical methods for modeling water quality in distribution systems: A comparison," Journal of Water Resources planning and management, vol. 122, no. 2, pp. 137–146, 1996.

[12] W. M. Grayman, R. M. Clark, and R. M. Males, "Modeling distribution system water quality: dynamic approach," Journal of Water

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue 2, 2024

Resources 56 Planning and Management, vol. 114, no. 3, pp. 295–312, 1988.

[13] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in Proceedings of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2003, pp. 2–11.

[14] G. Luo, K. Yi, S.-W. Cheng, Z. Li, W. Fan, C. He, and Y. Mu, "Piecewise linear approximation of streaming time series data with max-error guarantees," in Proceedings of the IEEE International Conference on Data Engineering, 2015, pp. 173–184.

[15] E. O. Brigham and E. O. Brigham, The fast Fourier transform. PrenticeHall Englewood Cliffs, NJ, 1974, vol. 7.

[16] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer," 1997.

[17] M. W. LeChevallier, T. Evans, and R. J. Seidler, "Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water." Applied and environmental microbiology, vol. 42, no. 1, pp. 159–167, 1981.

[18] L. Monteiro, D. Figueiredo, S. Dias, R. Freitas, D. Covas, J. Menaia, and S. Coelho, "Modeling of chlorine decay in drinking water supply systems using epanet msx," Procedia Engineering, vol. 70, pp. 1192–1200, 2014.

[19] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, no. 1, pp. 49–67, 2006.

[20] W. Zhang, K. Zhang, P. Gu, and X. Xue, "Multi-view embedding learning for incompletely labeled data," in Proceedings of the International Joint Conference on Artificial Intelligence, 2013, pp. 1910–1916.

[21] Y. Liu, L. Zhang, L. Nie, Y. Yan, and D. S. Rosenblum, "Fortune teller: Predicting your career path," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 201–207.

[22] S. Sun, "A survey of multi-view machine learning," Neural Computing and Applications, vol. 23, no. 7-8, pp. 2031–2038, 2013.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue 2, 2024

[23] C. Xu, D. Tao, and C. Xu, “A survey on multi-view learning,” arXiv 57 preprint arXiv:1304.5634, 2013.

[24] J. Zhang and J. Huan, “Inductive multi-task learning with multiple view data,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining, 2012, pp. 543–551.

[25] L. Zhao, Q. Sun, J. Ye, F. Chen, C.-T. Lu, and N. Ramakrishnan, “Multitask learning for spatio-temporal event forecasting,” in Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1503–1512.