BIBLIOGRAPHY

[1] P. Bazaanah and R. A. Mothapo, *Sustainability of drinking water and sanitation delivery systems in rural communities of the Lepelle Nkumpi Local Municipality, South Africa*, vol. 26, no. 6. Springer Netherlands, 2024. doi: 10.1007/s10668-023-03190-4.

[2] United Nations Children's Fund (UNICEF) and World Health Organization (WHO), "Summary progress update 2021 : SDG 6 — water and sanitation for all," *UN-Water Integr. Monit. Initiat.*, pp. 1–58, 2021, [Online]. Available: https://www.unwater.org/new-data-on-global-progress-towards-ensuring-water-and-sanitation-for-all-by-2030/

[3] WHO, *Global water, sanitation and hygiene Annual report 2022*. 2022.

[4] A. A. Aydamo, S. R. Gari, and S. T. Mereta, "Access to Drinking Water, Sanitation, and Hand Hygiene Facilities in the Peri-Urban and Informal Settlements of Hosanna Town, Southern Ethiopia," *Environ. Health Insights*, vol. 17, 2023, doi: 10.1177/11786302231193604.

[5] M. Anindita, A. Rahman, M. A. Alim, C. Xiong, S. Hossain, and A. Sathasivan, "Microbial ecology of harvested rainwater: Assessing quality, antimicrobial resistance and geographical variation," *J. Clean. Prod.*, vol. 486, p. 144439, Jan. 2025, doi: 10.1016/J.JCLEPRO.2024.144439.

[6] World Health Organization, *Guidelines for drinking-water quality: fourth edition incorporating the first addendum. Geneva: World Health Organization*, vol. 55. 2017. doi: 10.5005/jp/books/11431_8.

[7] A. C. C. Fortes, P. R. G. Barrocas, and D. C. Kligerman, "Water quality indices: Construction, potential, and limitations," *Ecol. Indic.*, vol. 157, no. July, 2023, doi: 10.1016/j.ecolind.2023.111187.

[8] S. Zhong *et al.*, "Machine Learning: New Ideas and Tools in Environmental Science and Engineering," *Environ. Sci. Technol.*, vol. 55, no. 19, pp. 12741–12754, 2021, doi: 10.1021/acs.est.1c01339.

[9] N. T. Anh, L. D. Can, N. T. Nhan, B. Schmalz, and T. Le Luu, "Influences of key factors on river water quality in urban and rural areas: A review," *Case Stud. Chem. Environ. Eng.*, vol. 8, no. July, p. 100424, 2023, doi: 10.1016/j.cscee.2023.100424.

[10] G. Özsezer and G. Mermer, "Prediction of drinking water quality with machine learning models: A public health nursing approach," *Public Health Nurs.*, vol. 41, no. 1, pp. 175–191, 2024, doi: 10.1111/phn.13264.

[11] D. Chatterjee, P. Ghosh, A. Banerjee, and S. S. Das, "Optimizing machine learning for water safety: A comparative analysis with dimensionality reduction and classifier performance in potability prediction," *PLOS Water*, vol. 3, no. 8 August, pp. 1–25, 2024, doi: 10.1371/journal.pwat.0000259.

[12] X. Li *et al.*, "A Fusion XGBoost Approach for Large-Scale Monitoring of Soil Heavy Metal in Farmland Using Hyperspectral Imagery," *Agronomy*, vol. 15, no. 3, 2025, doi: 10.3390/agronomy15030676.

[13] A. Krishnaraj and R. Honnasiddaiah, "Remote sensing and machine learning based framework for the assessment of spatio-temporal water quality in the Middle Ganga Basin," *Environ. Sci. Pollut. Res.*, vol. 29, no. 43, pp. 64939–

64958, 2022, doi: 10.1007/s11356-022-20386-9.
[14] A. K. Rai and S. S. Kumar, "Optimization-driven XGBoost model with metaheuristic algorithms for assessing compressive strength of high-performance concrete," *Asian J. Civ. Eng.*, vol. 26, no. 8, pp. 3401–3421, 2025, doi: 10.1007/s42107-025-01379-8.
[15] X. Liu, Y. Hu, X. Li, R. Du, Y. Xiang, and F. Zhang, "An Interpretable Model for Salinity Inversion Assessment of the South Bank of the Yellow River Based on Optuna Hyperparameter Optimization and XGBoost," *Agronomy*, vol. 15, no. 1, 2025, doi: 10.3390/agronomy15010018.
[16] K. Kelayakan *et al.*, "Classification of Drinking Water Source Suitability in West Java Using XGBoost and Cluster Analysis Based on SHAP Values *," vol. 8, no. 2, pp. 202–214, 2024.
[17] I. Marihot, R. Hartanto, and R. Halim, "ScienceDirect ScienceDirect Comparative Analysis of XGBoost , Random Forest , and Comparative Analysis of XGBoost , Random Forest , and Logistic Regression for Classifying Jakarta ' s Air Pollution Index Logistic Regression for Classifying Jakarta ' s Air," *Procedia Comput. Sci.*, vol. 269, pp. 108–120, 2025, doi: 10.1016/j.procs.2025.08.264.
[18] C. G. L. Pringandana and K. Kusnawi, "A Comparative Analysis of Hyperparameter-Tuned XGBoost and LightGBM for Multiclass Rainfall Classification in Jakarta," *J. Tek. Inform.*, vol. 6, no. 4, pp. 2467–2483, 2025, doi: 10.52436/1.jutif.2025.6.4.4965.
[19] T. Toharudin *et al.*, "Boosting Algorithm to Handle Unbalanced Classification of PM2.5Concentration Levels by Observing Meteorological Parameters in Jakarta-Indonesia Using AdaBoost, XGBoost, CatBoost, and LightGBM," *IEEE Access*, vol. 11, no. March, pp. 35680–35696, 2023, doi: 10.1109/ACCESS.2023.3265019.
[20] G. Kirubavathi, C. A. Dhanyatha Shree, and R. Nithish Kumar, "Enhanced DNS Cybersecurity: Optimization of XGBoost Using Automated Hyperparameter Tuning via Optuna," *4th IEEE Int. Conf. Mob. Networks Wirel. Commun. ICMNWC 2024*, 2024, doi: 10.1109/ICMNWC63764.2024.10872404.
[21] H. Incekara, I. Hakkı, C. Mücahid, M. Saritas, and M. Koklu, "Classification of almond kernels with optuna hyper-parameter optimization using machine learning methods," 2025.
[22] B. A. Dada, N. I. Nwulu, and S. O. Olukanmi, "Bayesian optimization with Optuna for enhanced soil nutrient prediction: a comparative study with genetic algorithm and particle swarm optimization," *Smart Agric. Technol.*, vol. 12, no. June, p. 101136, 2025, doi: 10.1016/j.atech.2025.101136.
[23] Y. Zhao *et al.*, "Interpretable data-driven chemometric approach for predicting non-optically active water quality parameters using ultraviolet-visible-near infrared absorption spectroscopy and physical-chemical measurements," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 331, p. 125768, Apr. 2025, doi: 10.1016/J.SAA.2025.125768.
[24] K. Shaw, M. Vogel, N. Andriessen, T. Hardeman, C. C. Dorea, and L. Strande, "Towards globally relevant, small-footprint dewatering solutions: Optimal conditioner dose for highly variable blackwater from non-sewered sanitation," *J. Environ. Manage.*, vol. 321, no. May, p. 115961, 2022, doi:

10.1016/j.jenvman.2022.115961.
[25] L. Andrade, J. O'Dwyer, E. O'Neill, and P. Hynds, "Surface water flooding, groundwater contamination, and enteric disease in developed countries: A scoping review of connections and consequences," *Environ. Pollut.*, vol. 236, pp. 540–549, 2018, doi: 10.1016/j.envpol.2018.01.104.
[26] A. Kelly, E. P. Widiyanto, and H. Irsyad, "Sistem Cerdas Pengkategorian Surat Undangan Elektronik Tender Pekerjaan Dengan AutoML," vol. 5, no. 4, pp. 211–221, 2024, doi: 10.47065/bit.v5i2.1501.
[27] M. A. Pienaar and K. D. Naidoo, "Classification and predictive models using supervised machine learning: A conceptual review," *South. African J. Crit. Care*, vol. 41, no. 1, pp. 2–8, 2025, doi: 10.7196/SAJCC.2025.v411.2937.
[28] F. Emmert-Streib and M. Dehmer, "Taxonomy of machine learning paradigms: A data-centric perspective," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 12, no. 5, 2022, doi: 10.1002/widm.1470.
[29] J. Peng, E. C. Jury, P. Dönnes, and C. Ciurtin, "Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges," *Front. Pharmacol.*, vol. 12, 2021, doi: 10.3389/fphar.2021.720694.
[30] Z. Karami Lawal *et al.*, "Optimized Ensemble Methods for Classifying Imbalanced Water Quality Index Data," *IEEE Access*, vol. 12, pp. 178536–178551, 2024, doi: 10.1109/ACCESS.2024.3502361.
[31] K. Maharana, S. Mondal, and B. Nemade, "A review : Data pre-processing and data augmentation techniques," vol. 3, no. April, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
[32] Q. A. Hidayaturrohman and E. Hanada, "Impact of Data Pre-Processing Techniques on XGBoost Model Performance for Predicting All-Cause Readmission and Mortality Among Patients with Heart Failure," *BioMedInformatics*, vol. 4, no. 4, pp. 2201–2212, 2024, doi: 10.3390/biomedinformatics4040118.
[33] D. F. Wicaksono, R. S. Basuki, and D. Setiawan, "Peningkatan Performa Model Machine Learning XGBoost Classifier melalui Teknik Oversampling dalam Prediksi Penyakit AIDS," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 736, 2024, doi: 10.30865/mib.v8i2.7501.
[34] N. Mungoli, "Adaptive Ensemble Learning: Boosting Model Performance through Intelligent Feature Fusion in Deep Neural Networks," 2023, [Online]. Available: http://arxiv.org/abs/2304.02653
[35] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, and S. M. Almufti, "eXtreme Gradient Boosting Algorithm with Machine Learning: a Review," *Acad. J. Nawroz Univ.*, vol. 12, no. 2, pp. 320–334, 2023, doi: 10.25007/ajnu.v12n2a1612.
[36] M. Zhu *et al.*, "A review of the application of machine learning in water quality evaluation," *Eco-Environment Heal.*, vol. 1, no. 2, pp. 107–116, 2022, doi: 10.1016/j.eehl.2022.06.001.
[37] T. Kavzoglu and A. Teke, "Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost)," *Arab. J. Sci. Eng.*, vol. 47, no. 6, pp. 7367–7385, 2022, doi: 10.1007/s13369-022-06560-8.

[38] N. MacNell *et al.*, “Implementing machine learning methods with complex survey data: Lessons learned on the impacts of accounting sampling weights in gradient boosting,” *PLoS One*, vol. 18, no. 1 January, 2023, doi: 10.1371/journal.pone.0280387.
[39] A. Mehdary, A. Chehri, A. Jakimi, and R. Saadane, “Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection,” *Sensors*, vol. 24, no. 4, 2024, doi: 10.3390/s24041230.
[40] B. Bischl *et al.*, “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 13, no. 2, 2023, doi: 10.1002/widm.1484.
[41] T. Chen and C. Guestrin, “XGBoost : A Scalable Tree Boosting System”.
[42] R. Y. Pratama and A. Baita, “Prediksi Stunting pada Anak Balita Menggunakan Algoritma Extreme Gradient Boosting dan Bayesian Optimization,” vol. 7, no. 2, pp. 175–189, 2025.
[43] K. Fujita, T. Hiyama, K. Wada, and T. Aihara, “Machine learning – based muscle mass estimation using gait parameters in community-dwelling older adults : A cross-sectional study”.
[44] T. R. Noviandy, R. Setiawan, and R. Sufri, “Optimizing Machine Learning Models for Cardiovascular Drug Candidate Screening with XGBoost and Successive Halving,” vol. 13, no. 2, pp. 41–50, 2025.
[45] F. Ad and D. Kurniawan, “Application of ADASYN and Optuna in the XGBoost Algorithm for Stunting Detection,” vol. 10, no. 1, pp. 1006–1014, 2026.
[46] S. Widodo, H. Brawijaya, and S. Samudi, “Stratified K-fold cross validation optimization on machine learning for prediction,” *Sinkron*, vol. 7, no. 4, pp. 2407–2414, 2022, doi: 10.33395/sinkron.v7i4.11792.
[47] T. Kavzoglu and A. Teke, “Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost),” *Bull. Eng. Geol. Environ.*, vol. 81, no. 5, 2022, doi: 10.1007/s10064-022-02708-w.
[48] S. Shekhar, A. Bansode, and A. Salim, “A Comparative study of Hyper-Parameter Optimization Tools,” *2021 IEEE Asia-Pacific Conf. Comput. Sci. Data Eng. CSDE 2021*, 2021, doi: 10.1109/CSDE53843.2021.9718485.
[49] Z. Wang, H. Li, and Q. Wang, “An Approach to Truck Driving Risk Identification: A Machine Learning Method Based on Optuna Optimization,” *IEEE Access*, vol. 13, pp. 42723–42732, 2025, doi: 10.1109/ACCESS.2025.3547445.
[50] R. Manoj, “Optuna — An Automatic Hyperparameter Optimization Framework.” [Online]. Available: https://medium.com/@publiciscommerce/optuna-an-automatic-hyperparameter-optimization-framework-f64638621ff7
[51] L. H. Lai *et al.*, “The Use of Machine Learning Models with Optuna in Disease Prediction,” *Electron.*, vol. 13, no. 23, 2024, doi: 10.3390/electronics13234775.
[52] S. Desire, K. Sande, and M. Theobald, *Regression vs . Medical LLMs : A Comprehensive Study for CVD and Mortality Risk Prediction*, vol. 1, no. 1. Association for Computing Machinery, 2026.

[53] P. Mithulesh, B. Agilan, S. K. N. R. S, and K. Vani, "Drought Prediction using Landsat-8 Images and Remote Sensing," no. Acrs, pp. 1–14, 2024.

[54] Z. Zhao, G. Zhong, F. Diao, and P. Ding, "A Feature Engineering and XGBoost Framework for Prediction of TOC from Conventional Logs in the Dongying Depression , Bohai Bay Basin," pp. 1–22, 2026.

[55] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.

[56] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," *Lect. Notes Networks Syst.*, vol. 724 LNNS, pp. 15–25, 2023, doi: 10.1007/978-3-031-35314-7_2.

[57] S. Tewari, "Deploying Machine Learning Models : A Comparative Study of Flask , FastAPI , and Streamlit," pp. 1–7, 2025, doi: 10.13140/RG.2.2.24413.06880.

[58] N. Sarangpure, V. Dhamde, A. Roge, J. Doye, S. Patle, and S. Tamboli, "Automating the Machine Learning Process using PyCaret and Streamlit".

[59] E. Heitkemper, S. Hulse, and B. Bekemeier, "The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) Dashboard for Health Equity in Rural Public Health : Usability Evaluation Corresponding Author :," vol. 11, pp. 1–12, 2024, doi: 10.2196/51666.