

ICRAMCS 2026

THE EIGHTH EDITION OF THE INTERNATIONAL CONFERENCE ON
RESEARCH IN APPLIED MATHEMATICS AND COMPUTER SCIENCE

April 23-24-25, 2026 | Marrakech, Morocco



Improving Imbalanced Water Quality Classification Using SMOTE

Communication Info

Authors:

Norashikin NASARUDDIN^{1,2}
Nurulkamal MASSERAN²
Wan Mohd Razi IDRIS²
Ahmad Zia UL-SAUFIE³

¹Universiti Teknologi Mara
(UiTM) Kedah Branch, 08400,
Merbok, Kedah, Malaysia

²Universiti Kebangsaan
Malaysia, 43600, Bangi,
Selangor, Malaysia

³Universiti Teknologi Mara
(UiTM), 40450, Shah Alam,
Selangor, Malaysia

Keywords:

- (1) Water Quality
- (2) SMOTE
- (3) Class Imbalance

Abstract

Class imbalance reduces predictive reliability in multi-class water quality classification, particularly when minority pollution levels are underrepresented. Synthetic Minority Oversampling Technique (SMOTE) is widely used to address this issue [1], though its performance depends on classifier selection [2]. This study compares SMOTE, ADASYN, and Borderline-SMOTE using imbalanced river data from Kedah, Malaysia (2020–2023), where 74% of samples belong to the majority class. Decision Tree, Random Forest, and XGBoost were evaluated, as ensemble models are effective under skewed distributions [3]. Performance was measured using F1-score and sensitivity, recommended for imbalanced learning [4]. Without resampling, minority F1-scores were below 35%; oversampling significantly improved minority detection, with SMOTE–XGBoost yielding the most consistent gains. These results support the use of carefully selected oversampling strategies in environmental monitoring systems [5].

© ICRAMCS 2026 Proceedings ISSN: 2605-7700

References

- [1] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP, SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, 16, 2002, 321–357.
- [2] Krawczyk B, Learning from imbalanced data: open challenges and future directions, *Progress in Artificial Intelligence*, 11, 2022, 1–22.
- [3] Chen T, Guestrin C, XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785–794.
- [4] Tharwat A, Classification assessment methods, *Applied Computing and Informatics*, 17, 2021, 168–192.
- [5] Xu L, Jing X, Yang Y, National surface water quality classification evaluation based on SMOTE-GA-CatBoost method, *China Environmental Science*, 43, 2023, 3848–3856.