Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.35794

Development of a Machine Learning-Based Seawater Intrusion Vulnerability Assessment Framework Incorporating Anthropogenic Hydrogeological Factors

LINN KIM[1], TAEYONG KIM[1], SUNGTAEK LIM[2] AND MINJUNE YANG[1,3]

[1]Pukyong National University
[2]Gyeongnam Regional Headquarters, Korea Rural Community Corporation
[3]Wible Co., Ltd.

Seawater intrusion (SWI) in coastal aquifers is a hydrogeological phenomenon that displaces the freshwater–saltwater interface, degrading groundwater quality. To assess SWI vulnerability, various index-based models have been developed, with GALDIT being widely used. However, the conventional GALDIT model applies pre-assigned expert-judgment weights, limiting its ability to reflect spatial heterogeneity in factor importance. Furthermore, anthropogenic hydrological factors such as groundwater abstraction are not integrated into this framework. This study aims to develop an ML-based framework to recalibrate GALDIT weights from data and integrate anthropogenic hydrological factors through a stepwise variable expansion approach. The study area is Tongyeong, South Korea, a coastal region characterized by a complex coastline. Both groundwater and subsurface seawater abstraction are in operation in the region. To investigate this, three stages were applied:the conventional six GALDIT parameters were recalibrated (Stage 1); groundwater pumping volume was incorporated to account for anthropogenic hydrological stress (Stage 2); and seawater abstraction was added as a mitigation variable to capture saltwater diffusion reduction (Stage 3). Random Forest (RF), XGBoost (XGB), and LightGBM (LGBM) algorithms were employed for model training, with Stratified Out-of-Fold cross-validation applied to address data imbalance and evaluate generalization performance. Model performance was evaluated against field-measured electrical conductivity values. model yielded $R^2 = 0.027$, MAE = 2711.67, and RMSE = 6506.80, while ML algorithms showed higher accuracy: XGB ($R^2 = 0.635$), RF ($R^2 = 0.562$), and LGBM ($R^2 = 0.244$). The XGB-based model improved across stages: Stage 1 ($R^2 = 0.635$, MAE = 947.41, RMSE = 3047.72), Stage 2 ($R^2 = 0.669$, MAE = 939.28, RMSE = 2903.15), and Stage 3 ($R^2 = 0.774$, MAE = 780.54, RMSE = 2396.33). These results show that ML-based recalibration of GALDIT weights and stepwise integration of anthropogenic factors improve SWI vulnerability assessment. The proposed model provides a quantitative management tool that can be implemented for identifying vulnerable zones in coastal aquifers.