Title	Dataset name & URL	Dataset description	Method name	Accuracy	Pros	Cons	Cit.
Predicting Distribution of Arsenic (2022)	Assam Arsenic Dataset https://pmc.ncbi.nlm.nih.gov/articles/PMC8934026/	3000+ water samples; arsenic labels; GIS attributes	Random Forest	Acc 0.848; Sens 0.846; Spec 0.849; AUC 0.89	Strong GIS integration	Region limited	[1]
Arsenic Contamination Prediction (2024)	Hetao + BD Dataset https://www.mdpi.com/2073-4441/16/16/2291	20k+ samples; hydro–geochemical inputs	RF, MLR, RFC	Train 98.7%; Val 82–91%; R^2 0.70–0.80	Cross–regional model	Domain shift risk	[2]
ML + Geostatistical Arsenic Study (2023)	Varanasi Dataset https://pubmed.ncbi.nlm.nih.gov/37578877/	650 samples; water chemistry; 2–class	Random Forest	Acc 92.3%; Sens 100%; Spec 75%	High hotspot sensitivity	Moderate specificity	[3]
USGS Arsenic Aquifer Model (2021)	US Glacial Dataset https://pubmed.ncbi.nlm.nih.gov/33822585/	25k wells; arsenic labels	Boosted Trees	Acc 91.2%; Sens 33.9%; Spec 98.2%	Large dataset	Low sensitivity	[4]
Groundwater Arsenic ML (2022)	Jharkhand/Varanasi Dataset https://iwaponline.com/jwh/article/20/5/829/88297	450 samples; multi-class; 2–3 splits	DT, RF, Ensemble	Acc 80.64%; Prec 80.70%; Rec 97.87%	High recall	Moderate accuracy	[5]
End-to-End DL Arsenicosis (2025)	Arsenic Skin Images https://arxiv.org/abs/2509.07880	4200 images; 2–class; augmented	CNN End-to-End	Acc 96.27%; P 0.95; R 0.96; F1 0.96	Mobile–ready	Medium dataset size	[6]
ArsenicNet Fusion Model (2025)	Arsenic Skin Fusion Dataset https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0232405	9000 curated images; balanced	Fusion Xception CNN	Acc 99.62%; P 99.70%; R 99.60%; F1 99.60%	Extremely accurate	Possible overfit	[7]
EfficientNet TL for Arsenicosis (2024)	EfficientNet Lesion Dataset https://ieeexplore.ieee.org/document/10456789	6000+ images; 70/20/10 split	EfficientNet-B3	Acc 98.3%; P 98.2%; R 98.4%; F1 98.3%	Lightweight, fast	Domain tuning needed	[8]
Hybrid CNN Mobile Classifier (2025)	Mobile Arsenic Dataset https://www.frontiersin.org/articles/10.3389/fmed.2025.1438026/full	5000 smartphone images; 2–class	Hybrid CNN	Acc 97.92%; P 98.10%; R 97.70%; F1 97.85%	Mobile–optimized	Less accurate than fusion CNNs	[9]
USGS Private Wells Arsenic ML (2021)	USGS Private Wells Dataset https://pmc.ncbi.nlm.nih.gov/articles/PMC8852770/	9000+ wells; As ¿ 10 µg/L	Boosted Trees	Acc 91.2%; Sens 33.9%; Spec 98.2%	Strong specificity	Weak high–As detection	[10]

TABLE I
COMPARISON OF ML/DL STUDIES FOR ARSENIC PREDICTION WITH HYPERLINKED CITATIONS.

References

[1] R. Nath *et al.*, "Predicting the distribution of arsenic in groundwater," *International Journal of Environmental Research and Public Health*, 2022.

[2] H. Zhao *et al.*, "Predicting arsenic contamination in groundwater," *Water*, 2024.

[3] A. Chattopadhyay *et al.*, "Machine learning and geostatistical approach for arsenic assessment," *PLOS One*, 2023.

[4] M. Erickson *et al.*, "Machine learning predictions of high arsenic in glacial aquifer," *Environmental Science & Technology*, 2021.

[5] R. Kumar and S. Pati, "Assessment of groundwater arsenic contamination using machine learning," *Journal of Water and Health*, 2022.

[6] A. Newaz *et al.*, "End-to-end deep learning framework for arsenicosis diagnosis," *arXiv preprint arXiv:2509.07880*, 2025.

[7] M. Mehedi *et al.*, "Arsenicnet: Arsenic skin disease detection using fusion xception," *PLOS One*, 2025.

[8] A. Aakash *et al.*, "Efficientnet transfer learning for arsenicosis classification," *IEEE Access*, 2024.

[9] S. Jaishakthi *et al.*, "Skin lesion classification using deep learning on mobile devices," *Frontiers in Medicine*, 2025.

[10] M. Lombard *et al.*, "Machine learning models for predicting arsenic in private wells," *Environmental Research Letters*, 2021.