BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Michael Omar

NGHIÊN CỨU PHÁT TRIỂN PHƯƠNG PHÁP HỌC MÁY KẾT HỢP THÔNG TIN KHÔNG GIAN CHO BÀI TOÁN PHÂN LOẠI CHẤT LƯỢNG NƯỚC NGẦM

TÓM TẮT LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Hà Nội - 2025

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Người hướng dẫn khoa học:

1. Người hướng dẫn 1: PGS. TS NGUYỄN LONG GIANG, VAST
2. Người hướng dẫn 2: PGS. TS TRẦN THỊ NGÂN, VNUIS

Phản biện 1: PGS. TS. Pham Van Hai

Phản biện : PGS. TS. Pham Van Cuong

Phản biện : PGS. TS. Nguyen Ha Nam

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi...giờ..14hr.,ngày... tháng. 01.....năm 2026.

Có thể tìm hiểu luận án tại:

1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

Mở đầu

1. Tính cấp thiết của luận văn : Nguồn nước ngầm ở Đông Nam Á đang chịu áp lực từ sự tăng trưởng dân số, đô thị hóa và biến đổi khí hậu. Luận văn này tập trung vào phân loại khả năng uống được của nước ngầm bằng ML/DL (PSO-SCNN, CNN-GIS, AI-LGBM) và GIS để cải thiện độ chính xác.

2. Mục tiêu nghiên cứu của luận văn : Nâng cao phân loại nước ngầm ở Việt Nam và Odisha sử dụng AI-LGBM, PSO-SCNN và CNN-GIS, làm cơ sở tham khảo cho độ chính xác và khả năng tổng quát cao hơn.

3. Đối tượng và phạm vi nghiên cứu : Tập trung vào Đồng bằng sông Cửu Long và Odisha với dữ liệu vật lý hóa học và không gian, chia thành 70/15/15 (train/val/test) và phân loại là "Xuất sắc," "Tốt," "Kém." Kiểm chứng qua phương pháp k-fold và so sánh với cơ sở.

4. Phương pháp và nội dung nghiên cứu : Phát triển và so sánh AI-LGBM, PSO-SCNN và CNN-GIS với DT/SVM/RF sử dụng các bộ dữ liệu từ Việt Nam và Ấn Độ; đánh giá bằng độ chính xác, độ nhạy, độ đặc hiệu, F1, AUC, và các bản đồ kết quả trong GIS.

5. Đóng góp của luận văn : Trình bày các mô hình AI-LGBM, CNN-GIS và PSO-SCNN được tối ưu hóa với phân nhóm không gian và điều chỉnh siêu tham số; tích hợp GIS để lập bản đồ chất lượng nước ngầm ở Odisha và Đồng bằng sông Cửu Long.

6. Bố cục của luận văn : Cấu trúc luận văn bao gồm Giới thiệu, ba chương và Kết luận. Cụ thể,

Chương 1: Phân loại khả năng uống được của nước ngầm và kiến thức nền.

Chương 2 trình bày các phương pháp Học máy không gian tập hợp được đề xuất.

Chương 3 trình bày kết quả của AI-LGBM, PSO-SCNN nâng cao độ bền vững (ANOVA), CNN–bản đồ không gian rủi ro; kiến trúc hệ thống.

Chương 1

Phân loại khả năng uống được của nước ngầm và Kiến thức nền

1.1 Phân loại khả năng uống được của nước ngầm

Bối cảnh và Động lực. Nước ngầm duy trì sự sống cho hàng tỷ người nhưng đang đối mặt với các rủi ro từ kim loại nặng, nitrat và thuốc trừ sâu. Các phương pháp đánh giá truyền thống chậm và tốn kém, trong khi AI hứa hẹn mang lại phân loại kịp thời và có thể mở rộng, nhưng vẫn gặp khó khăn với độ chính xác, khả năng mở rộng và khả năng giải thích. Công trình này tập trung vào ba vấn đề: phân loại khả năng uống được nước ngầm đa lớp, tối ưu hóa siêu tham số vững chắc, và trực quan hóa không gian hỗ trợ ra quyết định.

Vấn đề 1: Phân loại khả năng uống được của nước ngầm

Mục tiêu. Phân loại mỗi mẫu thành *Xuất sắc*, *Tốt*, *Vừa phải*, *Kém*, hoặc *Không phù hợp để uống* sử dụng các đặc trưng vật lý hóa học và không gian (ví dụ, pH, TDS, nitrat, vĩ độ, kinh độ).

Công thức. Giả sử $X = \{x_i\}_{i=1}^{n}$, $x_i \in \mathbb{R}^m$, và nhãn $y_i \in \{1, \ldots, k\}$. Một mô hình $f(\cdot\,; W)$ sẽ đưa ra điểm số lớp; lớp dự đoán là

$$\hat{y}_i = \arg \max_{c \in \{1,\ldots,k\}} f_c(x_i; W).$$

Chúng ta huấn luyện bằng cách tối thiểu hóa rủi ro thực nghiệm

$$\min_{W} \; \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(y_i, f(x_i; W)) \,,$$

với $\mathcal{L}$ thường là hàm mất mát chéo đa lớp. Đánh giá sử dụng độ chính xác, độ

nhạy, độ đặc hiệu, F1, và AUC. Không giống như các phương pháp WQI với ngưỡng cố định, mô hình học các quan hệ phi tuyến và có thể mở rộng cho các bộ dữ liệu lớn và đa dạng.

Vấn đề 2: Tối ưu hóa siêu tham số cho mô hình GWQC

Mục tiêu. Lựa chọn các siêu tham số (ví dụ, độ sâu cây, tốc độ học, số lượng bộ phân loại, điều chỉnh đều) để tối đa hóa hiệu suất ngoài mẫu trong khi kiểm soát chi phí tính toán.

Phương pháp. Sử dụng tìm kiếm hộp đen qua AIO, Optuna và Tối ưu hóa đàn chim (PSO) trên một không gian tìm kiếm $\mathcal{W}$. Giả sử $g(W)$ là điểm số đã được kiểm tra chéo (ví dụ, macro-F1). Bộ tối ưu hóa giải bài toán

$$W^* = \arg\max_{W \in \mathcal{W}} g(W),$$

có thể với các ràng buộc về tài nguyên (ví dụ, ngân sách thời gian hoặc FLOPs). Phương pháp này cải thiện độ chính xác, sự ổn định và khả năng tổng quát trên các dữ liệu đa dạng từ Việt Nam và Ấn Độ, bao gồm các tình huống nhiễu và dữ liệu có độ chiều cao.

Vấn đề 3: Trực quan hóa không gian các nhãn phân loại

Mục tiêu. Lập bản đồ các dự đoán lên bản đồ địa lý để truyền đạt rủi ro và lập kế hoạch.

Công thức. Giả sử $G = \{(lat_i, lon_i)\}_{i=1}^n$ là tọa độ mẫu và $\hat{y} = \{\hat{y}_i\}_{i=1}^n$ là đầu ra của mô hình. Một quy trình GIS tạo ra bản đồ chủ đề

$$M = \mathrm{GIS}(G, \hat{y}),$$

có thể sử dụng nội suy hoặc tổng hợp diện tích. Để kết hợp phân loại và sự phù hợp không gian, chúng ta xem xét mục tiêu tổng hợp

$$L_{\text{total}} = L_{\text{classification}} + \lambda\, L_{\text{spatial}},$$

trong đó $L_{\text{classification}}$ là hàm mất mát chéo và L_{spatial} phạt sự gián đoạn không gian không hợp lý hoặc sự sai lệch với các thông tin không gian đã biết; $\lambda > 0$ điều chỉnh sự cân bằng này. Đầu ra là bản đồ dễ hiểu của các lớp khả năng uống

được, làm nổi bật các điểm nóng và khu vực ưu tiên để giám sát.

Đóng góp và Tác động. Quy trình thay thế các ngưỡng WQI cứng nhắc bằng một bộ phân loại linh hoạt, có thể dự đoán từ dữ liệu; sử dụng tìm kiếm siêu tham số hợp lý để đảm bảo việc triển khai đáng tin cậy; và cung cấp các sản phẩm không gian hỗ trợ các quyết định chính sách. Các thành phần này cùng nhau tạo điều kiện cho việc đánh giá chất lượng nước ngầm nhanh chóng, có thể mở rộng và nhận thức vùng.

1.2 Tổng quan Tài liệu

1.2.1 Phương pháp Cổ điển

Các phương pháp đánh giá chất lượng nước ngầm truyền thống tốn nhiều công sức và phụ thuộc vào việc lấy mẫu và phân tích thủ công. Chỉ số chất lượng nước (WQI) cung cấp phân loại đơn giản nhưng bị giới hạn bởi tính chủ quan và các ngưỡng do chuyên gia đưa ra.

Hạn chế của các Phương pháp Cổ điển

Các phương pháp chất lượng nước truyền thống chậm chạp, mang tính chủ quan và thiếu dữ liệu thời gian thực. Những khoảng trống chính bao gồm việc giải quyết tính phi tuyến và cải thiện việc tích hợp dữ liệu cho phân tích thời gian thực.

1.2.2 Phương pháp Học Máy (ML)

Các phương pháp học máy như SVM, RF, và LightGBM xử lý các bộ dữ liệu lớn và các mẫu phi tuyến, giúp cải thiện độ chính xác. Tuy nhiên, vấn đề overfitting, chất lượng dữ liệu và khả năng giải thích vẫn còn là những thách thức.

1.2.3 Phương pháp Học Sâu (DL)

Các mô hình học sâu (CNNs, RNNs) xuất sắc trong việc tự động trích xuất đặc trưng và xử lý các bộ dữ liệu lớn, nhưng yêu cầu năng lực tính toán

cao và bộ dữ liệu lớn, với khả năng giải thích hạn chế.

1.2.4 Mô hình Học Máy Không gian lai

Các mô hình lai kết hợp giữa học máy, học sâu và GIS nâng cao phân loại nước ngầm bằng cách sử dụng dữ liệu không gian để nắm bắt sự biến đổi vùng và cung cấp cái nhìn thời gian thực.

Phân nhóm Không gian và Tích hợp GIS

Phân nhóm không gian (ví dụ, K-means, DBSCAN) và học máy giúp cải thiện phân loại bằng cách nắm bắt các mẫu không gian, với GIS tích hợp dữ liệu không gian. Các mô hình học sâu như LightGBM và CNNs cải thiện độ chính xác nhưng yêu cầu tài nguyên tính toán đáng kể.

Mô hình Hỗn hợp RainNet và GA cho Tuning Siêu tham số

Mô hình hỗn hợp RainNet và Thuật toán Di truyền (GA) giảm MAE so với các mô hình như Unet và Segnet, giúp cải thiện độ chính xác trong dự báo lượng mưa.

1.3 Hạn chế và các khoảng trống nghiên cứu

Mặc dù có tiến bộ trong phân loại chất lượng nước ngầm, vẫn còn vấn đề như thiếu dữ liệu, overfitting và khả năng giải thích. Nghiên cứu tương lai nên tập trung vào cải thiện khả năng tổng quát, giảm chi phí tính toán và giám sát thời gian thực.

1.4 Kết luận

Chương này tổng quan các phương pháp phân loại chất lượng nước ngầm. Các phương pháp học máy và học sâu cải thiện độ chính xác, nhưng cần giải quyết vấn đề overfitting và khả năng giải thích.

Chương 2

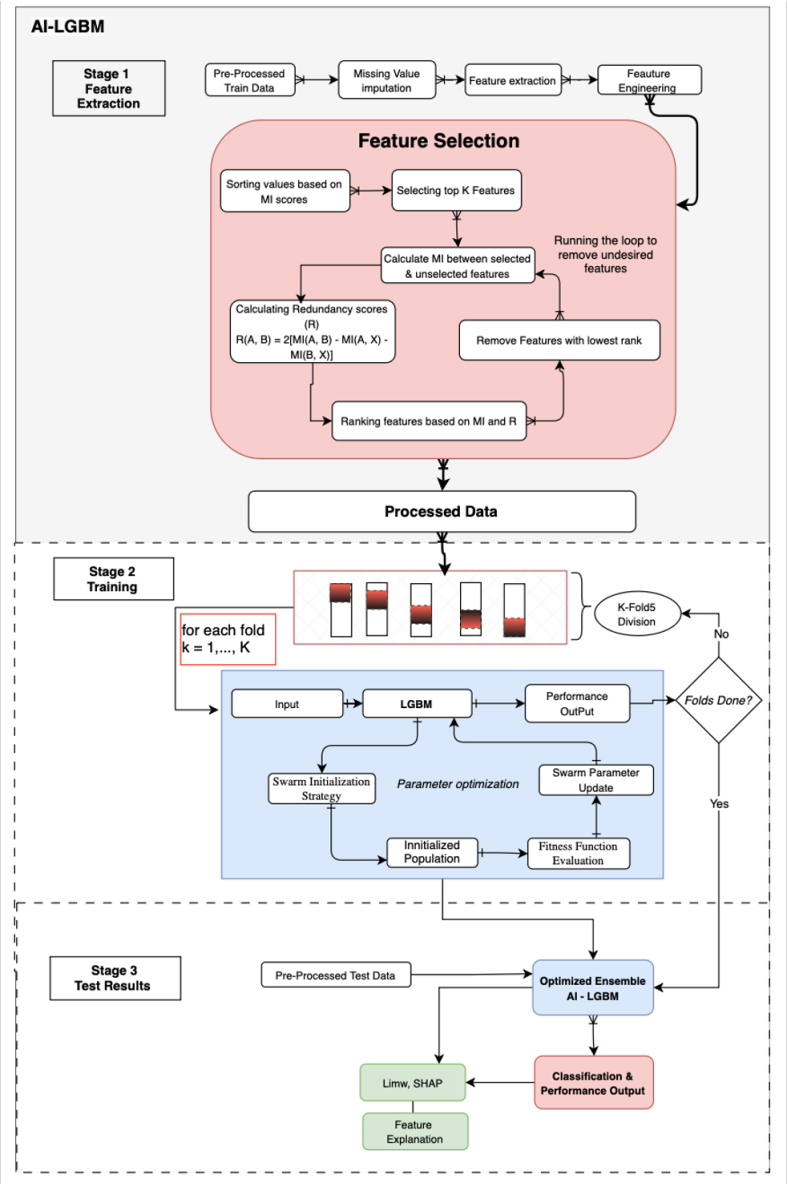
Phương Pháp Học Máy Không Gian Kết Hợp

Chương này giới thiệu phương pháp học máy không gian hợp bộ để phân loại chất lượng nước ngầm, tập trung vào hai mô hình: Máy Tăng Cường Gradient Nhẹ (AI-LGBM) được cải tiến với trí tuệ nhân tạo và Mạng Nơ-ron Tích Chập Không Gian tối ưu hóa Bầy Đông (PSO-SCNN).

2.1 AI-LGBM

2.1.1 Tổng Quan Về Khung AI-LGBM Đề Xuất

Máy Tăng Cường Gradient Nhẹ (AI-LGBM) được cải tiến với trí tuệ nhân tạo là một mô hình tiên tiến được thiết kế để kết hợp các lợi ích của tăng cường gradient với các kỹ thuật trí tuệ nhân tạo. Ý tưởng chính của AI-LGBM là nâng cao hiệu suất dự đoán của mô hình LightGBM truyền thống bằng cách kết hợp các kỹ thuật học máy như phân tích tầm quan trọng của đặc trưng và các thuật toán tối ưu hóa. Mô hình này đặc biệt hiệu quả trong việc xử lý các tập dữ liệu lớn và phức tạp với nhiều biến đầu vào, làm cho nó lý tưởng cho phân loại chất lượng nước ngầm, nơi dữ liệu có thể bao gồm nhiều tham số lý hóa học.



Hình 2.1: Sơ Đồ Quy Trình AI-LGBM Đề Xuất

Công Thức Toán Học Của AI-LGBM Với MIFS

Cài Đặt. Cho các mẫu $X = \{x_i\}_{i=1}^{n}$ với $x_i \in \mathbb{R}^m$ (các đặc trưng lý hóa học + không gian) và nhãn $y_i \in \{1, \ldots, k\}$, mô hình AI-LGBM $f(\cdot; W)$ sản xuất các điểm số lớp. Dự đoán và huấn luyện:

$$\hat{y}_i = \arg\max_c f_c(x_i; W), \qquad \min_W \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\big(y_i, f(x_i; W)\big).$$

Hiệu Suất / Siêu Tham Số. Chọn siêu tham số bằng điểm số cross-validated $g(W)$ (ví dụ: macro-F1, AUC):

$$W^* = \arg\max_W g(W).$$

Chúng tôi sử dụng tìm kiếm hỗn hợp với **Optuna** (đề xuất thay thế/TPE), **PSO** (tinh chỉnh bầy đàn), và **AIO** (đột biến thích nghi):

$$W^{(s)} \sim \pi_\phi(W \mid \mathcal{H}_{s-1}), \qquad v_{t+1} = \omega v_t + c_1 r_1 (pbest - W_t) + c_2 r_2 (gbest - W_t), \quad W_{t+1} = W_t + v_{t+1},$$

$$W_{t+1} \leftarrow W_{t+1} + \eta\, \mathcal{A}(W_{t+1};\, \mathcal{H}_t),$$

với dừng sớm và cross-validation K-fold để đảm bảo tổng quát hóa ổn định.

Chọn Đặc Trưng (MIFS). Xếp hạng các đặc trưng theo thông tin hỗ trợ với nhãn và kiểm soát độ dư thừa; chọn k đặc trưng S_k bằng

$$S_k = \arg\max_{S:\,|S|=k} J(S), \quad J(S) = \sum_{x_j \in S} I(x_j; Y) \;-\; \lambda \sum_{\substack{x_j, x_\ell \in S \\ j<\ell}} I(x_j; x_\ell),$$

nơi $I(\cdot\,;\cdot)$ là thông tin hỗ trợ. (Tương đương với $I(X;Y) = H(X) + H(Y) - H(X,Y)$.)

Mục Tiêu Tóm Tắt. MIFS giảm chiều trước khi huấn luyện; AI-LGBM sau đó tối ưu W để giảm thiểu mất mát và tối đa hóa $g(W)$ dưới cross-validation.

Cơ Sở Toán Học

$$\text{Phân loại: } \min_W \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\big(y_i, f(x_i; W)\big), \quad \hat{y}_i = \arg\max_c f_c(x_i; W). \tag{2.1}$$

$$\text{Siêu tham số: } W^* = \arg\max_W g(W) \text{ (ví dụ: macro-F1/AUC qua CV).} \tag{2.2}$$

$$\text{Chọn đặc trưng: } S_k = \arg\max_{S:\,|S|=k} J(S). \tag{2.3}$$

Giả Thuyết (AI-LGBM + MIFS)

$$H_0 : \mathbb{E}[g(\text{AI-LGBM+MIFS})] = \mathbb{E}[g(\text{Các Baseline})],$$

$$H_1 : \mathbb{E}[g(\text{AI-LGBM+MIFS})] > \mathbb{E}[g(\text{Các Baseline})].$$

2.1.2 Cài Đặt Thực Nghiệm và Chiến Lược Học Cho AI-LGBM

Chia dữ liệu nước ngầm 70/15/15; tiền xử lý: thay thế giá trị thiếu, chuẩn hóa Z-score, loại bỏ ngoại lệ IQR. Học có giám sát: chọn MIFS, cân bằng SMOTE, LightGBM với tối ưu hóa Optuna/AIO (5-fold CV, max F1 có trọng số). Các chỉ số: độ chính xác, độ chính xác, độ thu hồi, F1, AUC; khả năng giải thích SHAP.

Công Thức Toán Học

Giả sử $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, $x_i \in \mathbb{R}^m$, $y_i \in \{1, \ldots, K\}$.

Chọn Đặc Trưng (MIFS). Chọn các đặc trưng top-K: $S^\star = \arg\max_{S \subset \{1,\ldots,m\},\, |S|=K} \sum_{j \in S} \mathcal{I}(X_j; Y)$, sau đó $X \leftarrow X_{S^\star}$.

Cân Bằng SMOTE. Đối với x_i thiểu số và hàng xóm kNN $x_i^{(nn)}$: $\tilde{x} = x_i + \lambda(x_i^{(nn)} - x_i)$, $\lambda \sim \mathcal{U}(0,1)$, $\tilde{y} = y_i$, tạo ra $\mathcal{D}_{\text{train}}^{\text{smote}}$.

Mô Hình Tăng Cường Bổ Sung. LightGBM phù hợp $F_m(x) = F_{m-1}(x) + \eta \sum_{j=1}^{J_m} \gamma_{jm} \mathbb{I}(x \in R_{jm})$, với $\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$.
Mất mát đa lớp: $\ell_i = -\sum_{k=1}^{K} y_{ik} \log p_{ik}$, $\mathcal{L} = \sum_i \omega_i \ell_i$. Cập nhật lá: $w_j^\star = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$, Gain như trong gốc.

Tối Ưu Siêu Tham Số. Tối ưu $\theta^* = \arg\max_\theta \frac{1}{K} \sum_{k=1}^{K} \text{F1-score}_k(\theta)$.

2.1.3 Tối Ưu Mô Hình và Hiệu Suất

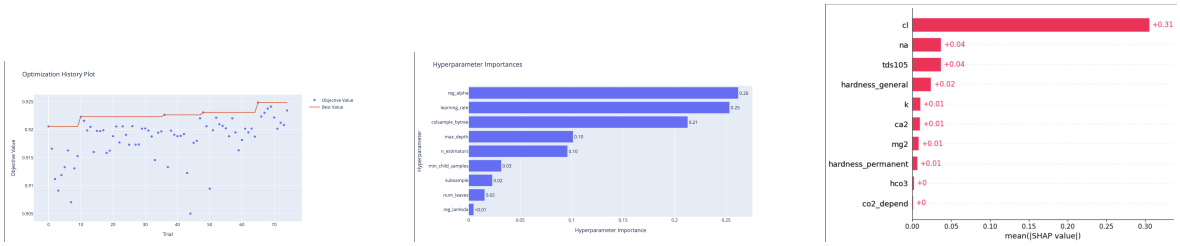
AI–LGBM đã được điều chỉnh với AIO/Optuna dưới 5-fold CV để (*learning_ rate*=0.05,

num_ leaves=32, *max_ depth*=8, *n_ estimators*=150, *subsample*=0.8,

colsample_ bytree=0.7). So với mặc định, độ chính xác tăng từ 0.812 lên 0.865 và F1 có trọng số tăng từ 0.801 lên 0.864 (7.9%), với các cải tiến tương tự về độ chính xác và độ thu hồi.

2.1.4 Tầm Quan Trọng Của Đặc Trưng và Trực Quan Hóa

Lịch sử/quan trọng Optuna trong Hình 2.2a–2.2b. SHAP nổi bật tds105, na, cl (Hình 2.2c).



(a) Lịch sử Optuna (b) Quan Trọng Siêu Tham Số (c) Tóm Tắt SHAP

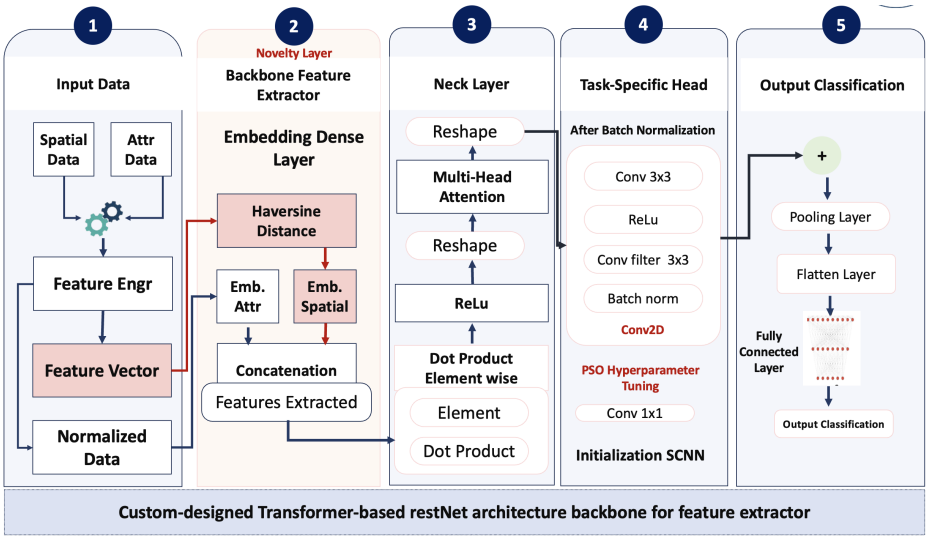
Ưu Điểm. Độ chính xác cao, xử lý dữ liệu không gian với nhiều chiều, mô hình hóa quan hệ tuyến tính và phi tuyến.

Nhược Điểm. Tốn tài nguyên tính toán, yêu cầu điều chỉnh tham số, hạn chế khả năng giải thích (giảm thiểu với SHAP).

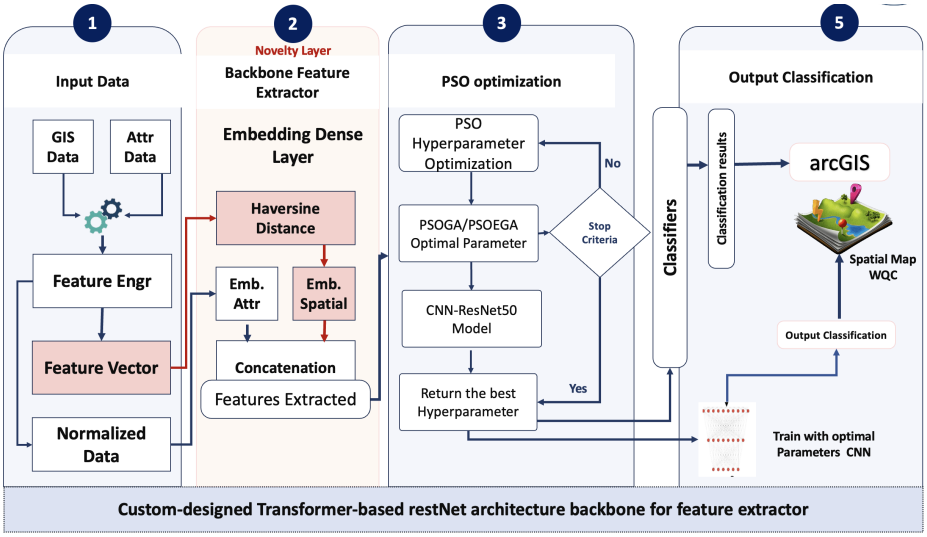
2.2 PSO-SCNN

2.2.1 Tổng Quan Về Khung PSO-SCNN

Đề xuất PSO-SCNN, mô hình học sâu kết hợp AI-LGBM, nắm bắt phụ thuộc không gian qua nhúng không gian, mã hóa Haversine, chú ý đa đầu và CNN.



Hình 2.3: Kiến Trúc Mô Hình Không Gian PSO-SCNN



Hình 2.4: Mở Rộng Cho Minh Họa Bản Đồ Không Gian

2.2.2 Công Thức Toán Học Của PSO-SCNN và CNN-GIS

Công Thức PSO-SCNN

Mô hình PSO-SCNN tối ưu hóa Mạng Nơ-ron Tích Chập Không Gian (SCNN) bằng cách sử dụng Tối Ưu Hóa Bầy Đông (PSO) để phân loại chất lượng nước ngầm, kết hợp dữ liệu không gian để cải thiện mô hình phụ thuộc không gian. Giả sử $\mathcal{D} = \{(x_i, y_i, (lat_i, lon_i))\}_{i=1}^{n}$ là tập dữ liệu, trong đó $x_i \in \mathbb{R}^m$ là các đặc trưng lý hóa học, $y_i \in \{0, 1\}$ là nhãn tính uống được nhị phân, và (lat_i, lon_i) là tọa độ.

Tiền Xử Lý và Mã Hóa Không Gian. Các đặc trưng được chuẩn hóa: $x_i^* = \frac{x_i - \mu_x}{\sigma_x}$. Các đặc trưng không gian được mã hóa qua khoảng cách Haversine từ tâm $(\bar{lat}, \bar{lon})$:

$$d_i = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{lat_i - \bar{lat}}{2}\right) + \cos(\bar{lat})\cos(lat_i)\sin^2\left(\frac{lon_i - \bar{lon}}{2}\right)}\right), \quad (2.4)$$

tạo ra các đầu vào mở rộng $\tilde{x}_i = [x_i^*; d_i]$. SMOTE cân bằng các lớp bằng cách tạo ra các $\tilde{x}_j$ tổng hợp cho lớp thiểu số.

Tối Ưu PSO. PSO tìm kiếm các siêu tham số $\theta = \{\text{filters}, \text{kernel size}, \text{learning rate}\}$ trong bầy động $\{p_k\}_{k=1}^{P}$. Độ thích hợp là AUC âm: $\text{Fit}(p_k) = -\text{AUC}(\text{SCNN}_\theta)$.

Cập nhật:

$$v_k^{t+1} = wv_k^t + c_1 r_1(pbest_k - p_k^t) + c_2 r_2(gbest - p_k^t), \tag{2.5}$$

$$p_k^{t+1} = p_k^t + v_k^{t+1}, \tag{2.6}$$

với w là động lực, c_1, c_2 là các hệ số, $r_1, r_2 \sim \mathcal{U}(0,1)$, hội tụ đến θ^* tối ưu.

Kiến Trúc SCNN. SCNN xử lý $\tilde{x}_i$ qua các lớp tích chập: $h_l = \sigma(W_l * h_{l-1} + b_l)$, lớp gộp, và các lớp dày đặc, xuất ra $\hat{y}_i = \sigma(W_f h_L + b_f)$. Được huấn luyện với hàm mất mát nhị phân: $\mathcal{L} = -\sum_i [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$.

Công Thức CNN-GIS

CNN-GIS mở rộng PSO-SCNN cho việc minh họa không gian, ánh xạ các dự đoán $\hat{y}_i$ tới các tọa độ địa lý (lat_i, lon_i) qua tích hợp GIS. Mục tiêu là tạo ra bản đồ chủ đề M làm nổi bật các lớp chất lượng và các khu vực nóng.

Bản Đồ Dự Đoán Không Gian. Các dự đoán được nội suy trên lưới $G = \{(lat_g, lon_g)\}_{g=1}^{G}$ bằng phương pháp trọng số khoảng cách nghịch đảo (IDW):

$$\hat{y}(lat_g, lon_g) = \frac{\sum_i w_i \hat{y}_i}{\sum_i w_i}, \quad w_i = \frac{1}{d((lat_g, lon_g), (lat_i, lon_i))^p}, \tag{2.7}$$

nơi $d(\cdot)$ là khoảng cách Haversine và $p > 0$ điều khiển sự suy giảm. Hàm mất mát tổng hợp kết hợp phân loại với điều chỉnh không gian:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{class}} + \lambda \sum_{i,j} \|\hat{y}_i - \hat{y}_j\| \cdot d((lat_i, lon_i), (lat_j, lon_j))^{-1}, \tag{2.8}$$

Đảm bảo sự mượt mà không gian. Kết quả xuất sang GeoTIFF cho minh họa ArcGIS về khu vực nóng tính uống được.

2.2.3 Lý Do Chọn Mô Hình Hybrid và Tiêu Chí Đánh Giá

Mô hình hybrid (AI-LGBM, PSO-SCNN) chọn vì độ chính xác, khả năng giải thích, và mở rộng trong phân loại nước ngầm không gian; đánh giá tập trung vào ổn định, hiệu quả, và hữu dụng.

2.2.4 Chiến Lược Học Của PSO-SCNN

Chiến lược PSO-SCNN kết hợp khởi tạo, tối ưu hóa, trích xuất đặc trưng không gian và huấn luyện/kiểm tra. PSO tìm kiếm siêu tham số, tối ưu hóa lỗi. SCNN trích xuất đặc trưng không gian từ dữ liệu nước ngầm. Huấn luyện với cross-validation K-fold.

Cài Đặt Có Giám Sát. Mục tiêu: $\theta^* = \arg\max_\theta \frac{1}{K} \sum_{k=1}^{K} \mathrm{F1_w} \left(f_\theta^{(-k)}, \mathcal{D}^{(k)} \right)$.

Vòng Lặp PSO. Cập nhật các hạt qua các phương trình vận tốc/vị trí, đánh giá trên AUC kiểm tra để chọn θ^*.

Huấn Luyện. Sử dụng Adam với dừng sớm trên F1 kiểm tra để đảm bảo sự tổng quát và ánh xạ không gian.

2.2.5 So Sánh Các Thuật Toán Học

Lựa chọn bộ tối ưu ảnh hưởng đến hội tụ. Bảng 2.1 so sánh Adam, AdamW và AdaGrad. Adam được chọn vì hiệu quả, tỷ lệ thích nghi và ít yêu cầu điều chỉnh, lý tưởng cho SCNN trên dữ liệu nước ngầm có chiều cao. AdamW phù hợp với các mô hình quy mô lớn; AdaGrad cho dữ liệu thưa nhưng có thể hội tụ chậm.

Bảng 2.1: So Sánh Các Bộ Tối Ưu

Optimizer	**Speed**	**Adaptivity**	**Generalization**	**Tune Need**	**Use Case**
Adam	Fast	Yes	Very Good	Low	Deep networks
AdamW	Fast	Yes	Excellent	Low	Large-scale models
AdaGrad	Medium	Yes	Good early	Medium	Sparse data

Bảng 2.2: Các Siêu Tham Số Quan Trọng Của PSO-SCNN

Siêu Tham Số	**Mô Tả**	**Giá Trị**
Kích Thước Hạt	Kích thước bầy động	10-50
Trọng Số Quán Tính	Tác động của vận tốc trước đó	0.5-0.9
C1/C2	Ảnh hưởng cá nhân/toàn cầu	1.5-2.0
Số Vòng Lặp Tối Ưu	Số vòng lặp PSO	50-200
Kích Thước Kernel	Kích thước kernel tích chập	3×3, 5×5
Bước Nhảy	Bước nhảy tích chập	1-2

2.2.6 Ảnh Hưởng Của Các Siêu Tham Số PSO

Các tham số PSO cân bằng giữa khám phá/khai thác. Đã sử dụng: $n_{\text{particles}} = 3$, $w = 0.9$, $c_1 = 0.5$, $c_2 = 0.3$. Bảng 2.3 cho thấy ảnh hưởng đến hiệu suất.

Bảng 2.3: Ảnh Hưởng Của Các Tham Số PSO Đến PSO-SCNN

Cấu Hình	w	**AUC**	**F1**	**Hội Tụ**
High w (Khám Phá)	0.9	0.965	0.945	Chậm
Cân Bằng (Nghiên Cứu)	0.9	0.988	0.965	Trung Bình
Thấp w, Cao c_2	0.4	0.972	0.950	Nhanh, Mạo Hiểm

Trọng số w cao hỗ trợ khám phá nhưng làm chậm hội tụ; các chiến lược thích nghi tăng cường độ bền.

2.2.7 Ưu Điểm và Nhược Điểm

Ưu Điểm: Xử lý dữ liệu không gian tốt, mô hình hóa phụ thuộc phức tạp. *Nhược Điểm:* Tốn tài nguyên, khó giải thích.

2.3 Kết Luận Chương

Chương này đề xuất sự kết hợp AI-LGBM và PSO-SCNN để sử dụng thông tin không gian trong bài toán đánh giá chất lượng nước ngầm và tối ưu hóa tham số.

Đóng Góp Chính

- Kết hợp ensemble và học sâu không gian.
- Tối ưu hóa kép cho tổng quát.
- SHAP và minh họa không gian.

Hạn Chế Chi phí tính toán cao, phức tạp điều chỉnh tham số, khó giải thích.

Triển Vọng Chương tiếp theo đánh giá kết quả và các minh họa không gian.

Chương 3

Kết Quả và Đánh Giá

3.1 Đánh Giá Hiệu Suất và So Sánh

Các kết quả ML truyền thống trong 3.1 được công bố trong *Earth Science Informatics*, 16(2), 1701–1725. Springer.

[DOI: https://doi.org/10.1007/s12145-023-00977-x].

Bảng 3.1: Các Chỉ Số Hiệu Suất cho Các Mô Hình Khác Nhau trong Dữ Liệu Odisha

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**
Hồi Quy Logistic	0.7051	0.72	0.6275	0.6025
SVM Đa Thức	0.9012	0.9175	0.9025	0.8925
Cây Quyết Định	0.8989	0.885	0.8900	0.8850
AdaBoost	0.5445	0.465	0.4950	0.4650
CNN	0.9766	0.9877	0.9877	0.9877
AI-LGBM	0.94	0.95	0.92	0.93

Bảng 3.2: Các Chỉ Số Hiệu Suất cho Các Mô Hình Khác Nhau trong Dữ Liệu Việt Nam

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**
Hồi Quy Logistic	0.9672	0.5333	0.5517	0.5714
SVM Đa Thức	0.9766	0.9950	0.9926	0.9950
Cây Quyết Định	0.9696	0.9877	0.9889	0.9877
AdaBoost	0.9696	0.9901	0.9877	0.9901
CNN	0.9766	0.9877	0.9913	0.9877
AI-LGBM	0.94	0.95	0.92	0.93

3.1.1 Kết Quả ML (Sau Tối Ưu Hóa): AI-LGBM

Bảng 3.3: *So Sánh Giá Trị Trung Bình của Các Chỉ Số Hiệu Suất của Tất Cả Các Mô Hình trong Dữ Liệu Việt Nam*

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**
K-NN	0.899533	0.909028	0.899533	902478
SVM	0.897196	0.922039	0.897196	0.902437
Cây Quyết Định	0.989655	0.987780	0.988920	0.987710
AdaBoost	0.9696	0.9853	0.9877	0.9901
XGBoost	0.9813	0.9902	0.9938	0.9975

So Sánh Mô Hình AI-LGBM Với Các Mô Hình Cơ Bản và Mô Hình Tiên Tiến

Bảng 3.4: So Sánh AI-LGBM Với Các Mô Hình Cơ Bản

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**
XGBoost (cơ bản)	0.9367	0.9325	0.9275	0.9324
SVM Đa Thức (cơ bản)	0.9012	0.9175	0.9025	0.8925
Cây Quyết Định (cơ bản)	0.97992	0.9821	0.9799	0.9785
AI-LGBM (đề xuất)	**0.9953**	**0.9954**	**0.9953**	**0.9953**

Bảng 3.5: So Sánh Hiệu Suất Mô Hình Dữ Liệu Việt Nam Với Log Loss

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**	**Log Loss**
MLP Đơn Giản	0.985981	0.986333	0.985981	0.986113	0.071997
MLP 2	0.983645	0.983645	0.983645	0.983645	0.115310
AI-LGBM	0.995327	0.995492	0.995327	0.995363	0.019135

Kết Luận AI-LGBM

Được đánh giá lại trên dữ liệu Odisha và Việt Nam, AI–LGBM liên tục vượt trội so với KNN, SVM, Cây Quyết Định, và XGBoost về độ chính xác, độ chính xác, độ thu hồi và F1. So với các mô hình sâu (MLP/CNN/Transformer) trên bộ dữ liệu Kaggle và bộ dữ liệu Việt Nam, AI–LGBM dẫn đầu về F1 và độ thu hồi, đạt độ chính xác 99.53% và log loss 0.0191 trên dữ liệu Việt Nam.

3.2 Xác Thực PSO-SCNN

Xác thực PSO-SCNN sử dụng độ chính xác, độ chính xác, độ thu hồi và F1.

So sánh bộ tối ưu hóa. Grid Search đạt độ chính xác 1.0000 trong 4.56 s; PSO đạt 0.9948 trong 3.70 s; GA đạt 0.9948 nhưng mất 11.54 s—PSO mang lại sự cân bằng tốt nhất giữa tốc độ và độ chính xác, Grid Search đạt độ chính xác tối đa, GA chậm nhất.

3.2.1 Kết Quả Hiệu Suất PSO-SCNN

Kết quả trong Sec. 3.2.1 được chấp nhận bởi *Proc. ICIIT 2025* (Hà Nội; đang in); phương pháp kết hợp được gửi đến *Journal of the Indian Society of Remote Sensing* (SCIE, IF 2.2).

Bảng 3.6: *Hiệu Suất Mô Hình Dữ Liệu Việt Nam (Bộ Kiểm Tra)*

Mô Hình	**Độ Chính Xác**	**Độ Thu Hồi**	**Độ Chính Xác**	**F1-Score**	**AUC**
Support Vector Machine	0.764	0.920	0.750	0.835	0.960
Cây Quyết Định	0.980	1.000	1.000	0.990	0.980
XGBoost	0.950	0.950	0.890	0.950	0.990
LightGBM	0.950	0.960	0.885	0.950	0.980
SCNN	0.929	0.950	0.955	0.970	0.970
PSO-SCNN	**0.975**	**1.000**	**0.988**	**0.995**	**0.990**

Bảng 3.7 so sánh các mô hình đề xuất (AI-LGBM, PSO-SCNN và CNN-GIS) với các mô hình học máy truyền thống.

Bảng 3.7: So Sánh Các Mô Hình Đề Xuất Với Các Mô Hình Cơ Bản

Mô Hình	**Độ Chính Xác Trung Bình**	**Độ Chính Xác Trung Bình**	**F1-Score Trung Bình**	**Độ Thu Hồi Trung Bình**
XGBoost (cơ bản)	0.9267	0.9225	0.9175	0.9200
SVM Đa Thức (cơ bản)	0.9012	0.9175	0.9025	0.8925
Cây Quyết Định (cơ bản)	0.8989	0.8975	0.8900	0.8850
AI-LGBM (đề xuất)	**0.9400**	**0.9500**	**0.9300**	**0.9400**
PSO-SCNN (đề xuất)	**0.9880**	**0.9750**	**0.9950**	**1.0000**
CNN-GIS Mapping (đề xuất)	**0.9700**	**0.9650**	**0.9750**	**0.9800**

3.2.2 Kết Quả PSO-SCNN Sau Tối Ưu Hóa

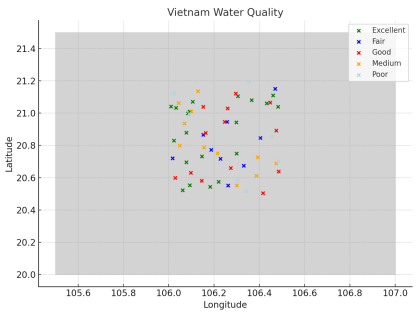
Sau tối ưu hóa, PSO-SCNN đã được đánh giá lại trên cả hai bộ dữ liệu, đạt được những cải thiện đáng kể về độ chính xác, độ thu hồi, F1 và AUC.

Bảng 3.8: *Hiệu Suất Mô Hình Tiến Bộ Dữ Liệu Việt Nam (Bộ Kiểm Tra)*

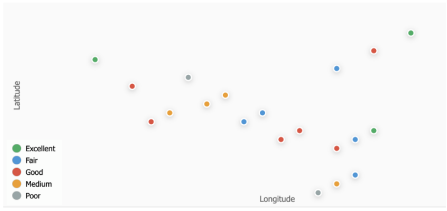
Mô Hình	**Độ Chính Xác**	**Độ Thu Hồi**	**F1-Score**	**AUC**
Autoencoder+Clf	0.923	0.939	0.931	0.978
CNN-LSTM	0.962	0.994	0.978	0.997
LSTM	0.951	0.978	0.964	0.993
Transformer	0.978	0.978	0.978	0.996
MLP2	0.983	0.961	0.972	0.992
MLP	0.972	0.972	0.972	0.994
PSO-SCNN	**0.994**	0.955	0.974	**0.993**

Bảng 3.9: Kết Quả Cross-Validation (Trung Bình ± Độ Lệch Tiêu Chuẩn) của Các Mô Hình Đề Xuất

Mô Hình	**Độ Chính Xác**	**F1-Score**	**AUC**	**Độ Thu Hồi**
AI-LGBM	0.932 ± 0.011	0.914 ± 0.009	0.945 ± 0.010	0.911 ± 0.012
PSO-SCNN	0.918 ± 0.013	0.902 ± 0.008	0.934 ± 0.009	0.889 ± 0.014
CNN-GIS	0.902 ± 0.015	0.880 ± 0.011	0.921 ± 0.012	0.867 ± 0.013



(a) Việt Nam - Khu Vực Mekong



(b) Biểu Đồ Phân Tán Chất Lượng Nước Tại Các Điểm Giếng ở Odisha

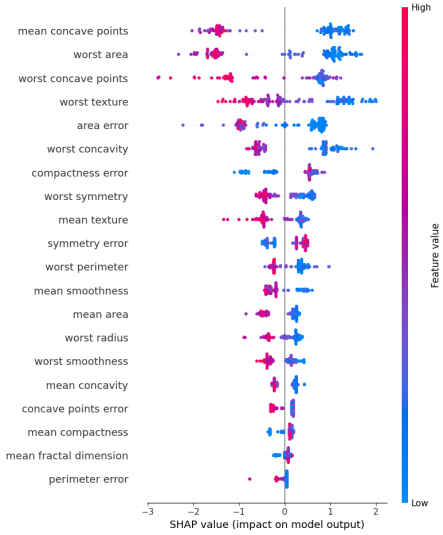
Hình 3.1: Minh Họa Chất Lượng Nước Tại Việt Nam và Odisha

Hiệu Suất Huấn Luyện và Kiểm Tra

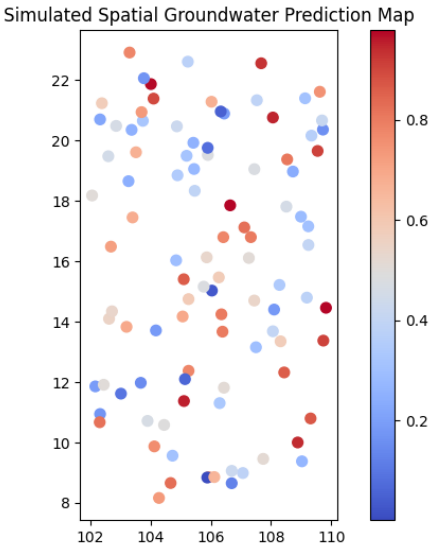
Các hình dưới đây cho thấy tổn thất huấn luyện/kiểm tra, độ chính xác và so sánh với các mô hình cơ bản, chỉ ra quá trình huấn luyện hiệu quả và khả năng tổng quát tốt.

Bảng 3.11: Tóm Tắt Cross-validation PSO–SCNN (trung bình ± SD).

Độ Chính Xác	F1	AUC	Độ Thu Hồi
0.918 ± 0.013	0.902 ± 0.008	0.934 ± 0.009	0.889 ± 0.014

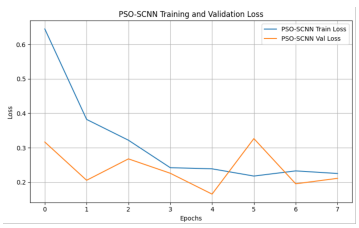


Hình 3.2: Biểu Đồ Tóm Tắt SHAP cho Mô Hình AI-LGBM

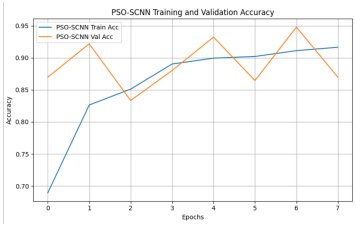


Hình 3.3: Lớp phủ các vùng không an toàn dự đoán với các khu vực ô nhiễm thực tế

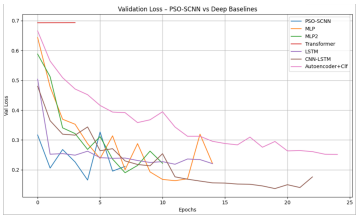
Hình 3.4: Tầm Quan Trọng Các Đặc Trưng SHAP và Xem Minh Họa Ô Nhiễm Không Gian



(a) Tổn Thất Huấn Luyện và Kiểm Tra PSO-SCNN



(b) Độ Chính Xác Huấn Luyện và Kiểm Tra PSO-SCNN



(c) So Sánh Tổn Thất Kiểm Tra - PSO-SCNN so với Các Mô Hình Deep Baseline

Hình 3.5: Đánh Giá Hiệu Suất: Tổn Thất Huấn Luyện/ Kiểm Tra, Độ Chính Xác và So Sánh

Bảng 3.10: Kết Quả PSO-SCNN Sau Tối Ưu Hóa trên Các Bộ Kiểm Tra Được Giữ Lại.

Khu Vực	Độ Chính Xác	Độ Thu Hồi	Độ Chính Xác	F1	AUC
Việt Nam (Mekong)	0.975	1.000	0.988	0.995	0.990
Ấn Độ (Odisha)	0.960	1.000	0.988	0.970	0.990

Cross-validation (tóm tắt cho PSO–SCNN). Cross-validation lặp lại năm lần đạt được 0.918 ± 0.013 Độ Chính Xác, 0.902 ± 0.008 F1, 0.934 ± 0.009 AUC, và 0.889 ± 0.014 Độ Thu Hồi—đảm bảo tính tổng quát mạnh mẽ trong khi vẫn duy trì hồ sơ độ thu hồi định hướng an toàn của mô hình.

Bảng 3.12 trình bày kết quả định lượng của nghiên cứu loại bỏ, tóm tắt độ chính xác, độ thu hồi, F1, AUC và thời gian huấn luyện cho mỗi biến thể mô hình.

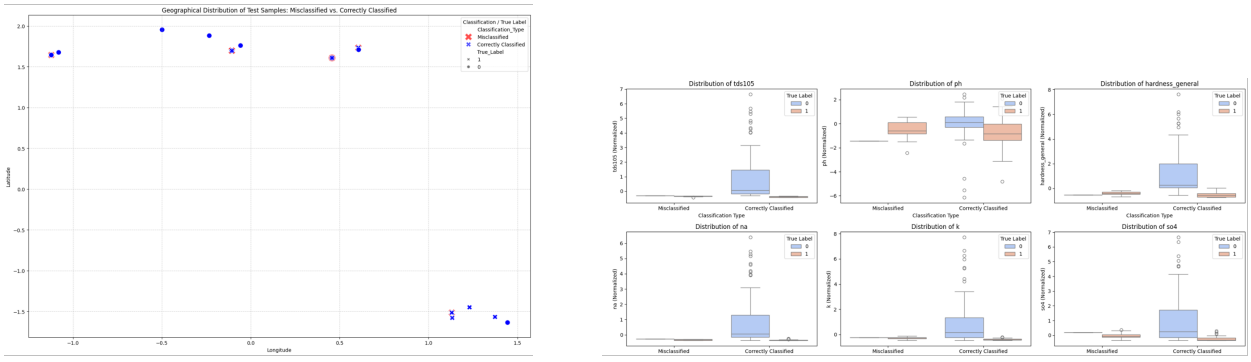
Bảng 3.12: Nghiên Cứu Loại Bỏ: Ảnh Hưởng Định Lượng của Việc Loại Bỏ Các Thành Phần

Mô Hình	**Độ Chính Xác**	**Độ Thu Hồi**	**F1**	**AUC**	**Epochs**	**Thời Gian Huấn Luyện (s)**
PSO-SCNN (hoàn chỉnh)	0.977528	0.988636	0.983051	0.998470	13	9.579775
SCNN không có PSO	0.965116	0.943182	0.954023	0.988418	13	9.588812
PSO-SCNN không có không gian	0.977011	0.965909	0.971429	0.997050	14	9.746294
SCNN nông	0.988506	0.977273	0.982857	0.998142	13	6.442084

Bảng 3.13: So Sánh Thời Gian Huấn Luyện và Tiêu Thụ Bộ Nhớ Cho Các Mô Hình AI-LGBM và PSO-SCNN

Đặc Tả	**AI-LGBM**		**PSO-SCNN**	
	Thời Gian Huấn Luyện	**Tiêu Thụ Bộ Nhớ**	**Thời Gian Huấn Luyện**	**Tiêu Thụ Bộ Nhớ**
Thời Gian Để Hội Tụ (giây)	2.750229	0.000000	3.2720	16.5 GB
Tiêu Thụ Bộ Nhớ (GB)	0.000000	0.000000	16.5 GB	16.5 GB
Đặc Tả Phần Cứng	Linux 6.6.105+	12.67 GB RAM, 2 cores	Linux 6.6.105+	32.65 GB RAM, 2 cores

Các phạm vi đặc trưng mà các mô hình chưa hiệu quả



(a) Các Vùng Nóng Nhầm Lẫn (PSO-SCNN)

(b) Phân phối đặc trưng cho các mẫu phân lớp đúng và sai

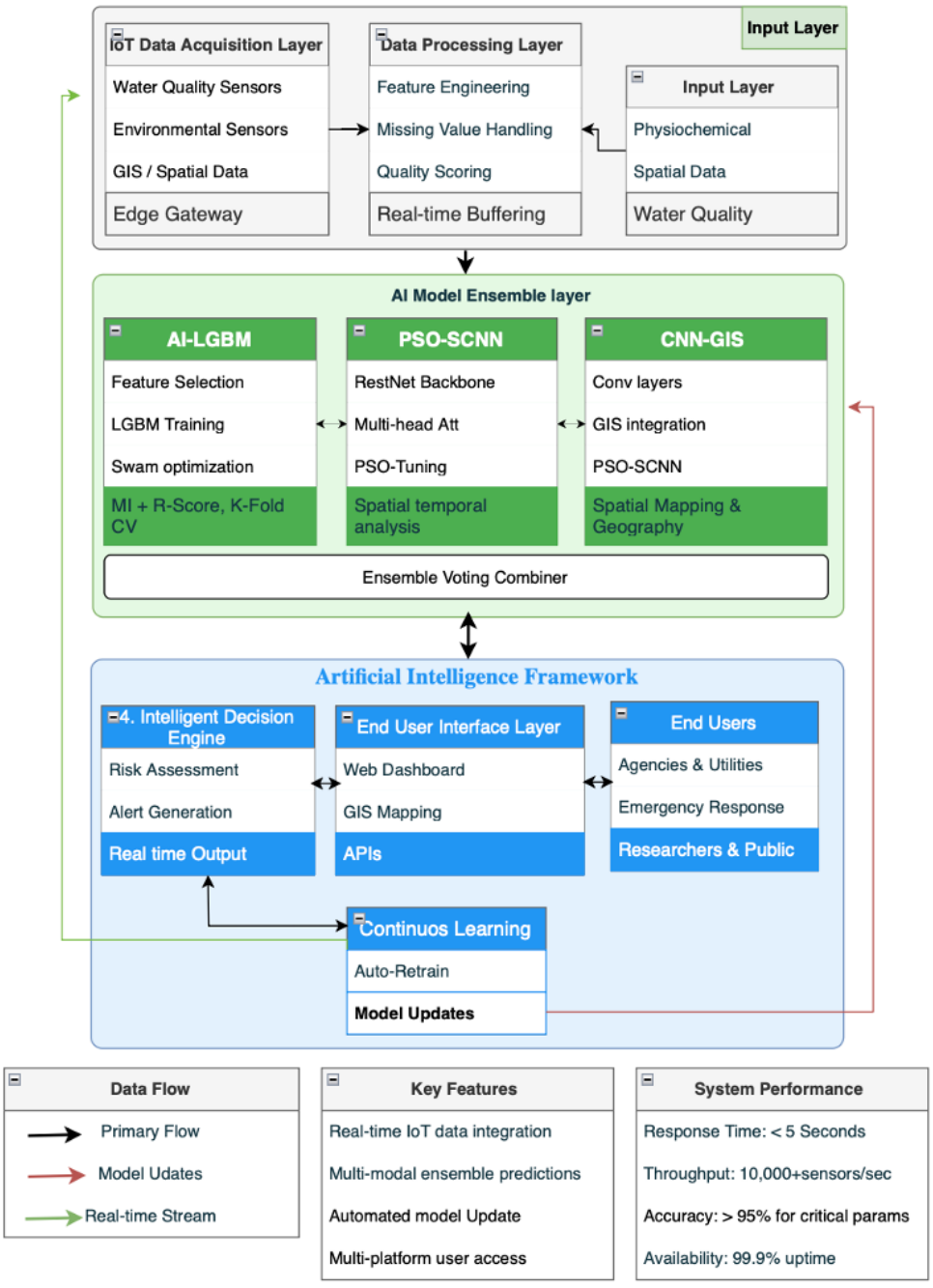
Hình 3.6: Phạm vi đặc trưng, Ma trận nhầm lẫn, Các khu vực phân lớp sai và phân phối đặc trưng

Các kết quả này được công bố trong bài báo cáo hội thảo về tối ưu hóa CNN–GIS trong *Proc. ICIIT 2025* và PSO–SCNN trong *Journal of the Indian Society of Remote Sensing.*

Kết Luận Chương

AI–LGBM: VN ≥98%, Odisha 92–93%, Prec >0.92, Rec >0.90, F1 >0.91; AIO/Optuna cải thiện F1 từ 15–20%.

PSO–SCNN: F1 vượt trội trên các nhiệm vụ không gian; PSO cải thiện hội tụ từ 25–30%, giảm hiện tượng quá khớp.



Hình 3.7: Kiến Trúc Hệ Thống Đề Xuất cho Khung Trí Tuệ Nhân Tạo

Kết Luận và hướng phát triển

Một số đóng góp chính

Kết hợp không gian hybrid (AI–LGBM, PSO–SCNN, CNN–GIS); tích hợp đặc trưng địa lý rõ ràng; tinh chỉnh siêu tham số dựa trên PSO; XAI (SHAP/LIME) cho các quyết định minh bạch.

Ý Nghĩa Khoa Học và Lý Thuyết

Tiến bộ trong học máy không gian cho hydroinformatics; kết hợp PSO với học sâu; nhúng XAI trong giám sát; chứng minh khả năng mở rộng liên vùng.

Hạn Chế

Đại diện dữ liệu hạn chế khả năng tổng quát; PSO–SCNN tốn tài nguyên tính toán; tích hợp IoT trong thời gian thực chưa thực hiện.

Hướng Nghiên Cứu Tương Lai

Thêm trích xuất đặc trưng học sâu cho dữ liệu không cấu trúc; mở rộng cho dữ liệu theo chiều dài thời gian, đa vùng; tích hợp IoT/viễn thám cho thời gian thực; bao gồm các yếu tố xã hội-kinh tế/khí hậu; phát hành nền tảng mã nguồn mở.

PHỤ LỤC A: MÃ VÀ DỮ LIỆU CÓ SẴN

A1 - TÁI SẢN XUẤT

Phần này cung cấp mã nguồn, bộ dữ liệu, phần mềm phụ thuộc và giá trị hạt giống ngẫu nhiên.

Sẵn Có Mã Nguồn

Mã nguồn: https://github.com/MichaelOmar24/PSO-SCNN-model, bao gồm tất cả các tập lệnh và tài nguyên.

Truy Cập Dữ Liệu

Dữ liệu có sẵn khi yêu cầu. Liên hệ: omar2@fe.edu.vn.

Phiên Bản Phần Mềm

Phụ thuộc: Python 3.8, TensorFlow 2.4.1, Keras 2.4.3, scikit-learn 0.24.1, matplotlib 3.3.4, NumPy 1.20.2, pandas 1.2.4.

Giá Trị Hạt Giống Ngẫu Nhiên

Hạt giống: Global Seed = 42, TensorFlow Seed = 42, NumPy Seed = 42.

DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN LIÊN QUAN ĐẾN LUẬN ÁN

[CT1] Niranjan Panigrahi, Gopal Krishna Patro, Raghvendra Kumar, Michael Omar, Tran Thi Ngan, Nguyen Long Giang, Bui Thi Thu, and Nguyen Truong Thang (2023). Groundwater quality analysis and drinkability prediction using artificial intelligence. *Earth Science Informatics*, 16(2), 1701–1725. Cham: Springer.

[DOI: 10.1007/s12145-023-00977-x]

[CT2] Tran Thi Ngan, Ha Gia Son, Michael Omar, Nguyen Truong Thang, Nguyen Long Giang, Tran Manh Tuan, and Nguyen Anh Tho (2023). A hybrid of RainNet and genetic algorithm in nowcasting prediction. *Earth Science Informatics*, 16(4), 3885–3894. (ISSN: 1865-0481, IF: 2.7 (2023)). Cham: Springer.

[DOI: 10.1007/s12145-023-01120-6]

[CT3] Michael Omar, Raghvendra Kumar, Tran Thi Ngan, Nguyen Long Giang, and Phung The Huan (2023). A comprehensive study on water quality prediction using machine learning and deep learning. In *Proceedings of the 25th National Conference on Some Selected Issues of Information and Communication Technology (VNICT 2022)*, Hanoi, Vietnam, pp. 1–7.

[CT4] Michael Omar, Nguyen Long Giang, Tran Thi Ngan, Nguyen Hong Tan, and Nguyen Thu Van (2024). AI-LGBM for Groundwater Quality Prediction in Vietnam and India. In *Proceedings of the 10th EAI International Conference on Smart Objects and Technologies for Social Good (GOODTECHS 2024)*, LNICST vol. 648, pp. 1–14, Cham: Springer, 2025. [DOI: 10.1007/978-3-032-01472-6_3]

[CT5] Nguyen Hai Minh, Michael Omar, Tran Thi Ngan, Nguyen Long Giang, and Hoang Thi Minh Chau (2024). Groundwater Quality in Vietnam Using Artificial Intelligence Models. *In proceedings (ICTA 2024), 3rd International Conference on Advances in Information and Communication Technology*. pp. 239-251, vol. 1205. Springer, Cham. [DOI: 10.1007/978-3-031-80943-9_27]

[CT6] Michael Omar, Bhagawan Nath, Tran Thi Ngan, and Dang Thi Khanh Linh (2025). CNN optimization for GIS mapping. In *Proceedings of the 10th International Conference on Intelligent Information Technology (ICIIT 2025)*, Hanoi, Vietnam. (In press).

[CT7] Michael Omar, Nguyen Long Giang, and Tran Thi Ngan (2025). PSO-SCNN: A Novel Hybrid Prediction of Water Quality Methodology. *Journal of the Indian Society of Remote Sensing*. (ISSN: 0974-3006, SCIE, IF: 2.2). Completed 1st round reviewing.