

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA MÔI TRƯỜNG

ĐỒ ÁN TỐT NGHIỆP
- CAPSTONE PROJECT

NGÀNH: MÔI TRƯỜNG
CHUYÊN NGÀNH: QUẢN LÝ TÀI NGUYÊN
VÀ MÔI TRƯỜNG

ĐỀ TÀI:

**ỨNG DỤNG CÁC MÔ HÌNH HỌC MÁY ĐỂ DỰ
BÁO MỨC ĐỘ TÍCH LŨY VÀ RỦI RO SINH
THÁI CỦA MỘT SỐ KIM LOẠI NẶNG TRONG
TRẦM TÍCH MẶT TẠI CỬA AN HÒA, SÔNG
TRƯỜNG GIANG, TỈNH QUẢNG NAM**

Người hướng dẫn: PGS. TS. LÊ PHƯỚC CƯỜNG
THS. NGUYỄN THÀNH TRUNG

Sinh viên thực hiện: NGUYỄN ANH ĐÀO

Số thẻ sinh viên: 117200033

Lớp: 20QLMT

Đà Nẵng, 06/2025

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA MÔI TRƯỜNG

ĐỒ ÁN TỐT NGHIỆP

NGÀNH: MÔI TRƯỜNG
CHUYÊN NGÀNH: QUẢN LÝ TÀI NGUYÊN
VÀ MÔI TRƯỜNG

ĐỀ TÀI:

ỨNG DỤNG CÁC MÔ HÌNH HỌC MÁY ĐỂ DỰ
BÁO MỨC ĐỘ TÍCH LŨY VÀ RỦI RO SINH
THÁI CỦA MỘT SỐ KIM LOẠI NẶNG TRONG
TRẦM TÍCH MẶT TẠİ CỬA AN HÒA, SÔNG
TRƯỜNG GIANG, TỈNH QUẢNG NAM

Người hướng dẫn: PGS. TS. LÊ PHƯỚC CƯỜNG
THS. NGUYỄN THÀNH TRUNG

Sinh viên thực hiện: NGUYỄN ANH ĐÀO

Số thẻ sinh viên: 117200033

Lớp: 20QLMT

Đà Nẵng, 06/2025

TÓM TẮT

Tên đề tài: Ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam.

Sinh viên thực hiện: Nguyễn Anh Đào

Số thẻ SV: 117200033

Lớp: 20QLMT

GV hướng dẫn:

PGS.TS. Lê Phước Cường

ThS. Nguyễn Thành Trung

Nội dung đề án: Tiến hành khảo sát, đánh giá mức độ ô nhiễm kim loại nặng trong trầm tích mặt và rủi ro sinh thái tại khu vực cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam. Đề án thực hiện thu thập mẫu trầm tích tại các vị trí đặc trưng, phân tích thành phần kim loại nặng để đánh giá mức độ ô nhiễm, xác định các chỉ số rủi ro sinh thái và nguy cơ tác động đến hệ sinh thái thủy vực và sức khỏe con người. Đồng thời, ứng dụng các mô hình học máy để dự báo mức độ tích lũy kim loại nặng trong trầm tích và đánh giá độ tin cậy của mô hình. Trên cơ sở kết quả nghiên cứu, đề xuất các giải pháp kiểm soát nguồn phát thải kim loại nặng, hạn chế sự tích lũy trong trầm tích, cùng các chính sách quản lý tổng hợp nhằm giảm thiểu rủi ro sinh thái.

LỜI NÓI ĐẦU VÀ CẢM ƠN

Lời đầu tiên em xin trân trọng cảm ơn các thầy cô giáo khoa Môi trường, Trường Đại học Bách khoa, Đại học Đà Nẵng và Phân viện Khoa học An toàn vệ sinh Lao động và Bảo vệ môi trường miền Trung đã tạo điều kiện tốt nhất, thuận lợi nhất cho em trong thời gian học tập, nghiên cứu và hoàn thành Capstone Project này.

Em xin gửi lời cảm ơn chân thành đến PGS.TS. Lê Phước Cường - giảng viên Bộ môn Quản lý Môi Trường, trường Đại học Bách khoa, Đại học Đà Nẵng và ThS. Nguyễn Thành Trung – Phân viện phó, Phân viện Khoa học An toàn vệ sinh Lao động và Bảo vệ môi trường miền Trung, người đã tận tình hướng dẫn, chỉ bảo em trong suốt quá trình thực hiện đề tài thực tế : ***“Ứng dụng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam”***.

Đề tài này là bước đầu đi vào thực tế, kiến thức của em còn hạn chế và còn nhiều bất ngờ. Do vậy, không tránh khỏi những thiếu sót nên em rất mong nhận được những ý kiến đóng góp quý báu của quý Thầy, Cô để kiến thức của em trong lĩnh vực này được hoàn thiện hơn.

Sau cùng, em xin kính chúc quý Thầy, Cô trong khoa Môi Trường thật dồi dào sức khỏe để tiếp tục thực hiện sứ mệnh cao đẹp của mình là truyền đạt kiến thức cho thế hệ mai sau.

Trân trọng,

Đà Nẵng, ngày 06 tháng 06 năm 2025

Sinh viên thực hiện

Nguyễn Anh Đào

CAM ĐOAN

Em xin cam đoan đây là đề tài đồ án tốt nghiệp độc lập của riêng em. Các số liệu sử dụng phân tích trong đồ án tốt nghiệp có nguồn gốc rõ ràng, đã công bố theo đúng quy định. Các kết quả nghiên cứu trong đồ án tốt nghiệp phân tích một cách trung thực, khách quan và phù hợp với thực tiễn. Các kết quả chưa từng công bố trong bất kỳ đồ án tốt nghiệp nào khác.

Sinh viên thực hiện

Nguyễn Anh Đào

MỤC LỤC

TÓM TẮT
LỜI NÓI ĐẦU VÀ CẢM ƠNi
CAM ĐOAN..... ii
DANH MỤC HÌNHvi
DANH MỤC BẢNG viii
DANH MỤC CÁC TỪ VIẾT TẮTix
MỞ ĐẦU..... 1
1. Mục đích thực hiện đề tài..... 1
2. Đối tượng và phạm vi nghiên cứu 1
2.1. Đối tượng..... 1
2.2. Phạm vi nghiên cứu 1
3. Phương pháp 1
3.1. Khảo cứu tài liệu..... 1
3.2. Phương pháp 1
4. Cấu trúc đồ án tốt nghiệp 2
4.1. Thuyết minh..... 2
4.2. Bản vẽ..... 2
CHƯƠNG 1. TỔNG QUAN VỀ KHU VỰC NGHIÊN CỨU 3
1.1. Đặt vấn đề..... 3
1.2. Tổng quan chất lượng môi trường tại cửa An Hòa sông Trường Giang 4
1.2.1. Vị trí địa lý..... 4
1.2.2. Chế độ thủy văn khu vực nghiên cứu 5
1.2.3. Tình hình kinh tế - xã hội khu vực nghiên cứu 5
1.3. Tình hình các nguồn thải công nghiệp tại khu vực huyện Núi Thành 7
1.3.1. Các khu công nghiệp 7
1.3.2. Tình hình thu gom, xử lý nước thải công nghiệp 7
1.4. Các nguồn thải chính gây ô nhiễm..... 8
1.5. Tổng quan về học máy..... 9

1.5.1. Lịch sử ra đời và các khái niệm cơ bản	9
1.5.2. Ứng dụng các mô hình học máy trong kỹ thuật môi trường	10
1.6. Ảnh hưởng của biến đổi khí hậu đến sự tích lũy kim loại nặng trong trầm tích	12
1.6.1. Ảnh hưởng của thủy triều	12
1.6.2. Ảnh hưởng của xói lở bờ biển	12
1.6.3. Ảnh hưởng của xâm nhập mặn	13
CHƯƠNG 2. KHẢO SÁT THỰC ĐỊA, ĐÁNH GIÁ MỨC ĐỘ Ô NHIỄM KIM LOẠI NẶNG TRONG TRẦM TÍCH VÀ RỦI RO SINH THÁI.....	14
2.1. Phương pháp nghiên cứu	14
2.1.1. Phương pháp thu thập và nghiên cứu tài liệu	14
2.1.2. Phương pháp tham vấn ý kiến cộng đồng	14
2.1.3. Phương pháp lấy mẫu và phân tích mẫu	15
2.2. Khảo sát thực tế cộng đồng dân cư tại khu vực nghiên cứu.....	16
2.2.1. Mục tiêu khảo sát.....	16
2.2.2. Đối tượng và phương pháp khảo sát.....	16
2.2.3. Kết quả khảo sát	17
2.2.4. Tổng hợp và phân tích tình hình mắc bệnh tại khu vực nghiên cứu	19
2.3. Phân tích các chỉ tiêu trong nước và trầm tích.....	21
2.3.1. Mục đích	21
2.3.2. Đối tượng.....	21
2.3.3. Phạm vi nghiên cứu	21
2.3.4. Nội dung	22
2.3.5. Phương pháp và thiết bị quan trắc	23
CHƯƠNG 3. ỨNG DỤNG MÔ HÌNH HỌC MÁY ĐỂ DỰ BÁO MỨC ĐỘ TÍCH LŨY KIM LOẠI NẶNG TRONG TRẦM TÍCH VÀ PHÂN TÍCH ĐỘ TIN CẬY CỦA MÔ HÌNH.....	36
3.1. Cơ sở dữ liệu đầu vào	36
3.1.1. Nguồn dữ liệu	36
3.1.2. Tiền xử lý dữ liệu	36
3.1.3. Phân chia dữ liệu	36

3.2. Lựa chọn và xây dựng mô hình học máy.....	38
3.2.1. Tải và cài đặt phần mềm.....	38
3.2.2. Các mô hình được sử dụng.....	40
3.2.3. Quy trình thực hiện mô hình	43
3.3. Xây dựng mô hình học máy	45
3.3.1. Làm sạch môi trường và cài đặt thư viện.....	45
3.3.2. Đọc dữ liệu và kiểm tra	45
3.3.3. Phân tích phân phối và mối tương quan.....	46
3.3.4. Chia dữ liệu thành tập huấn luyện và kiểm tra.....	46
3.3.5. Thiết lập công thức mô hình và kiểm tra chéo	46
3.3.6. Huấn luyện các mô hình.....	47
3.3.7. Đánh giá và so sánh mô hình.....	47
3.3.8. Trực quan hóa kết quả mô hình.....	48
3.3.9. Lựa chọn mô hình tối ưu	48
3.4. Đánh giá hiệu suất mô hình	48
3.4.1. Các chỉ số đánh giá.....	48
3.4.2. So sánh hiệu suất giữa các mô hình.....	49
3.5. Chạy dự báo nồng độ các kim loại nặng.....	61
CHƯƠNG 4. ĐỀ XUẤT GIẢI PHÁP QUẢN LÝ Ô NHIỄM VÀ GIẢM THIỂU RỦI RO SINH THÁI.....	64
4.1. Biện pháp quản lý	64
4.2. Biện pháp kỹ thuật	65
KẾT LUẬN	67
KIẾN NGHỊ	69
TÀI LIỆU THAM KHẢO.....	71
PHỤ LỤC 1
PHỤ LỤC 2
PHỤ LỤC 3

DANH MỤC HÌNH

Hình 1.1 Bản đồ khu vực nghiên cứu.....	4
Hình 2.1 Chất lượng môi trường không khí tại cửa An Hòa.....	17
Hình 2.2 Chất lượng môi trường đất tại cửa An Hòa.....	18
Hình 2.3 Chất lượng môi trường nước và trầm tích tại cửa An Hòa.....	18
Hình 2.4 Tỷ lệ mắc các bệnh của người dân tại cửa An Hòa.....	19
Hình 2.5 Thu thập số liệu tại bệnh viện đa khoa Quảng Nam.....	20
Hình 2.6 Khu vực nghiên cứu và các vị trí lấy mẫu.....	21
Hình 2.7 Một số hình ảnh lấy mẫu tại các vị trí.....	22
Hình 2.8 Phân tích các chỉ tiêu trong phòng thí nghiệm.....	24
Hình 2.9 Kết quả phân tích chỉ tiêu pH trong nước.....	27
Hình 2.10 Kết quả phân tích chỉ tiêu TSS trong nước.....	28
Hình 2.11 Kết quả phân tích chỉ tiêu NO ₂ trong nước.....	28
Hình 2.12 Kết quả phân tích chỉ tiêu NH ₄ ⁺ trong nước.....	29
Hình 2.13 Kết quả phân tích chỉ tiêu Cl ⁻ trong nước.....	29
Hình 2.14 Kết quả phân tích chỉ tiêu BOD ₅ trong nước.....	30
Hình 2.15 Kết quả phân tích chỉ tiêu COD trong nước.....	30
Hình 2.16 Kết quả phân tích chỉ tiêu As trong trầm tích.....	32
Hình 2.17 Kết quả phân tích chỉ tiêu Pb trong trầm tích.....	33
Hình 2.18 Kết quả phân tích chỉ tiêu Hg trong trầm tích.....	33
Hình 2.19 Kết quả phân tích chỉ tiêu Cd trong trầm tích.....	34
Hình 2.20 Kết quả phân tích chỉ tiêu Zn trong trầm tích.....	34
Hình 3.1 Biểu tượng R-Studio trên màn hình Desktop.....	39
Hình 3.2 Màn hình làm việc của R-Studio.....	39
Hình 3.3 Sơ đồ quy trình thực hiện mô hình học máy.....	44
Hình 3.4 Kết quả sau khi chạy kiểm tra dữ liệu.....	49
Hình 3.5 Biểu đồ phân phối của từng biến.....	50
Hình 3.6 Biểu đồ tương quan giữa các biến.....	50
Hình 3.7 Kết quả sau khi chạy mô hình.....	51
Hình 3.8 So sánh MAE giữa các mô hình.....	51

Hình 3.9 So sánh RMSE giữa các mô hình	52
Hình 3.10 So sánh R^2 giữa các mô hình	52
Hình 3.11 MAE trên tập Train theo các tỷ lệ	56
Hình 3.12 MAE trên tập Test theo các tỷ lệ	57
Hình 3.13 RMSE trên tập Train theo các tỷ lệ	57
Hình 3.14 RMSE trên tập Test theo các tỷ lệ	58
Hình 3.15 R^2 trên tập Train theo các tỷ lệ	59
Hình 3.16 R^2 trên tập Test theo các tỷ lệ	59

DANH MỤC BẢNG

Bảng 1.1 Quy hoạch các khu công nghiệp thuộc Khu KTM Chu Lai	7
Bảng 2.1 Vị trí các điểm lấy mẫu tại khu vực nghiên cứu	16
Bảng 2.2 Số liệu người mắc các bệnh về da và đường hô hấp tại Núi Thành	20
Bảng 2.3 Các chỉ tiêu và phương pháp phân tích	23
Bảng 2.4 Kết quả phân tích các chỉ tiêu trong nước.....	25
Bảng 2.5 Kết quả phân tích các chỉ tiêu trong trầm tích	31
Bảng 3.1 Chỉ số rủi ro Risk Index (RI)	38
Bảng 3.2 Kết quả chạy các mô hình theo các tỷ lệ khác nhau từ tập dữ liệu Training .	54
Bảng 3.3 Kết quả chạy các mô hình theo các tỷ lệ khác nhau từ tập dữ liệu Testing...	55
Bảng 3.4 Nồng độ các kim loại dự báo	63

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Giải thích
KCN	: Khu công nghiệp
ML	: Machine Learning (Học máy)
QCVN	: Quy chuẩn Việt Nam
BTNMT	: Bộ Tài nguyên Môi trường
BOD	: Biochemical Oxygen Demand (Nhu cầu oxy sinh hóa)
COD	: Chemical Oxygen Demand (Nhu cầu oxy hóa học)
CRAN	: Comprehensive R Archive Network
KNN	: K-Nearest Neighbors (K- Điểm dữ liệu gần nhất)
RF	: Random Forest (Rừng ngẫu nhiên)
SVM	: Support vector machine (Máy hỗ trợ vector)
MAE	: Mean Squared Error (Sai số tuyệt đối trung bình)
RMSE	: Root Mean Squared Error (Sai số toàn phương trung bình)
R^2	: R- Square (Độ tin cậy)
RI	: Risk Index (Chỉ số rủi ro sinh thái)
Igeo	: Geo-accumulation Index (Chỉ số địa hoá)
PLI	: Pollution Load Index (Chỉ số ô nhiễm)

MỞ ĐẦU

1. Mục đích thực hiện đề tài

Nghiên cứu đánh giá mức độ ô nhiễm kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam. Ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của kim loại nặng, từ đó đề xuất các biện pháp quản lý và giảm thiểu ô nhiễm.

2. Đối tượng và phạm vi nghiên cứu

2.1. Đối tượng

- Trầm tích mặt tại cửa An Hòa, sông Trường Giang.
- Rủi ro sinh thái do ô nhiễm kim loại nặng đối với hệ sinh thái thủy vực và sức khỏe con người.

2.2. Phạm vi nghiên cứu

- Khảo sát, thu thập mẫu trầm tích tại các vị trí đặc trưng trong khu vực nghiên cứu.
- Phân tích thành phần kim loại nặng trong mẫu trầm tích để đánh giá mức độ ô nhiễm.
- Ứng dụng các mô hình học máy để dự báo xu hướng tích lũy kim loại nặng.
- Đề xuất giải pháp kiểm soát ô nhiễm và giảm thiểu rủi ro sinh thái.

3. Phương pháp

3.1. Khảo cứu tài liệu

- Thu thập dữ liệu về địa lý, địa chất, thủy văn và hiện trạng môi trường tại cửa An Hòa, sông Trường Giang.
- Tổng quan về nguồn thải chính gây ô nhiễm kim loại nặng trong khu vực.
- Tổng hợp các nghiên cứu liên quan đến mô hình học máy trong đánh giá ô nhiễm môi trường.
- Xem xét tác động của biến đổi khí hậu (thủy triều, xói lở, xâm nhập mặn) đến sự tích lũy kim loại nặng trong trầm tích.

3.2. Phương pháp

- *Khảo sát thực địa*: Xác định vị trí lấy mẫu trầm tích đại diện cho khu vực nghiên cứu.

- *Phân tích mẫu*: Đo đạc các thông số kim loại nặng bằng thiết bị chuyên dụng và phân tích trong phòng thí nghiệm để đánh giá mức độ ô nhiễm.

- *Phân tích dữ liệu bằng mô hình học máy*: Ứng dụng các thuật toán học máy để dự báo mức độ tích lũy kim loại nặng và xác định các yếu tố ảnh hưởng đến quá trình tích lũy.

- *Đánh giá rủi ro sinh thái*: Tính toán các chỉ số rủi ro sinh thái và phân tích tác động đến hệ sinh thái thủy vực.

4. Cấu trúc đề án tốt nghiệp

4.1. Thuyết minh

- Chương 1: Tổng quan về khu vực nghiên cứu.
- Chương 2: Khảo sát thực địa, đánh giá mức độ ô nhiễm kim loại nặng trong trầm tích và rủi ro sinh thái.
- Chương 3: Ứng dụng mô hình học máy để dự báo mức độ tích lũy kim loại nặng trong trầm tích và phân tích độ tin cậy của mô hình.
- Chương 4: Đề xuất giải pháp quản lý ô nhiễm và giảm thiểu rủi ro sinh thái.

4.2. Bản vẽ

- BV1: Mặt bằng vị trí lấy mẫu quan trắc khu vực xung quanh cửa An Hòa, sông Trường Giang.
- BV2: Kết quả khảo sát ý kiến người dân về hiện trạng môi trường khu vực xung quanh cửa An Hòa, sông Trường Giang.
- BV3: Kết quả phân tích các chỉ tiêu trong nước.
- BV4: Kết quả phân tích các kim loại nặng trong trầm tích.
- BV5: Sơ đồ thực hiện mô hình học máy trong dự báo mức độ tích lũy kim loại nặng và rủi ro sinh thái.
- BV 6: Kết quả chạy mô hình học máy.

CHƯƠNG 1. TỔNG QUAN VỀ KHU VỰC NGHIÊN CỨU

1.1. Đặt vấn đề

Trầm tích là một thành phần quan trọng của hệ sinh thái thủy sinh, đóng vai trò như một bể chứa và nguồn phát tán các chất ô nhiễm trong môi trường nước. Trong đó, kim loại nặng là một nhóm chất ô nhiễm đặc biệt do tính chất bền vững, khả năng tích lũy sinh học và gây độc hại đối với các sinh vật cũng như con người. Việc ô nhiễm kim loại nặng trong trầm tích có thể xuất phát từ nhiều nguồn khác nhau, bao gồm hoạt động công nghiệp, nông nghiệp, nuôi trồng thủy sản và sinh hoạt của con người.

Hiện nay, vấn đề ô nhiễm kim loại nặng trong môi trường nước đang trở thành một thách thức lớn đối với nhiều quốc gia trên thế giới, trong đó có Việt Nam. Các kim loại nặng như chì (Pb), cadimi (Cd), thủy ngân (Hg), asen (As) và kẽm (Zn) khi tích lũy trong trầm tích có thể ảnh hưởng nghiêm trọng đến hệ sinh thái và sức khỏe con người thông qua chuỗi thức ăn. Đặc biệt, tại các khu vực cửa sông, nơi có sự giao thoa giữa nước ngọt và nước biển, trầm tích dễ bị ảnh hưởng bởi các hoạt động xả thải từ đất liền, dẫn đến sự tích lũy kim loại nặng trong thời gian dài.

Khu vực cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam là một trong những khu vực có tiềm năng phát triển kinh tế - xã hội mạnh mẽ, đặc biệt trong lĩnh vực nuôi trồng thủy sản và khai thác thủy sản tự nhiên. Tuy nhiên, sự gia tăng các hoạt động công nghiệp, nông nghiệp và sinh hoạt tại khu vực này đã và đang gây ra những tác động tiêu cực đến môi trường nước và trầm tích. Việc đánh giá mức độ tích lũy kim loại nặng trong trầm tích mặt tại khu vực này là rất quan trọng nhằm xác định nguy cơ ô nhiễm và rủi ro sinh thái đối với hệ sinh thái thủy vực.

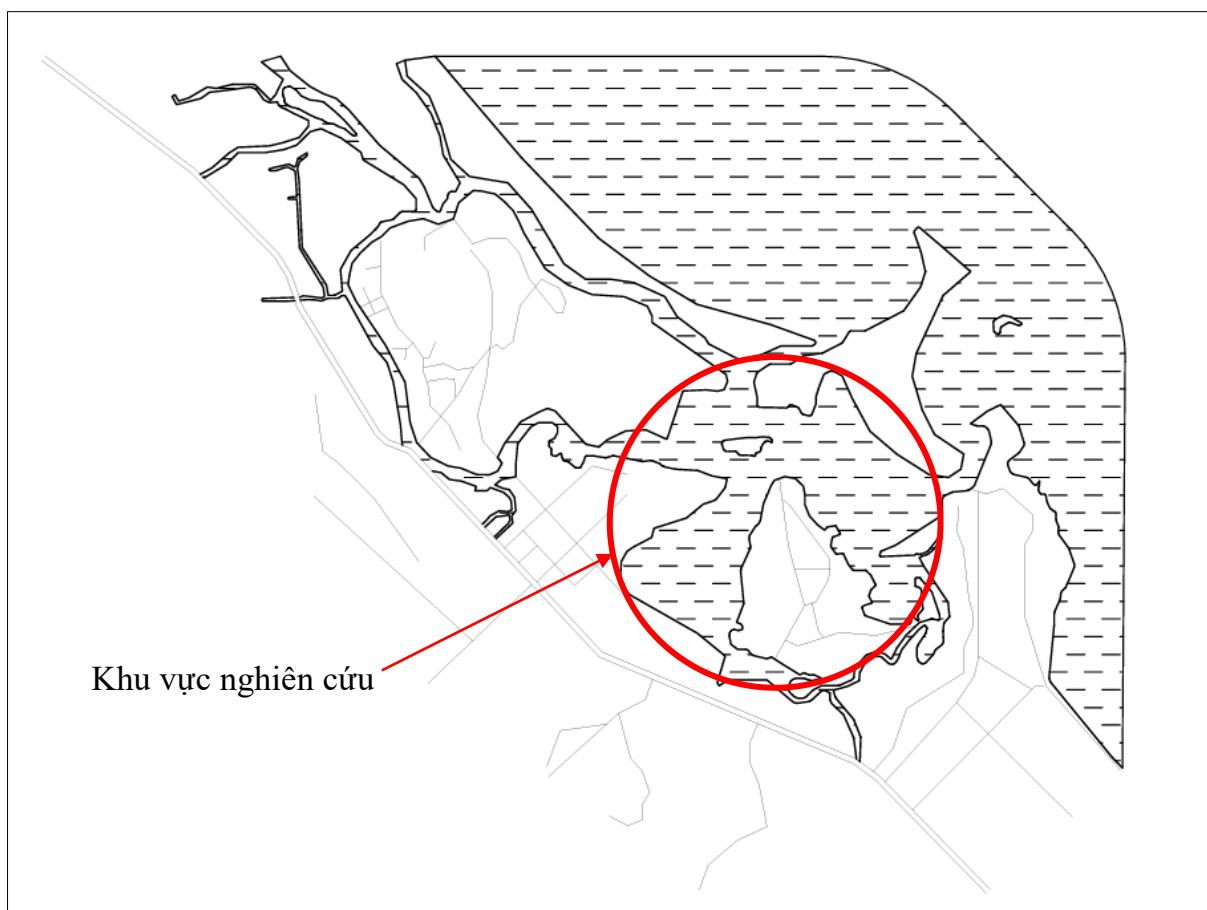
Dựa trên những vấn đề nêu trên, việc ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam là một hướng tiếp cận khoa học và hiệu quả. Phương pháp này giúp phân tích các yếu tố ảnh hưởng đến quá trình tích lũy kim loại nặng, đồng thời cung cấp công cụ dự báo đáng tin cậy để hỗ trợ công tác quản lý và bảo vệ môi trường. Trên cơ sở đó, đề tài “*Ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam*” được thực hiện nhằm cung cấp

thông tin khoa học phục vụ cho công tác kiểm soát ô nhiễm và bảo vệ hệ sinh thái khu vực nghiên cứu.

1.2. Tổng quan chất lượng môi trường tại cửa An Hòa sông Trường Giang

1.2.1. Vị trí địa lý [1]

Cửa An Hòa là khu vực phía Nam của sông Trường Giang trước khi đổ ra biển, thuộc địa phận huyện Núi Thành, tỉnh Quảng Nam.



Hình 1.1 Bản đồ khu vực nghiên cứu

Sông Trường Giang có chiều dài khoảng 60 km, chảy theo hướng Bắc - Nam, song song với đường bờ biển và bị ngăn cách với biển bởi một dải cồn cát kéo dài. Ở phía Bắc, sông nối với hạ lưu hệ thống sông Vu Gia - Thu Bồn và đổ ra biển qua cửa Đại (Hội An); phía Nam, sông kết hợp với hệ thống sông Tam Kỳ, thoát ra biển tại cửa Lở và cửa An Hòa.

Sông Tam Kỳ hình thành từ hợp lưu của 10 sông suối nhỏ có nguồn từ khu vực đồi núi phía Tây. Các dòng này chảy theo hướng Tây - Đông, hợp lưu tại khu vực Xuân Bình - Phú Thọ (xã Tam Trà, huyện Núi Thành), sau đó chuyển hướng Tây Bắc - Đông Nam và đổ ra biển qua cửa An Hòa. Diện tích lưu vực sông khoảng 800 km². Với điều

kiện lượng mưa cao và rừng đầu nguồn còn giữ được tương đối tốt, dòng chảy sông Tam Kỳ ổn định theo mùa. Lưu lượng lớn nhất ghi nhận là 20,7 m³/s.

1.2.2. Chế độ thủy văn khu vực nghiên cứu [1]

Vào mùa khô, chế độ dòng chảy của sông Trường Giang chịu ảnh hưởng chủ yếu từ dao động thủy triều. Khi triều lên, nước biển xâm nhập vào hệ thống sông qua các cửa, tạo ra dòng chảy hai chiều: đoạn sông phía bắc có xu hướng chảy về phía nam, trong khi đoạn sông phía nam lại chảy về hướng bắc. Ngược lại, khi triều xuống, dòng chảy ở đoạn phía nam chuyển động về phía Cửa Lở và An Hòa theo hướng nam, trong khi đoạn phía bắc đổ ra Cửa Đại theo hướng bắc. Đặc biệt, đoạn sông Trường Giang chảy qua địa phận huyện Thăng Bình, nằm ở khoảng giữa toàn tuyến, có hiện tượng dòng chảy không ổn định, có xu hướng dao động và phân vân giữa hai hướng.

Trong mùa mưa lũ, đặc biệt là khi xuất hiện các đợt lũ lớn, dòng chảy của sông Trường Giang phụ thuộc vào lưu lượng nước từ hai hệ thống sông chính là Vu Gia – Thu Bồn và Tam Kỳ – An Tân. Vị trí của đoạn sông có dòng chảy dẹt dằng có thể dịch chuyển theo hướng bắc hoặc nam, tùy thuộc vào tương quan cường độ dòng chảy của hai hệ thống sông nói trên.

Khu vực Cửa An Hòa chịu tác động trực tiếp từ thủy triều vùng ven biển. Chế độ thủy triều ở đây thuộc loại bán nhật triều không đều, trong đó nhật triều thường xuất hiện từ 10 đến 15 ngày mỗi tháng, các ngày còn lại là bán nhật triều. Mực nước triều trung bình tại khu vực dao động khoảng 1,2m. Biên độ triều trong các kỳ triều lớn dao động từ 1,0 đến 1,5m, còn trong kỳ triều kém dao động từ 0,4 đến 0,8m. Tốc độ dòng chảy trung bình trong khu vực là 0,2 – 0,3m/s, và có thể đạt cực đại từ 2 đến 2,5m/s.

1.2.3. Tình hình kinh tế - xã hội khu vực nghiên cứu

Trong giai đoạn 2016–2020, huyện Núi Thành đã có những bước phát triển đáng kể về kinh tế - xã hội. Bám sát định hướng và chỉ đạo của Trung ương và tỉnh Quảng Nam, huyện đã chủ động triển khai đồng bộ nhiều giải pháp phát triển, tập trung vào những mục tiêu mang tính chiến lược như cải thiện hạ tầng, thúc đẩy đầu tư và chuyển dịch cơ cấu kinh tế. Quá trình xây dựng nông thôn mới được đẩy mạnh, kết hợp với việc tháo gỡ khó khăn trong phát triển đô thị và tạo điều kiện thuận lợi cho hoạt động sản xuất - kinh doanh. Nhờ vậy, tốc độ tăng trưởng kinh tế bình quân hằng năm đạt trên 10%, đưa quy mô kinh tế huyện vào nhóm dẫn đầu của tỉnh. Tính đến năm 2020, tổng giá trị sản xuất của huyện đạt khoảng 80.095 tỷ đồng, gấp đôi so với năm 2015. Cơ cấu kinh tế chuyển biến theo hướng tích cực, trong đó ngành phi nông nghiệp chiếm tỷ trọng vượt trội, gần 95%. Thu ngân sách trên địa bàn tăng trưởng khá, với tổng thu giai đoạn 2016–

2020 đạt gần 49.200 tỷ đồng, chiếm hơn 43% tổng thu của tỉnh. Mức chi ngân sách được đảm bảo, tập trung cho các mục tiêu phát triển KT-XH và quốc phòng - an ninh.

Công nghiệp - xây dựng tiếp tục giữ vai trò là trụ cột chính trong tăng trưởng kinh tế. Tốc độ phát triển của lĩnh vực này đạt gần 9,6%/năm, đóng góp lớn vào giá trị sản xuất toàn huyện. Bên cạnh Khu kinh tế mở Chu Lai – trung tâm động lực công nghiệp của huyện – các cụm công nghiệp khác như Trảng Tôn và Nam Chu Lai cũng thu hút hàng loạt dự án sản xuất mới. Đến năm 2020, giá trị sản xuất của ngành đạt gần 71.000 tỷ đồng, gấp 2,3 lần so với năm 2015, chiếm khoảng 85% tổng giá trị sản xuất. Trên địa bàn có hơn 110 dự án công nghiệp đang hoạt động với tổng vốn đăng ký hơn 39.000 tỷ đồng, tạo công ăn việc làm ổn định cho hơn 17.000 người. Các cụm công nghiệp của huyện cũng ghi nhận mức thu hút đầu tư tốt với hàng chục dự án sản xuất, giải quyết việc làm cho hơn 1.300 lao động.

Lĩnh vực thương mại - dịch vụ phát triển nhanh cả về quy mô lẫn chất lượng, với tốc độ tăng trưởng bình quân gần 14% mỗi năm. Hạ tầng thương mại được mở rộng khắp vùng nông thôn và miền núi, giúp lưu thông hàng hóa thuận tiện và đáp ứng tốt nhu cầu tiêu dùng. Các dịch vụ như vận tải, viễn thông, tài chính - ngân hàng phát triển ổn định. Du lịch bước đầu được định hình theo hướng phát triển phía Nam tỉnh Quảng Nam, khai thác lợi thế về cảnh quan và lịch sử để thu hút du khách. Tính đến năm 2020, giá trị sản xuất của ngành thương mại - dịch vụ đạt khoảng 6.280 tỷ đồng, chiếm khoảng 10% quy mô kinh tế huyện.

Ngành nông - lâm - thủy sản tuy chiếm tỷ trọng nhỏ nhưng vẫn duy trì tốc độ tăng trưởng ổn định. Chính quyền địa phương chú trọng đầu tư cơ sở hạ tầng nông thôn, tổ chức lại sản xuất theo hướng hàng hóa, và áp dụng các mô hình canh tác hiệu quả. Diện mạo nông thôn từng bước khởi sắc, đời sống nhân dân được cải thiện rõ rệt. Hợp tác xã và tổ hợp tác hoạt động hiệu quả hơn, hỗ trợ người dân về vốn, kỹ thuật và đầu ra sản phẩm. Đến năm 2020, tổng giá trị sản xuất của ngành này đạt gần 2.880 tỷ đồng, với tốc độ tăng trưởng hơn 4,5% mỗi năm. Sản lượng lương thực có hạt đạt gần 44.000 tấn, tăng hơn 12% so với năm 2016. Huyện có 12/15 xã đạt chuẩn nông thôn mới, vượt chỉ tiêu nghị quyết Đại hội đề ra.

Song hành với tăng trưởng kinh tế là việc đầu tư mạnh mẽ vào hệ thống kết cấu hạ tầng. Nhiều công trình quan trọng như công viên, đường giao thông, chợ, khu dân cư và công trình thể thao được xây dựng mới hoặc nâng cấp. Đặc biệt, nhiều dự án quy mô lớn do Trung ương, tỉnh và các doanh nghiệp đầu tư như đường cao tốc Đà Nẵng – Quảng Ngãi, cảng biển Chu Lai, sân bay Chu Lai, nút giao thông 2 tầng và đường ven

biển Võ Chí Công... đã tạo điều kiện thuận lợi thúc đẩy liên kết vùng, phát triển công nghiệp - dịch vụ, cũng như góp phần thay đổi diện mạo đô thị Núi Thành một cách rõ rệt. [2]

1.3. Tình hình các nguồn thải công nghiệp tại khu vực huyện Núi Thành

1.3.1. Các khu công nghiệp

Ngày 05/6/2003 Thủ tướng Chính phủ ban hành Quyết định số 108/2003/QĐ-TTg về việc thành lập Khu kinh tế mở Chu Lai (Khu KTM Chu Lai) và điều chỉnh tại Quyết định 1737/QĐ-TTg ngày 13/12/2018, theo đó, tổng diện tích đất Khu kinh tế không thay đổi là 27.040 ha, nhưng ranh giới Khu kinh tế được điều chỉnh gồm 19 xã, phường trải dài trên 2 huyện và 1 thành phố: huyện Núi Thành, Thăng Bình và thành phố Tam Kỳ.

Bảng 1.1 Quy hoạch các khu công nghiệp thuộc Khu KTM Chu Lai

TT	Tên khu chức năng	Quy mô (ha)	Địa điểm
1	KCN Bắc Chu Lai	700	Xã Tam Hiệp, huyện Núi Thành
2	KCN Tam Anh	1.545	Xã Tam Anh Nam, huyện Núi Thành
3	KCN Tam Hiệp	530	Xã Tam Hiệp, huyện Núi Thành
4	KCN Tam Thăng	800	01 phần xã Tam Thăng, Tam Kỳ và 01 phần xã Bình Nam, huyện Thăng Bình
5	KCN khí – năng lượng, sản phẩm sau khí	410	Xã Tam Quang, huyện Núi Thành
6	KCN Nam Thăng Bình	655	Xã Bình Trung, Bình Nam, Bình Sa, huyện Thăng Bình
7	KCN Công nghệ cao Thăng Bình	310	Xã Bình Tú, Bình Trung, Bình Sa, huyện Thăng Bình
8	KCN Thaco Trường Hải	451	Xã Tam Hiệp, huyện Núi Thành
	Tổng cộng	5.401	

1.3.2. Tình hình thu gom, xử lý nước thải công nghiệp

Các khu công nghiệp trên địa bàn đã có những bước tiến quan trọng trong công tác bảo vệ môi trường thông qua việc đầu tư hệ thống xử lý nước thải tập trung. Thống kê

mới nhất cho thấy 85,7% số KCN đang hoạt động (6/7) đã xây dựng và vận hành các nhà máy xử lý nước thải chuyên nghiệp. Tổng công suất xử lý của các hệ thống này đạt khoảng 12.064 m³/ngày đêm, đáp ứng nhu cầu xử lý nước thải phát sinh từ hoạt động sản xuất công nghiệp.

Các hệ thống xử lý nước thải tại đây đều được thiết kế và vận hành theo tiêu chuẩn nghiêm ngặt. Điển hình như hệ thống tại KCN Bắc Chu Lai (1.900 m³/ngày đêm) đã hoạt động ổn định từ năm 2012, hay hệ thống hiện đại tại KCN Tam Thăng với công suất lên tới 28.000 m³/ngày đêm. Đáng chú ý, tất cả các hệ thống này đều đảm bảo chất lượng nước đầu ra đạt chuẩn cột A theo QCVN 40:2011/BTNMT, một số còn có khả năng tái sử dụng nước sau xử lý.

Công tác giám sát môi trường cũng được thực hiện chặt chẽ thông qua hệ thống quan trắc tự động liên tục, kết nối trực tiếp với cơ quan quản lý môi trường địa phương. Hầu hết các KCN (5/7) đã được cấp giấy phép xả thải hợp lệ, trong đó yêu cầu bắt buộc là các doanh nghiệp thành viên phải xử lý sơ bộ nước thải đạt chuẩn trước khi đầu nối vào hệ thống tập trung. Chỉ có một số ít trường hợp đặc biệt như Công ty CP Sô đa Chu Lai được phép xả thải riêng do đặc thù công nghệ sản xuất. [3]

1.4. Các nguồn thải chính gây ô nhiễm

Tại khu vực cửa An Hoà, xung quanh có rất nhiều hoạt động công nghiệp đang xả thải vào lưu vực sông Trường Giang như KCN Bắc Chu Lai, KCN Thaco Trường Hải, Bệnh viện Đa khoa Trung ương Quảng Nam,... đây là các nguồn thải gây nguy cơ phát thải các chất ô nhiễm hữu cơ khó phân huỷ tại khu vực nghiên cứu. Bên cạnh đó, hoạt động nạo vét luồng lạch đang thực hiện tại khu vực sẽ gây xáo trộn lớp trầm tích và ảnh hưởng đến kết quả nghiên cứu.

Tình trạng ô nhiễm tại sông Trường Giang đang trở nên nghiêm trọng, đặc biệt là vấn đề ứ đọng rác thải tại khu vực hạ lưu. Rác thải chủ yếu gồm bao ni lông, chai nhựa, ngư lưới cụ bị cuốn trôi và tích tụ thành từng đống dọc bờ biển, gây ô nhiễm môi trường và mất mỹ quan. Nguyên nhân chính là do sông Trường Giang chảy qua nhiều khu vực dân cư trước khi đổ ra biển tại xã Tam Hải, trở thành nơi tập kết rác từ các xã ven biển thuộc TP. Tam Kỳ và huyện Núi Thành. Ngoài ra, cửa biển Tam Hải là khu vực neo đậu của nhiều tàu thuyền đánh bắt xa bờ, khiến rác thải từ hoạt động ngư nghiệp như bao bì nhựa, đồ hộp, lưới đánh cá bị vứt bừa bãi, khó kiểm soát, làm trầm trọng thêm tình trạng ô nhiễm.

Hoạt động nuôi trồng thủy sản đang tạo ra áp lực lớn đối với môi trường nước sông, đặc biệt tại các khu vực nuôi trồng tập trung như hạ lưu các sông thuộc hệ thống sông

Tam Kỳ và đầu phía Nam sông Trường Giang (thuộc huyện Núi Thành). Nếu không có giải pháp xử lý nước thải hợp lý, ô nhiễm nước sông sẽ trở thành vấn đề nghiêm trọng, không chỉ ảnh hưởng trực tiếp đến nguồn nước cấp cho các ao nuôi mà còn tác động tiêu cực đến sự phát triển của ngành nuôi trồng thủy sản.

Ngoài ra, ô nhiễm nước sông tại khu vực này còn gián tiếp ảnh hưởng đến du lịch tại xã đảo Tam Hải do làm suy giảm chất lượng nước tại các khu vực cửa Lở, cửa An Hòa và vùng biển ven bờ. Bên cạnh đó, việc khai thác nước ngầm phục vụ nuôi trồng thủy sản cũng đang làm gia tăng tình trạng xâm nhập mặn vào nguồn nước ngầm – vốn đã khan hiếm tại vùng ven biển. Hiện tượng này đã bắt đầu xuất hiện tại một số khu vực ở xã Tam Tiến, ảnh hưởng đến nguồn nước sinh hoạt của người dân và có nguy cơ gia tăng nếu không có biện pháp kiểm soát chặt chẽ.[4]

1.5. Tổng quan về học máy

1.5.1. Lịch sử ra đời và các khái niệm cơ bản

a. Lịch sử ra đời

Học máy (Machine Learning) xuất hiện vào cuối thập niên 1980 và đầu những năm 1990, là một nhánh của trí tuệ nhân tạo (AI). Công nghệ này giúp máy tính có khả năng thực hiện các nhiệm vụ cụ thể bằng cách phân tích dữ liệu và đưa ra quyết định. Thay vì được lập trình cứng nhắc, các thuật toán trong học máy được xây dựng sao cho chúng có thể tự học và nâng cao hiệu suất khi xử lý thêm dữ liệu theo thời gian.

b. Các khái niệm

Học máy (Machine Learning) là lĩnh vực cho phép máy tính tự động học hỏi và cải thiện hiệu suất thực hiện một nhiệm vụ nhất định thông qua kinh nghiệm và dữ liệu. Đây là một nhánh quan trọng của trí tuệ nhân tạo (AI), giúp các hệ thống có thể học trực tiếp từ ví dụ, dữ liệu hoặc trải nghiệm mà không cần phải được lập trình chi tiết cho từng tác vụ cụ thể.

Thay vì viết các chương trình với các bước cố định, học máy sử dụng các thuật toán để phân tích thông tin đầu vào, rút ra quy luật và áp dụng vào việc đưa ra dự đoán hoặc quyết định đối với các tình huống tương tự. Khi hệ thống thực hiện một nhiệm vụ nhiều lần, hiệu quả xử lý sẽ được cải thiện nhờ khả năng học hỏi liên tục.

Việc huấn luyện máy tính trong học máy được thực hiện bằng cách cung cấp một lượng lớn dữ liệu kết hợp với các thuật toán phù hợp, từ đó cho phép hệ thống hiểu và thực hiện được nhiệm vụ một cách ngày càng chính xác hơn. Nếu không có học máy, trí

tuệ nhân tạo sẽ gặp nhiều hạn chế, vì học máy chính là yếu tố mang lại khả năng tự thích nghi và phát triển cho các hệ thống thông minh.

c. Phân loại học máy [5]

Học máy được phân làm 4 nhóm:

Học có giám sát (Supervised Learning) là phương pháp mà trong đó mô hình được huấn luyện dựa trên một tập dữ liệu đã được gán nhãn — nghĩa là mỗi đầu vào đều đi kèm với đầu ra mong muốn. Mục tiêu của mô hình là học cách ánh xạ từ dữ liệu đầu vào sang kết quả đầu ra, để từ đó có thể dự đoán chính xác cho dữ liệu mới. Đây là loại thuật toán được sử dụng phổ biến nhất trong học máy.

Học không giám sát (Unsupervised Learning) lại không có sẵn nhãn đầu ra. Mô hình sẽ phải tự phát hiện ra các cấu trúc ẩn trong dữ liệu, chẳng hạn như việc phân cụm các điểm dữ liệu có đặc điểm tương đồng, hoặc rút gọn số chiều dữ liệu để tối ưu việc lưu trữ và xử lý.

Học bán giám sát (Semi-supervised Learning) là phương pháp kết hợp giữa học có giám sát và học không giám sát. Trong trường hợp này, chỉ một phần nhỏ dữ liệu có nhãn, còn phần lớn là dữ liệu chưa được gán nhãn. Phương pháp này đặc biệt hữu ích khi việc gán nhãn tốn kém về thời gian và chi phí, chẳng hạn như với dữ liệu ảnh y tế đòi hỏi chuyên gia phân tích. Trong thực tế, nhiều bài toán học máy rơi vào nhóm này vì dữ liệu không nhãn thường có thể thu thập dễ dàng hơn từ các nguồn như internet.

Học tăng cường (Reinforcement Learning) là kỹ thuật mà trong đó hệ thống sẽ học cách hành động thông qua quá trình thử – sai để tối đa hóa phần thưởng trong một môi trường nhất định. Mỗi hành động của hệ thống sẽ được phản hồi bằng một phần thưởng hoặc hình phạt, và qua thời gian, hệ thống sẽ học được chiến lược tối ưu. Phương pháp này thường được ứng dụng trong các trò chơi hoặc bài toán điều khiển tự động, nơi cần đưa ra quyết định nối tiếp để đạt được kết quả tốt nhất.

1.5.2. Ứng dụng các mô hình học máy trong kỹ thuật môi trường

Trong bối cảnh biến đổi khí hậu và tác động của con người ngày càng gia tăng, việc giám sát, đánh giá và quản lý tài nguyên môi trường trở nên cấp thiết hơn bao giờ hết. Các phương pháp truyền thống trong nghiên cứu môi trường thường đòi hỏi nhiều thời gian, nhân lực và chi phí cao, trong khi khả năng phân tích dữ liệu lớn còn hạn chế. Sự phát triển của các mô hình học máy (Machine Learning - ML) đã mở ra những hướng đi mới, giúp tối ưu hóa quá trình thu thập, xử lý và dự báo dữ liệu môi trường một cách chính xác và hiệu quả hơn.

Ứng dụng của học máy trong lĩnh vực môi trường rất đa dạng, từ dự báo chất lượng không khí, chất lượng nước, giám sát tài nguyên thiên nhiên, cho đến phát hiện các hiện tượng cực đoan như cháy rừng, xói lở đất hay lũ lụt. Các mô hình ML có khả năng phát hiện quy luật tiềm ẩn trong dữ liệu, từ đó giúp các nhà khoa học và nhà quản lý đưa ra quyết định chính xác hơn trong việc bảo vệ môi trường và phát triển bền vững.

Nhiều nghiên cứu đã chứng minh rằng ML không chỉ nâng cao hiệu quả giám sát môi trường, mà còn giúp cảnh báo sớm các rủi ro và đề xuất giải pháp ứng phó kịp thời. Để minh họa cho tính ứng dụng của ML trong thực tế, có thể xem xét hai nghiên cứu điển hình: dự báo chất lượng môi trường tại Bán đảo Cà Mau và Hội An, trong đó các mô hình học máy đã được triển khai thành công để phân tích và dự báo xu hướng môi trường. [6]

Nghiên cứu về dự báo chất lượng nước mặt tại Bán đảo Cà Mau đã sử dụng các mô hình học máy để xác định chỉ số chất lượng nước (WQI) dựa trên một số thông số quan trắc tối thiểu. Việc này giúp giảm chi phí phân tích mẫu nhưng vẫn đảm bảo độ chính xác cao trong đánh giá chất lượng nước. Các thuật toán như **Gradient Boosting** và **XGBoost** được áp dụng để tối ưu hóa quá trình dự báo, trong đó Gradient Boosting đạt **hệ số xác định $R^2 = 0,973$** , chứng tỏ khả năng mô phỏng chính xác xu hướng biến đổi chất lượng nước theo thời gian và không gian. [7]

Trong nghiên cứu tại Hội An, các mô hình học máy được sử dụng để phân tích và dự báo chỉ số chất lượng nước dưới đất (**GWQI**) tại khu vực lân cận bãi rác Cẩm Hà. Việc đánh giá chất lượng nước dưới đất có ý nghĩa quan trọng trong việc phát hiện nguy cơ ô nhiễm từ kim loại nặng và hóa chất độc hại, giúp đưa ra các biện pháp bảo vệ nguồn nước. Các mô hình Random Forest, SVM, KNN và Cubist đã được so sánh, trong đó Cubist đạt $R^2 = 0,96$, chứng tỏ độ chính xác cao nhất trong dự báo chất lượng nước dưới đất.

Hai nghiên cứu trên cho thấy rằng việc ứng dụng học máy trong lĩnh vực môi trường là hết sức cần thiết. Học máy không chỉ mở ra khả năng xử lý và phân tích khối lượng lớn dữ liệu một cách hiệu quả, mà còn giúp phát hiện sớm các nguy cơ ô nhiễm và hỗ trợ các quyết định quản lý kịp thời. Trong bối cảnh các hệ sinh thái ngày càng chịu nhiều áp lực từ biến đổi khí hậu và hoạt động của con người, việc tích hợp các mô hình học máy vào công tác quan trắc và quản lý tài nguyên là một hướng đi cấp thiết nhằm hướng tới phát triển bền vững.

1.6. Ảnh hưởng của biến đổi khí hậu đến sự tích lũy kim loại nặng trong trầm tích

1.6.1. Ảnh hưởng của thủy triều

Thủy triều là một yếu tố quan trọng ảnh hưởng đến sự phân bố và động thái của trầm tích cũng như kim loại nặng trong môi trường nước. Sự dao động liên tục của mực nước do thủy triều có thể làm xáo trộn lớp trầm tích bề mặt, dẫn đến việc giải phóng kim loại nặng tích tụ lâu ngày vào cột nước. Quá trình này đặc biệt quan trọng ở các khu vực cửa sông, nơi có sự luân phiên giữa nước ngọt và nước biển, tạo ra sự thay đổi lớn trong điều kiện hóa học của trầm tích. Khi thủy triều lên, trầm tích có thể bị cuốn theo dòng nước và lắng đọng lại khi thủy triều xuống, làm thay đổi vị trí và nồng độ kim loại nặng trong trầm tích.

Ngoài ra, thủy triều còn ảnh hưởng đến điều kiện oxy hóa-khử của môi trường trầm tích. Trong điều kiện yếm khí (thiếu oxy), nhiều kim loại nặng như sắt (Fe) và mangan (Mn) có thể bị khử và hòa tan vào nước, làm gia tăng nồng độ kim loại trong cột nước. Ngược lại, khi môi trường có nhiều oxy hơn, các kim loại này có thể kết tủa và lắng xuống, tích lũy trở lại trong trầm tích. Sự thay đổi liên tục này khiến kim loại nặng trong trầm tích không ổn định, làm tăng nguy cơ ô nhiễm môi trường nước, đặc biệt khi có các hoạt động nhân sinh như nạo vét, giao thông thủy hay xây dựng công trình ven biển.

1.6.2. Ảnh hưởng của xói lở bờ biển

Xói lở bờ biển là một trong những hệ quả nghiêm trọng của biến đổi khí hậu, đặc biệt ở các vùng ven biển có tốc độ nước biển dâng cao và chịu tác động mạnh từ bão, sóng lớn. Khi bờ biển bị xói lở, một lượng lớn đất, cát và trầm tích bị cuốn ra môi trường nước, làm tăng lượng trầm tích lơ lửng và kéo theo sự vận chuyển của kim loại nặng. Các kim loại nặng có thể tồn tại trong lớp đất bề mặt của bờ biển, được tích lũy từ các nguồn ô nhiễm như chất thải công nghiệp, nông nghiệp và sinh hoạt. Khi đất bị xói mòn và trôi ra sông, biển, những kim loại này có thể hòa tan vào nước hoặc bám vào các hạt trầm tích nhỏ, di chuyển xa hơn so với vị trí ban đầu.

Ngoài ra, quá trình xói lở còn làm thay đổi mô hình lắng đọng của trầm tích. Khi một khu vực bị mất trầm tích do xói lở, vùng khác có thể nhận thêm lượng bùn cát lớn hơn, gây biến động lớn trong việc tích lũy kim loại nặng. Điều này đặc biệt quan trọng trong các hệ sinh thái ven biển, nơi mà sự cân bằng giữa lắng đọng và xói mòn quyết định mức độ ô nhiễm của trầm tích. Nếu kim loại nặng được vận chuyển đến các khu vực sinh thái nhạy cảm như rừng ngập mặn hay rạn san hô, chúng có thể gây ra những tác động tiêu cực đối với các loài sinh vật sống ở đó, ảnh hưởng đến toàn bộ chuỗi thức ăn trong hệ sinh thái biển.

1.6.3. Ảnh hưởng của xâm nhập mặn

Xâm nhập mặn là hiện tượng nước biển xâm lấn vào các vùng nước ngọt, làm thay đổi đáng kể các đặc tính hóa lý của môi trường nước và trầm tích. Sự thay đổi độ mặn có thể ảnh hưởng trực tiếp đến quá trình tích lũy và phát tán kim loại nặng trong trầm tích. Khi độ mặn tăng lên, sự trao đổi ion giữa nước biển và trầm tích có thể làm hòa tan một số kim loại nặng như chì (Pb), cadmium (Cd), đồng (Cu) và kẽm (Zn), khiến chúng trở nên dễ dàng di chuyển hơn trong cột nước. Điều này có thể dẫn đến sự gia tăng nồng độ kim loại nặng trong nước, làm tăng nguy cơ ô nhiễm đối với hệ sinh thái và nguồn nước sử dụng cho con người.

Ngoài ra, xâm nhập mặn còn ảnh hưởng đến thành phần vi sinh vật trong trầm tích. Các nhóm vi khuẩn yếm khí có vai trò quan trọng trong quá trình chuyển hóa kim loại nặng, có thể bị suy giảm hoặc thay đổi do độ mặn cao. Điều này dẫn đến sự thay đổi trong cơ chế cố định hoặc giải phóng kim loại nặng, từ đó làm biến động mức độ ô nhiễm của trầm tích theo thời gian. Nếu không có các biện pháp kiểm soát phù hợp, sự gia tăng của xâm nhập mặn do biến đổi khí hậu có thể làm tăng nguy cơ ô nhiễm kim loại nặng, ảnh hưởng đến hệ sinh thái thủy sinh và chất lượng nước ngầm ở các khu vực ven biển.

CHƯƠNG 2. KHẢO SÁT THỰC ĐỊA, ĐÁNH GIÁ MỨC ĐỘ Ô NHIỄM KIM LOẠI NẶNG TRONG TRẦM TÍCH VÀ RỦI RO SINH THÁI

2.1. Phương pháp nghiên cứu

2.1.1. Phương pháp thu thập và nghiên cứu tài liệu

Thu thập các tài liệu về tính chất, nguồn gây ô nhiễm và độc tính cũng như sự lan truyền của các hợp chất nghiên cứu trong môi trường bằng việc tham khảo các tài liệu trong các giáo trình, sổ tay, các bài báo, các đề tài nghiên cứu trong và ngoài nước có liên quan đến đối tượng nghiên cứu.

Thu thập các tài liệu liên quan đến phương pháp đánh giá rủi ro sinh thái của kim loại nặng trong trầm tích, cũng như các nghiên cứu ứng dụng mô hình học máy trong dự báo ô nhiễm môi trường, nhằm làm cơ sở lý thuyết và thực tiễn cho việc xây dựng và áp dụng mô hình trong nghiên cứu.

2.1.2. Phương pháp tham vấn ý kiến cộng đồng

Tham vấn ý kiến cộng đồng trong nghiên cứu được hiểu là quá trình trao đổi trực tiếp với các dân cư sống lân cận tại khu vực cửa An Hòa, đặc biệt là người dân đang khai thác nguồn lợi thủy sản tại khu vực này. Hoạt động tham vấn đã giúp tác giả nắm rõ hơn về khu vực nghiên cứu, làm cơ sở để tiến hành khảo sát, bố trí các vị trí lấy mẫu cho phù hợp với điều kiện, đặc điểm thủy văn của khu vực. Số lượng người dân tham gia phỏng vấn theo công thức trong thông tư 03/2025/TT-BLĐTĐBXH. [8]

Công thức xác định:

$$n_0 = \frac{N}{1 + N \times e^2}$$

Trong đó N là dân số của (N = 150340), chọn sai số cho phép e = 15%, Do giới hạn về thời gian, nguồn lực và mục tiêu nghiên cứu hiện tại chỉ nhằm đánh giá sơ bộ ban đầu, nên sai số 15% được chấp nhận trong trường hợp này. Mức sai số này thường được xem là phù hợp cho các nghiên cứu mang tính chất thăm dò, giúp tiết kiệm chi phí và thời gian trong khi vẫn đảm bảo độ tin cậy nhất định cho kết quả thu được.. Khi đó:

$$n_0 = \frac{150340}{1 + 150340 \times 0.15^2} = 45 \text{ (người)}$$

Để đảm bảo tính đại diện và đủ cơ sở kết luận về hiện trạng sức khỏe môi trường, sinh viên tiến hành khảo sát 50 người dân tại khu vực nghiên cứu. Kết quả thu thập được

sẽ mang tính tham khảo và sẽ giúp cung cấp những thông tin sơ bộ, hỗ trợ việc hình thành các đánh giá ban đầu cho nghiên cứu.

2.1.3. Phương pháp lấy mẫu và phân tích mẫu

❖ Lấy mẫu

- Thời gian lấy mẫu: Tháng 4/2025 và tháng 5/2025.

- Số lượng mẫu: 15 mẫu/đợt

- Mẫu trầm tích mặt: Mẫu được lấy với độ sâu 0 - 10 cm, được lấy bằng thiết bị lấy mẫu trầm tích kiểu Ponar. Dụng cụ lấy mẫu được thả xuống và gắp một lượng trầm tích xác định tùy thuộc vào thể tích của gầu, áp lực của nước và tính chất của trầm tích. Mẫu trầm tích được trộn đều bằng khay và chuyển vào bình chứa mẫu phù hợp. Mẫu vận chuyển và bảo quản theo TCVN 6663-15:2004. [9]

Sau khi lấy mẫu cần chuyển ngay các mẫu trầm tích được lấy túi zip. Mẫu trầm tích cần được phân tích trong khoảng thời gian 1 tuần. Nếu bảo quản mẫu trong thời gian lâu hơn, cần xử lý sơ bộ mẫu bằng cách làm khô không khí (tránh ánh nắng mặt trời) hoặc làm khô trong tủ làm khô lạnh. Sau đó mẫu cần được bảo quản trong tủ lạnh sâu. Cần tiến hành chiết mẫu và phân tích càng sớm càng tốt.

Các vị trí lấy mẫu trong phạm vi khu vực nghiên cứu, được lựa chọn dựa dựa trên các điều tra khảo sát thực tế trong quá trình đi lấy mẫu. Các điểm lấy mẫu được dựa trên cơ sở là những điểm gần với nguồn thải, những điểm có nguy cơ hoặc tiềm năng là những điểm có thể bị ô nhiễm do các hoạt động giao thông, khai thác, xả nước thải,...

Bảng 2.1 Vị trí các điểm lấy mẫu tại khu vực nghiên cứu

STT	Mẫu	Kinh độ	Vĩ độ	Mô tả vị trí lấy mẫu
1	S-1	15°28'01.1"N	108°38'01.9"E	Gần khu vực xả thải nhà máy Chu Lai
2	S-2	15°27'15.4"N	108°37'36.3"E	Gần cửa xả KĐT Vĩnh An Hòa
3	S-3	15°26'53.3"N	108°38'18.3"E	Gần cửa xả KĐT Vĩnh An Hòa
4	S-4	15°26'34.5"N	108°38'58.3"E	Cầu An Tân
5	S-5	15°27'51.2"N	108°38'43.3"E	Khu vực sông xã Tam Giang bờ phía Tây
6	S-6	15°28'25.9"N	108°39'00.1"E	Khu vực ngã 3 sông
7	S-7	15°28'27.6"N	108°39'09.0"E	Khu vực cảng cá An Hòa
8	S-8	15°28'00.5"N	108°39'34.9"E	Khu vực nuôi tôm xã Tam Giang
9	S-9	15°27'27.0"N	108°40'00.0"E	Khu vực nuôi tôm, cá xã Tam Giang
10	S-10	15°28'05.3"N	108°40'40.5"E	Khu vực dân cư An Hải Tây
11	S-11	15°28'19.2"N	108°40'51.1"E	Bến phà Tam Quang, nhiều thuyền neo đậu
12	S-12	15°28'30.8"N	108°41'04.5"E	Cảng Kỳ Hà, nhiều thuyền neo đậu
13	S-13	15°28'31.8"N	108°40'28.2"E	Gần bến phà Tam Hải
14	S-14	15°28'51.9"N	108°40'01.5"E	Cách bến phà Tam Hải 50m về phía Tây Nam
15	S-15	15°28'51.9"N	108°40'01.5"E	Xã Tam Hải

2.2. Khảo sát thực tế cộng đồng dân cư tại khu vực nghiên cứu

2.2.1. Mục tiêu khảo sát

Khảo sát cộng đồng dân cư tại khu vực cửa An Hòa nhằm ghi nhận các thông tin thực tế liên quan đến nhận thức và phản ánh của người dân về tình trạng ô nhiễm môi trường, ảnh hưởng của môi trường đến sức khỏe và sinh kế, cũng như sự thay đổi chất lượng môi trường trong những năm gần đây. Đây là nguồn thông tin quan trọng nhằm hỗ trợ quá trình đánh giá rủi ro sinh thái và lựa chọn vị trí lấy mẫu phù hợp.

2.2.2. Đối tượng và phương pháp khảo sát

Đối tượng khảo sát: Gồm 50 hộ dân sinh sống tại khu vực cửa An Hòa.

Phương pháp khảo sát: Phỏng vấn trực tiếp thông qua bảng hỏi cấu trúc kết hợp với phỏng vấn bán cấu trúc. Nội dung bảng hỏi tập trung vào việc ghi nhận nhận thức, đánh giá và phản ánh của người dân về hiện trạng môi trường, những ảnh hưởng đến

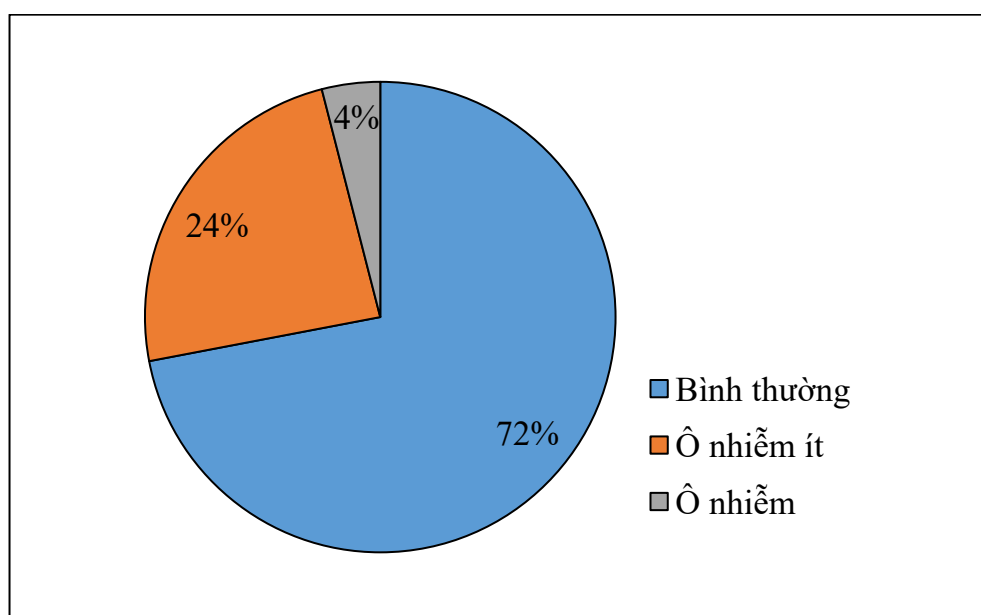
sức khỏe, sinh kế và những thay đổi môi trường trong những năm gần đây.

Tiêu chí lựa chọn đối tượng: Ưu tiên lựa chọn các hộ dân có thời gian cư trú lâu dài và có các hoạt động, công việc liên quan đến môi trường đất, nước tại khu vực Cửa An Hòa. Việc chọn mẫu nhằm đảm bảo sự đa dạng và khách quan trong thu thập ý kiến phản ánh từ cộng đồng dân cư.

Thời gian khảo sát: Tháng 4 năm 2025.

2.2.3. Kết quả khảo sát

❖ Môi trường không khí

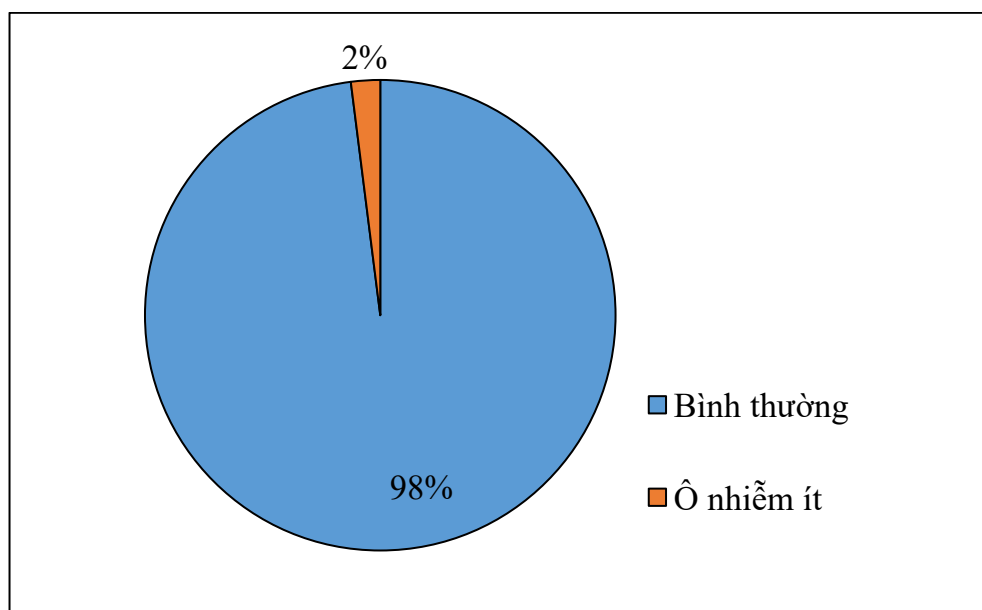


Hình 2.1 Chất lượng môi trường không khí tại cửa An Hòa

Theo kết quả khảo sát 72% người dân cho rằng chất lượng không khí bình thường, 24% đánh giá là ô nhiễm ít, 4% cho rằng không khí ô nhiễm

Nhận xét: Từ kết quả trên có thể thấy rằng đa số người dân nhận định chất lượng không khí ở mức chấp nhận được, tuy nhiên vẫn có một tỷ lệ nhất định phản ánh tình trạng ô nhiễm, trong đó 4% cho rằng không khí bị ô nhiễm rõ rệt. Các nguồn gây ô nhiễm chủ yếu từ hoạt động giao thông, khí thải, nước thải từ các nhà máy, khu công nghiệp xung quanh và từ hoạt động xây dựng cảng. Điều này cho thấy cần có biện pháp kiểm soát các nguồn phát thải nhằm đảm bảo chất lượng không khí ổn định và an toàn cho sức khỏe cộng đồng.

❖ **Môi trường đất**

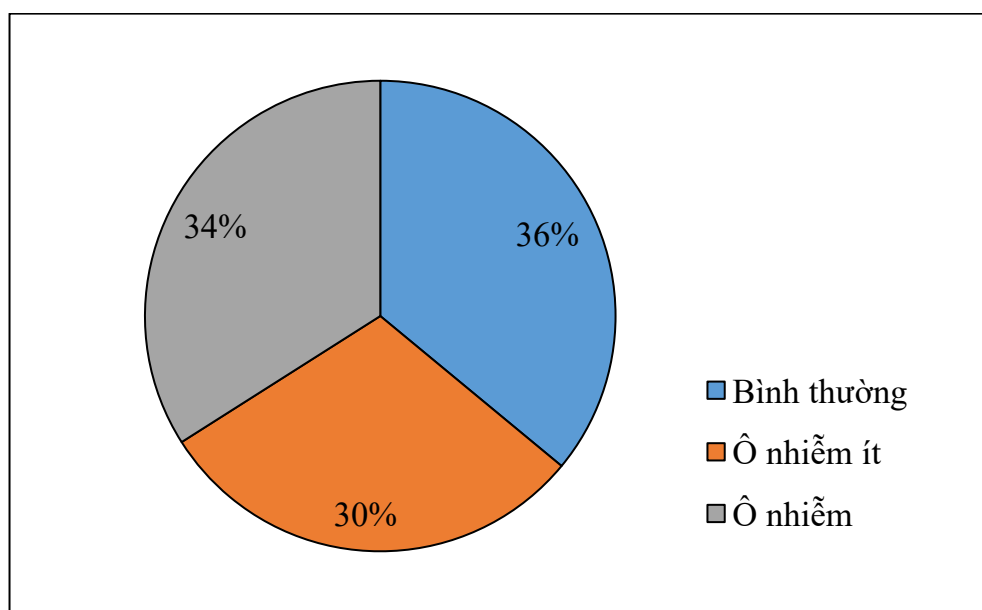


Hình 2.2 Chất lượng môi trường đất tại cửa An Hòa

Kết quả khảo sát cho thấy 98% người dân đánh giá chất lượng đất là bình thường, chỉ 2% cho rằng có hiện tượng ô nhiễm nhẹ, chủ yếu do hoạt động nuôi trồng thủy sản ven sông và việc xả thải sinh hoạt, chất thải rắn không đúng quy định từ các hộ dân.

Nhận xét: Phần lớn người dân cảm nhận rằng đất tại khu vực vẫn đảm bảo chất lượng tốt. Điều này cho thấy vấn đề ô nhiễm đất chưa phải là mối quan tâm chính tại thời điểm khảo sát.

❖ **Môi trường nước**

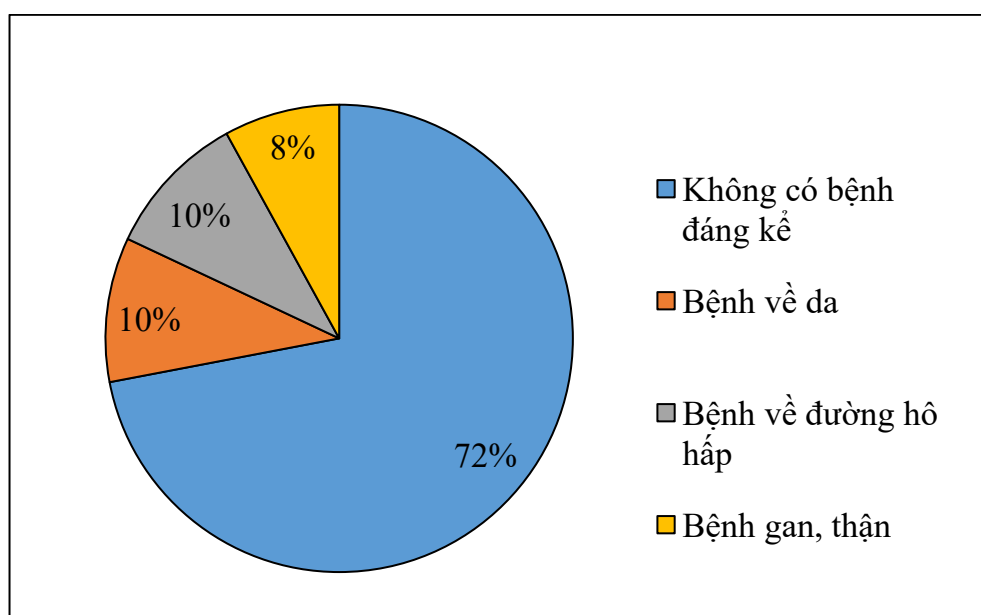


Hình 2.3 Chất lượng môi trường nước và trầm tích tại cửa An Hòa

Theo kết quả khảo sát, có 36% người dân cho rằng chất lượng nước và trầm tích ở mức bình thường, 30% đánh giá là ô nhiễm ít, và 34% nhận định khu vực đang bị ô nhiễm.

Nhận xét: Kết quả cho thấy phần lớn người dân nhận thấy môi trường nước và trầm tích có dấu hiệu ô nhiễm, với tổng cộng 64% đánh giá là ô nhiễm hoặc ô nhiễm ít. Điều này phản ánh sự lo ngại của cộng đồng về chất lượng môi trường thủy sinh tại khu vực, có thể ảnh hưởng đến hoạt động đánh bắt và nuôi trồng thủy sản. Theo phản ánh của người dân, các nguồn gây ô nhiễm chủ yếu bao gồm: hoạt động nuôi trồng thủy hải sản với mật độ cao, chất thải sinh hoạt chưa qua xử lý từ các khu dân cư ven sông, cũng như tác động từ các nhà máy, cơ sở sản xuất trong khu vực. Những yếu tố này góp phần làm suy giảm chất lượng nước và trầm tích, ảnh hưởng đến hệ sinh thái thủy vực và sức khỏe cộng đồng.

❖ Tỷ lệ mắc các bệnh



Hình 2.4 Tỷ lệ mắc các bệnh của người dân tại cửa An Hòa

Nhận xét: Kết quả khảo sát cho thấy 72% người dân không mắc bệnh đáng kể, trong khi 28% còn lại có các biểu hiện về bệnh da liễu (10%), hô hấp (10%), và gan, thận (8%). Điều này cho thấy một bộ phận dân cư có thể đang bị ảnh hưởng bởi chất lượng môi trường, đặc biệt là các bệnh lý liên quan đến ô nhiễm không khí, nước và điều kiện vệ sinh. Đây là cơ sở quan trọng để đánh giá rủi ro sinh thái và tác động sức khỏe cộng đồng trong nghiên cứu.

2.2.4. Tổng hợp và phân tích tình hình mắc bệnh tại khu vực nghiên cứu

Số liệu về số ca mắc các bệnh da liễu và hô hấp tại khu vực Núi Thành trong giai

đoạn từ tháng 1/2024 đến tháng 4/2025 được thu thập từ Bệnh viện Đa khoa Quảng Nam.

Bảng 2.2 Số liệu người mắc các bệnh về da và đường hô hấp tại Núi Thành

	Bệnh về đường hô hấp	Bệnh ngoài da
Điều trị nội trú	366	52
Khám bệnh	1231	189

Nhận xét: Qua so sánh giữa kết quả khảo sát người dân và số liệu từ bệnh viện tại khu vực Núi Thành, có thể thấy sự chênh lệch rõ rệt về tỷ lệ mắc các bệnh ngoài da và bệnh đường hô hấp. Cụ thể, khảo sát ghi nhận khoảng 10% dân số (tương đương khoảng 15.034 người) cho biết mắc bệnh ngoài da và 10% tương tự mắc bệnh đường hô hấp. Trong khi đó, số liệu từ bệnh viện chỉ ghi nhận 241 lượt khám bệnh ngoài da và 1.597 lượt khám bệnh về đường hô hấp trong thời gian từ đầu năm 2024 đến nay.

Sự chênh lệch này cho thấy có thể nhiều người dân không đến khám tại cơ sở y tế công lập mà lựa chọn tự điều trị tại nhà, điều trị bằng thuốc không kê đơn, hoặc đến khám tại các phòng khám tư nhân và bệnh viện khác trong tỉnh. Trong thực tế, trên địa bàn tỉnh có nhiều bệnh viện tuyến tỉnh và phòng khám tư nhân, đặc biệt là ở khu vực thành phố, có thể thu hút người dân đến khám thay vì tại địa phương. Ngoài ra, một số trường hợp có thể không đi khám do đánh giá mức độ bệnh không nghiêm trọng.



Hình 2.5 Thu thập số liệu tại bệnh viện đa khoa Quảng Nam

2.3. Phân tích các chỉ tiêu trong nước và trầm tích

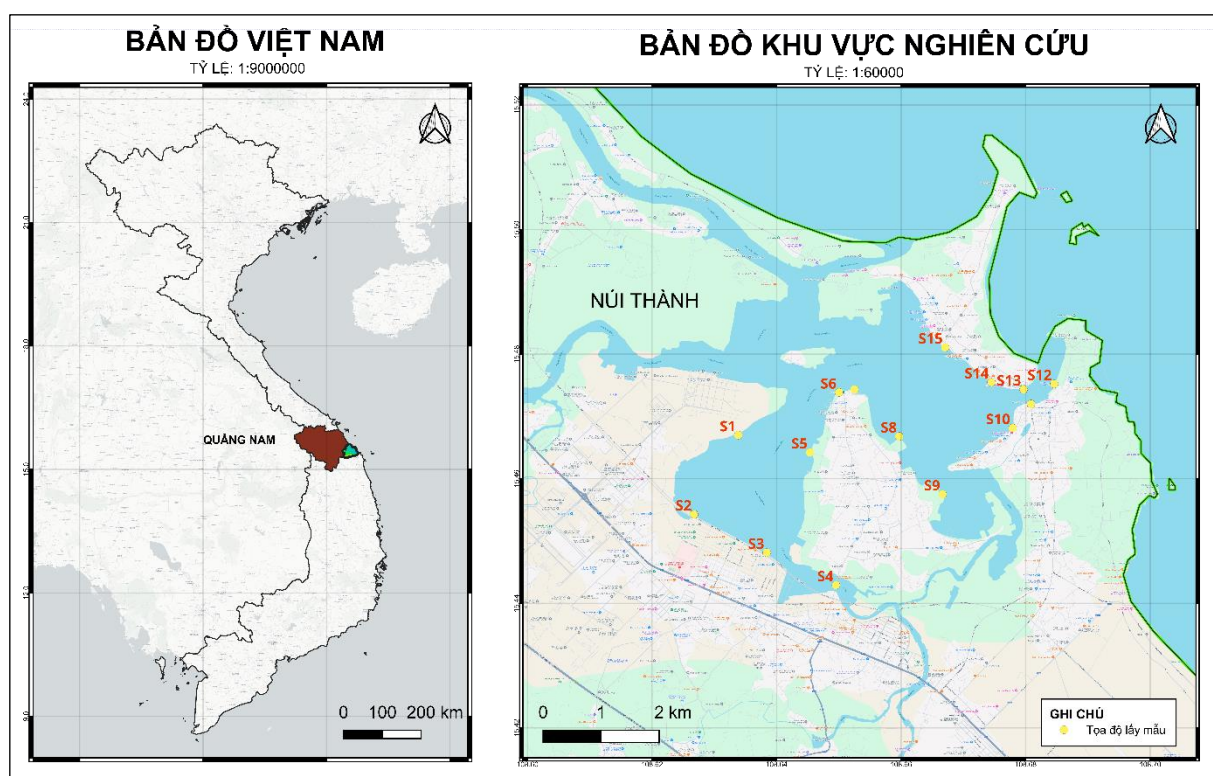
2.3.1. Mục đích

Nhằm có cái nhìn tổng quan ban đầu về hiện trạng chất lượng nước và trầm tích tại khu vực nghiên cứu, sinh viên đã tiến hành phân tích một số chỉ tiêu cơ bản trong nước tại các điểm quan trắc. Việc phân tích này nhằm đánh giá sơ bộ hiện trạng môi trường nước.

2.3.2. Đối tượng

Thông số chất lượng nước/trầm tích mặt tại khu vực cửa An Hòa

2.3.3. Phạm vi nghiên cứu



Hình 2.6 Khu vực nghiên cứu và các vị trí lấy mẫu

Nghiên cứu được thực hiện tại khu vực cửa An Hòa trong thời gian 04/2025. Mẫu nước được thu thập tại 15 vị trí đại diện cho chất lượng nước mặt trong khu vực khảo sát.

Điều kiện thời tiết: trời nắng, gió nhẹ.



Hình 2.7 Một số hình ảnh lấy mẫu tại các vị trí

2.3.4. Nội dung

Tiến hành lấy mẫu nước và trầm tích mặt tại các vị trí

Phân tích các thông số chất lượng nước/trầm tích mặt

Mẫu nước

- Phân tích các chỉ tiêu: Nhiệt độ, pH, TSS, BOD₅, COD, N-NH₄⁺, N-NO₂⁻, Cl⁻.
- Quy chuẩn so sánh: QCVN 08:2023/BTNMT - quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt.

Mẫu trầm tích

- Phân tích các thông số: Các kim loại nặng (As, Cd, Pb, Zn, Hg)

- Quy chuẩn so sánh : QCVN 43 :2017/BTNMT-quy chuẩn kỹ thuật quốc gia về trầm tích

2.3.5. Phương pháp và thiết bị quan trắc

- TCVN 6663-1:2011 (ISO 5667-2:2006): Chất lượng nước – Lấy mẫu – Phần 1: Hướng dẫn kỹ thuật lấy mẫu.

- TCVN 6663-6:2018: Chất lượng nước - Lấy mẫu - phần 6: Hướng dẫn lấy mẫu nước sông và suối.

- TCVN 6663-3:2003 (ISO 5667-3:1985): Chất lượng nước – Lấy mẫu – Phần 3: Hướng dẫn bảo quản và xử lý mẫu.

- TCVN 6663-13 : 2000 (ISO 5667-13 : 1997) Phần 13: hướng dẫn lấy mẫu bùn nước, bùn nước thải và bùn liên quan.

- TCVN 6663-15 : 2004 (ISO 5667-15 : 1999) Phần 15: Hướng dẫn bảo quản và xử lý mẫu bùn và trầm tích.

Bảng 2.3 Các chỉ tiêu và phương pháp phân tích

STT	Chỉ tiêu phân tích	Phương pháp	Dụng cụ và thiết bị
1	Nhiệt độ	Đo nhanh	-
2	pH	Đo nhanh	Bút đo Hanna HI 98107
3	Amoni (NH ₄ ⁺)	TCVN 6179-1:1996	Máy quang phổ LIUV 201 UV/Vis
4	Chloride (Cl ⁻)	TCVN 6194:1996	-
5	TSS	TCVN 6625:2000	Giấy lọc, tủ sấy, cân phân tích
6	Nitrit (NO ₂ ⁻)	TCVN 6178:1996	Máy quang phổ LIUV 201 UV/Vis
7	COD	TCVN 6491:1999	Bộ phân tích COD VARIO
8	BOD ₅	TCVN 6001-1:2008	Bộ phân tích BOD - FOC225E

Do không có điều kiện và trang thiết bị phù hợp để phân tích các chỉ tiêu kim loại nặng, các mẫu trầm tích được thu thập tại các điểm khảo sát đã được gửi đến Trung tâm Quan trắc và Phân tích môi trường Quảng Nam – một đơn vị có chuyên môn và đủ năng lực phân tích theo các phương pháp chuẩn. Kết quả phân tích được sử dụng làm cơ sở để đánh giá mức độ ô nhiễm kim loại nặng trong khu vực nghiên cứu.

Ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam



Hình 2.8 Phân tích các chỉ tiêu trong phòng thí nghiệm

❖ Dưới đây là kết quả phân tích một số chỉ tiêu chất lượng nước tại các điểm quan trắc thuộc khu vực nghiên cứu.

Bảng 2.4 Kết quả phân tích các chỉ tiêu trong nước

Chỉ tiêu Vị trí	Nhiệt độ	pH	Amoni (NH₄⁺)	Chloride (Cl⁻)	TSS	Nitrit (NO₂⁻)	BOD₅	COD
Đơn vị	°C	-	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L
S1	25,6	6,5	1	9,6	2,6	0,27	8,0	13,6
S2	25,4	6	1,1	9,6	1,8	0,32	8,0	8,2
S3	25,4	6,5	0,8	9,6	4,6	0,14	7,0	7,7
S4	25,3	5,5	1,3	11,0	7,5	0,27	7,2	9,7
S5	25,6	7	0,7	11,0	5,1	0,10	7,0	8,7
S6	25,1	7	1,8	12,4	11,4	0,11	7,4	9,5
S7	26,2	6	1,2	12,4	6,1	0,55	5,8	6,9
S8	26,1	7	0,6	11,0	4,2	0,19	5,0	9,0
S9	26,0	7	1,2	11,0	9,0	0,16	5,8	7,7
S10	26,5	6,5	1,5	12,7	6,4	0,3	8,8	9,7
S11	26,4	7	2,5	13,2	7,2	0,15	11,2	15,1
S12	26,6	5,5	0,7	13,8	3,1	0,12	5,6	6,7
S13	26,7	7	2,4	13,8	5,0	0,04	10,6	14,1
S14	26,9	7	0,6	13,2	7,2	0,11	5,6	6,9

Ứng dụng các mô hình học máy để dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng trong trầm tích mặt tại cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam

Chỉ tiêu		Nhiệt độ	pH	Amoni (NH ₄ ⁺)	Chloride (Cl ⁻)	TSS	Nitrit (NO ₂ ⁻)	BOD ₅	COD
Vị trí	S15	26,7	6,5	1,1	12,9	6,6	0,31	6,4	6,8
QCVN 08:2023/BTNMT	Mức A	-	6,5 – 8,5	-	-	≤ 25 mg/L	-	≤ 10 mg/L	≤ 10 mg/L
	Mức B	-	6,5 – 8,5	-	-	≤ 100 mg/L	-	≤ 15 mg/L	≤ 15 mg/L
	Mức C	-	6,5 – 8,5	-	-	> 100 mg/L và không có rác nổi	-	≤ 20 mg/L	≤ 20 mg/L
	Mức D	-	< 6,0 hoặc > 8,5	-	-	> 100 mg/L và có rác nổi	-	> 10 mg/L	> 20 mg/L
AHSK	-	-	0,3	250	-	0,05	-	-	

Ghi chú: QCVN 08:2023/BTNMT quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt.

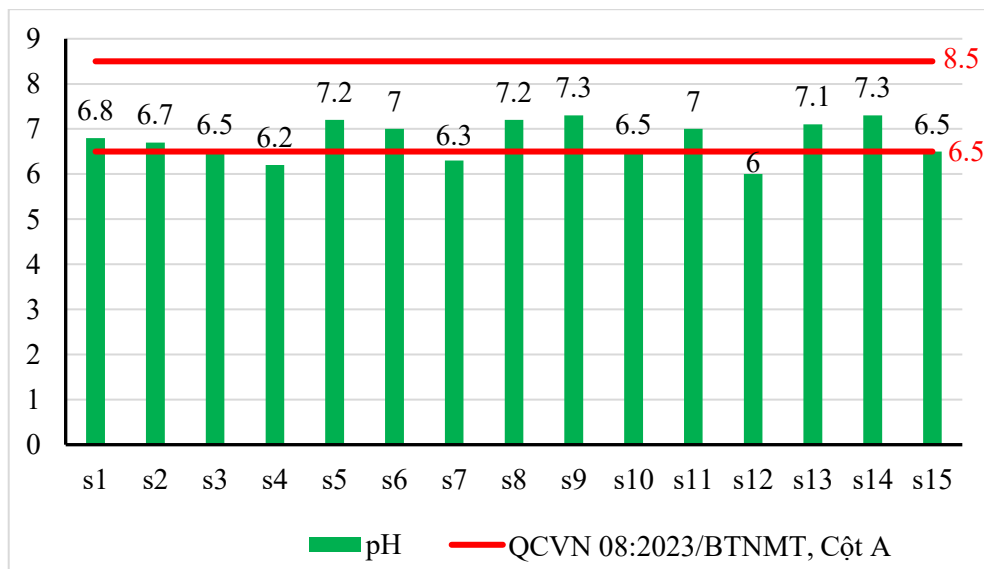
AHSK: Giá trị giới hạn tối đa các thông số ảnh hưởng tới sức khỏe con người.

Mức A: Chất lượng nước tốt. Hệ sinh thái trong môi trường nước có hàm lượng oxy hòa tan (DO) cao. Nước có thể được sử dụng cho mục đích cấp nước sinh hoạt, bơi lội, vui chơi dưới nước sau khi áp dụng các biện pháp xử lý phù hợp.

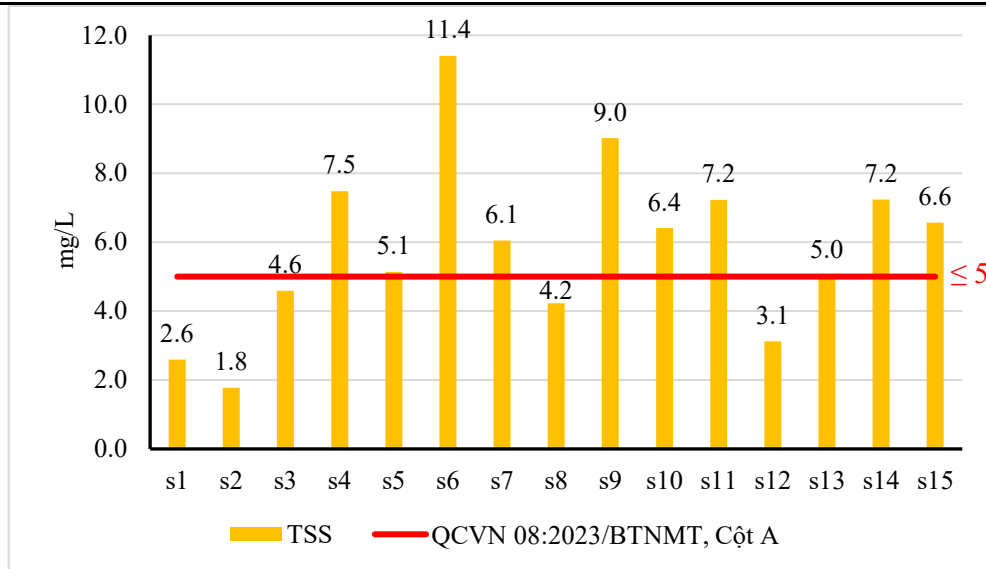
Mức B: Chất lượng nước trung bình. Hệ sinh thái trong nước tiêu thụ nhiều oxy hòa tan do một lượng lớn chất ô nhiễm. Nước có thể sử dụng cho mục đích sản xuất công nghiệp, nông nghiệp sau khi áp dụng các biện pháp xử lý phù hợp.

Mức C: Chất lượng nước xấu. Hệ sinh thái trong nước có lượng oxy hòa tan giảm mạnh do chứa một lượng lớn các chất ô nhiễm. Nước không gây mùi khó chịu, có thể được sử dụng cho các mục đích sản xuất công nghiệp sau khi áp dụng các biện pháp xử lý phù hợp.

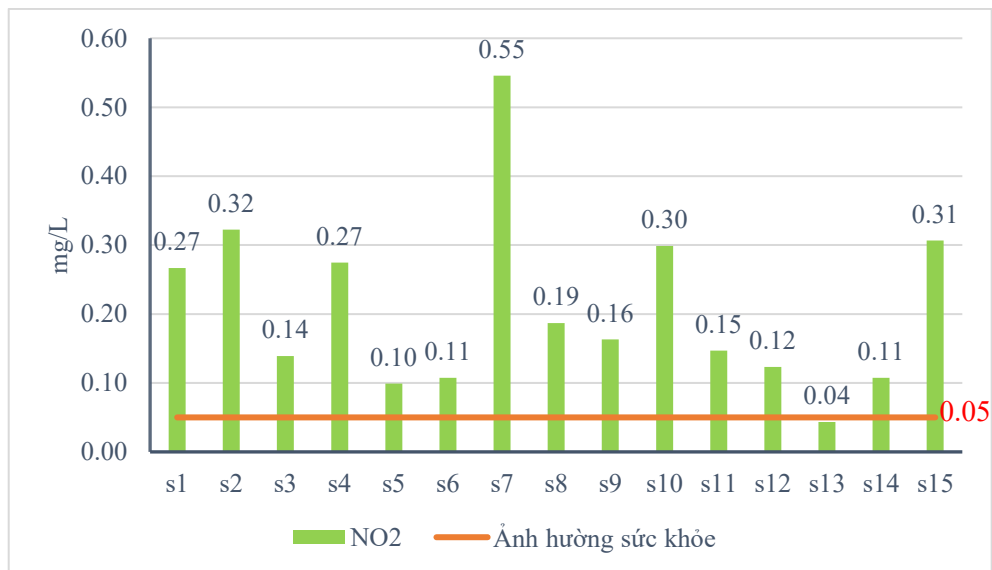
Mức D: Nước có chất lượng rất xấu, có thể gây ảnh hưởng lớn tới cá và các sinh vật sống trong môi trường nước do nồng độ oxy hòa tan thấp, nồng độ chất ô nhiễm cao. Nước có thể được sử dụng cho các mục đích giao thông thủy và các mục đích khác với yêu cầu nước chất lượng thấp.



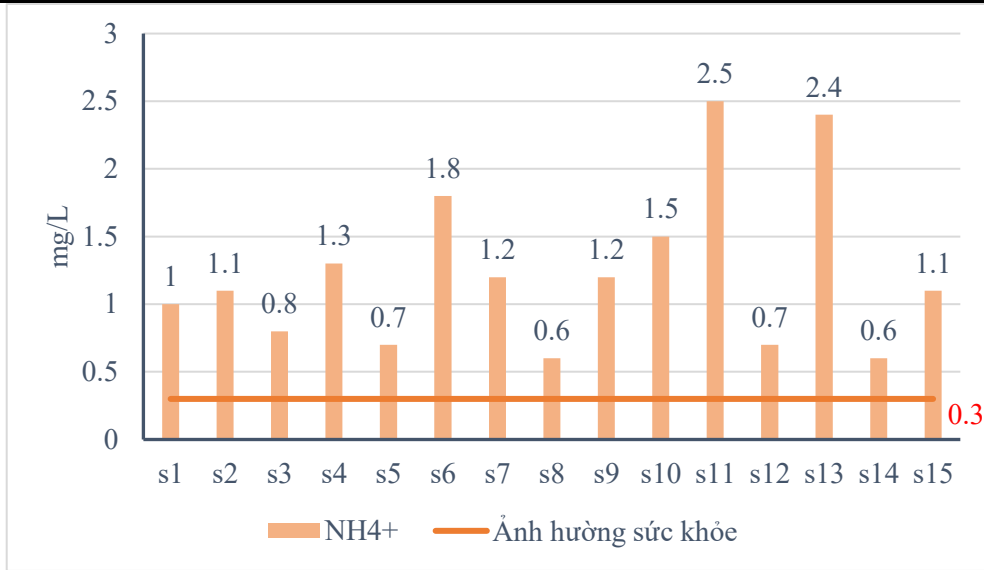
Hình 2.9 Kết quả phân tích chỉ tiêu pH trong nước



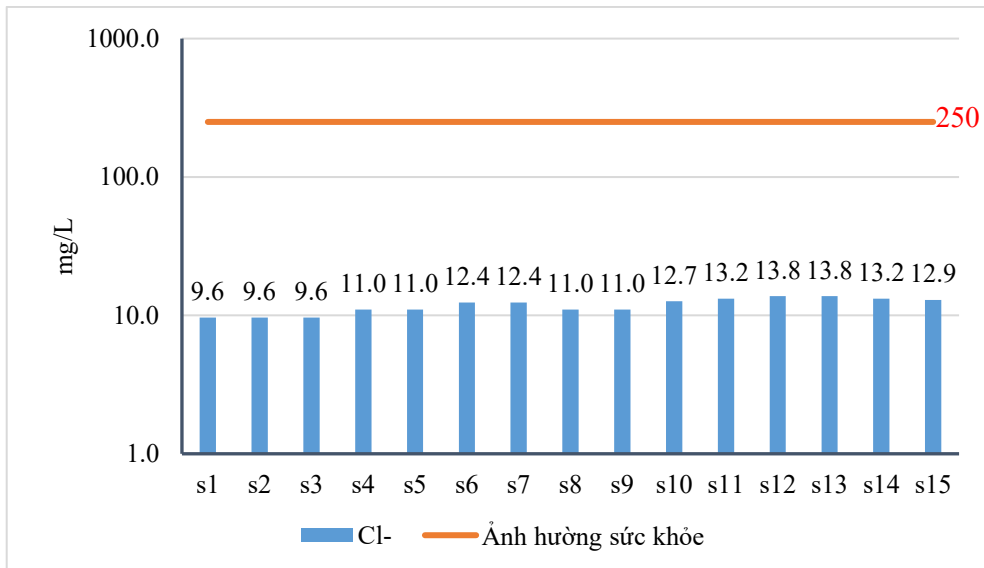
Hình 2.10 Kết quả phân tích chỉ tiêu TSS trong nước



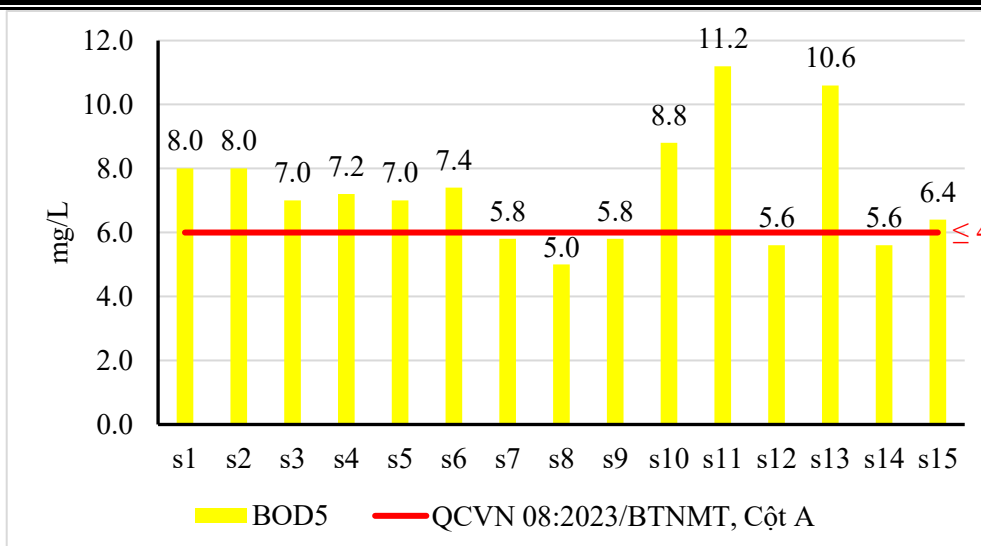
Hình 2.11 Kết quả phân tích chỉ tiêu NO₂ trong nước



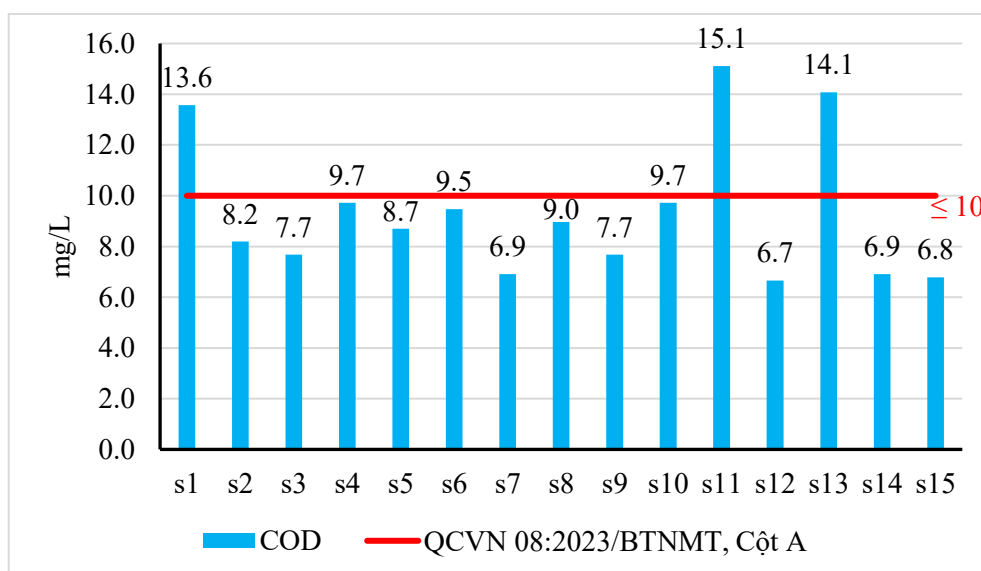
Hình 2.12 Kết quả phân tích chỉ tiêu NH₄⁺ trong nước



Hình 2.13 Kết quả phân tích chỉ tiêu Cl⁻ trong nước



Hình 2.14 Kết quả phân tích chỉ tiêu BOD₅ trong nước



Hình 2.15 Kết quả phân tích chỉ tiêu COD trong nước

Nhận xét: Kết quả phân tích giá trị pH tại tất cả 15 vị trí dao động từ 6,2 đến 7,3. Mặc dù phần lớn các vị trí nằm trong khoảng giới hạn cho phép (6,5–8,5 theo QCVN 08:2023/BTNMT, Cột A), một số vị trí như S1, S2, S3 và S4 có pH thấp hơn giới hạn dưới, cho thấy nước có xu hướng hơi axit nhẹ ở những khu vực này. Nồng độ tổng chất rắn lơ lửng (TSS) biến động lớn, từ 1,8 mg/L đến 11,4 mg/L, với phần lớn các điểm lấy mẫu (trừ S1, S2, S7 và S12) đều vượt ngưỡng cho phép là 5 mg/L. Đặc biệt, vị trí S6 có giá trị cao nhất lên tới 11,4 mg/L, cho thấy khả năng nước đang chịu ảnh hưởng từ dòng chảy mặt hoặc hoạt động gây xáo trộn đất đá.

Về hàm lượng các hợp chất nitơ, cả Nitrit (NO_2^-) và Amoni (NH_4^+) đều vượt xa giới hạn an toàn cho sức khỏe (lần lượt là 0,05 mg/L và 0,3 mg/L) tại tất cả các điểm

lấy mẫu. Vị trí S7 có giá trị Nitrit cao nhất là 0,55 mg/L, trong khi S11 có Amoni cao nhất lên tới 2,5 mg/L. Những kết quả này cho thấy dấu hiệu ô nhiễm hữu cơ và phân hủy hợp chất nitơ, có thể liên quan đến nguồn thải từ hoạt động nuôi trồng thủy sản và công nghiệp. Ngược lại, nồng độ Clorua (Cl^-) tại các điểm quan trắc dao động từ 9,6 mg/L đến 13,8 mg/L, thấp hơn nhiều so với ngưỡng giới hạn là 250 mg/L, chứng tỏ nước không bị ô nhiễm bởi muối hoặc nước biển.

Các chỉ tiêu oxy hóa cũng cho thấy sự ô nhiễm đáng kể. Giá trị COD dao động từ 6,7 mg/L đến 15,1 mg/L, với các vị trí S1, S4, S5, S10, S11 và S13 vượt mức cho phép là 10 mg/L (Cột A). Vị trí S11 tiếp tục là điểm có mức ô nhiễm cao nhất. Tương tự, BOD₅ tại tất cả các điểm lấy mẫu đều vượt quá giới hạn 4 mg/L (Cột A), dao động từ 5,0 mg/L đến 11,2 mg/L, trong đó S11 một lần nữa có giá trị cao nhất. Điều này phản ánh rõ ràng sự hiện diện của các hợp chất hữu cơ dễ phân hủy sinh học trong nguồn nước.

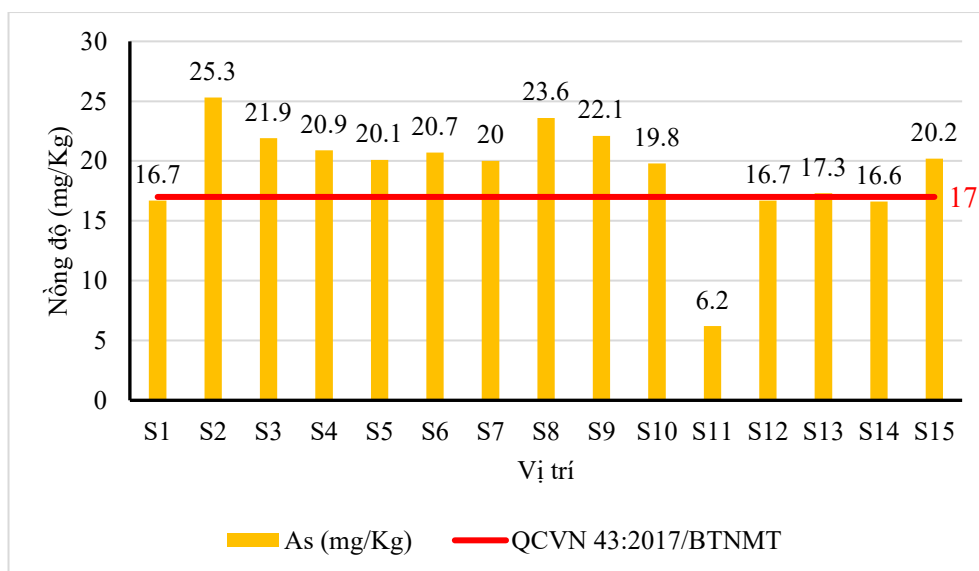
❖ Dưới đây là kết quả phân tích một số chỉ tiêu chất lượng nước tại các điểm quan trắc thuộc khu vực nghiên cứu.

Bảng 2.5 Kết quả phân tích các chỉ tiêu trong trầm tích

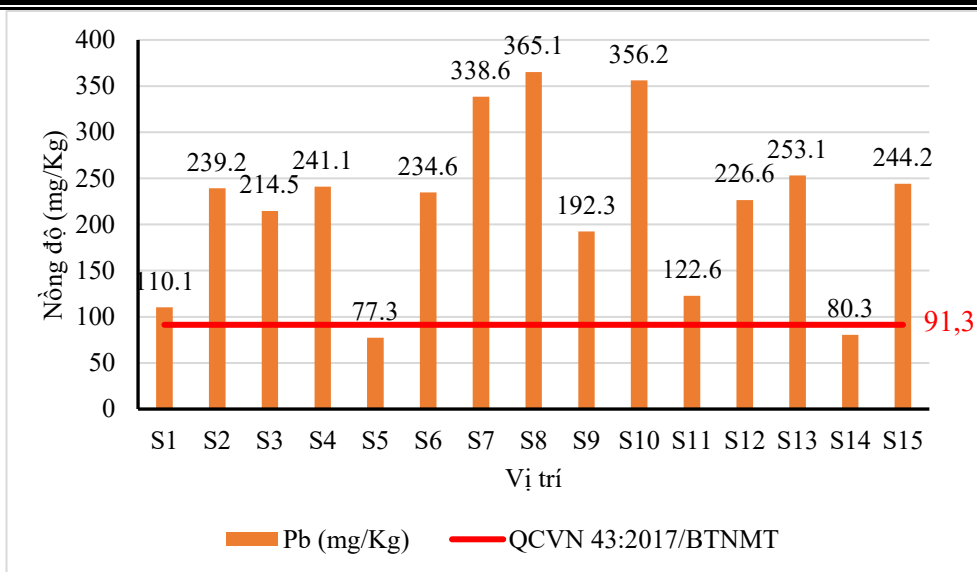
Chỉ tiêu Vị trí	As	Pb	Hg	Cd	Zn
	Đơn vị	mg/Kg	mg/Kg	mg/Kg	mg/Kg
S1	16,7	110,1	0,34	1,10	61,5
S2	25,3	239,2	0,30	1,60	88,3
S3	21,9	214,5	0,29	1,06	93,1
S4	20,9	241,1	0,29	1,02	83,4
S5	20,1	77,3	0,23	1,00	74,1
S6	20,7	234,6	0,26	1,31	82,9
S7	20	338,6	0,28	1,43	96,3
S8	23,6	365,1	0,28	1,30	92,3
S9	22,1	192,3	0,25	1,07	83,9
S10	19,8	356,2	0,15	1,16	95,7

Chỉ tiêu Vị trí	As	Pb	Hg	Cd	Zn
S11	6,2	122,6	0,18	0,09	82,1
S12	16,7	226,6	0,20	0,13	72,8
S13	17,3	253,1	0,20	1,86	81,6
S14	16,6	80,3	0,16	2,59	95
S15	20,2	244,2	0,19	2,86	91
QCVN 43:2017/BTNMT	17	91.3	0,5	3,5	315

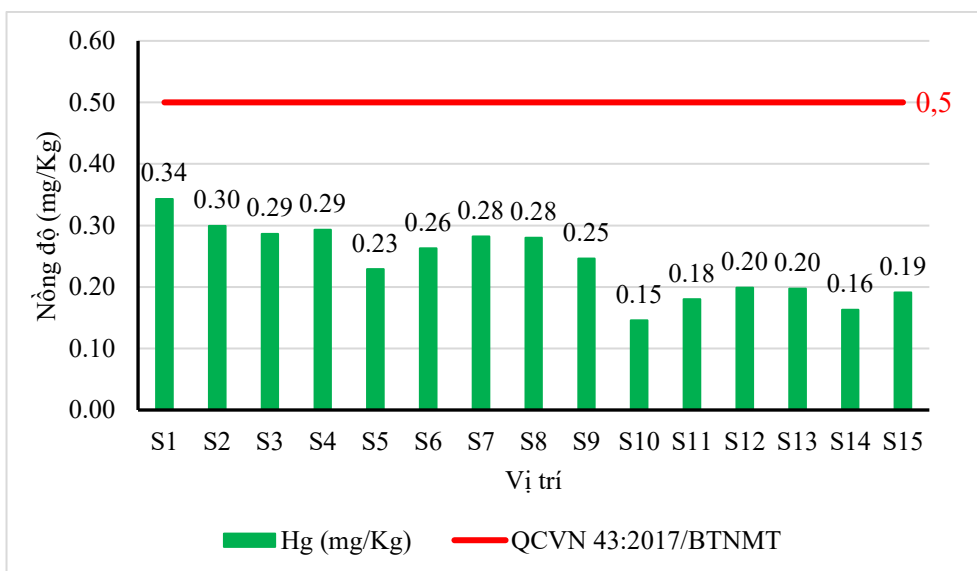
❖ **Ghi chú:** QCVN 43:2017/BTNMT quy chuẩn kỹ thuật quốc gia về chất lượng trầm tích.



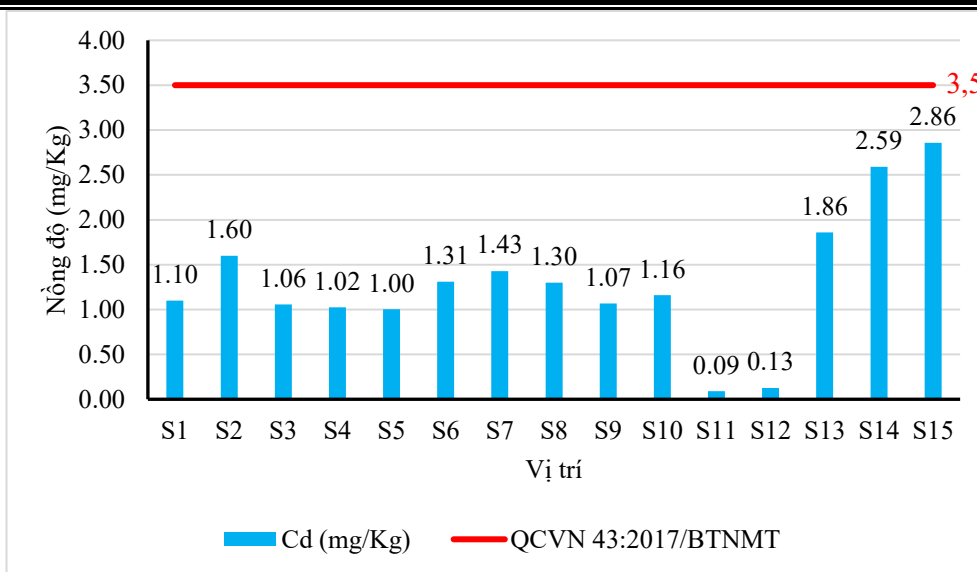
Hình 2.16 Kết quả phân tích chỉ tiêu As trong trầm tích



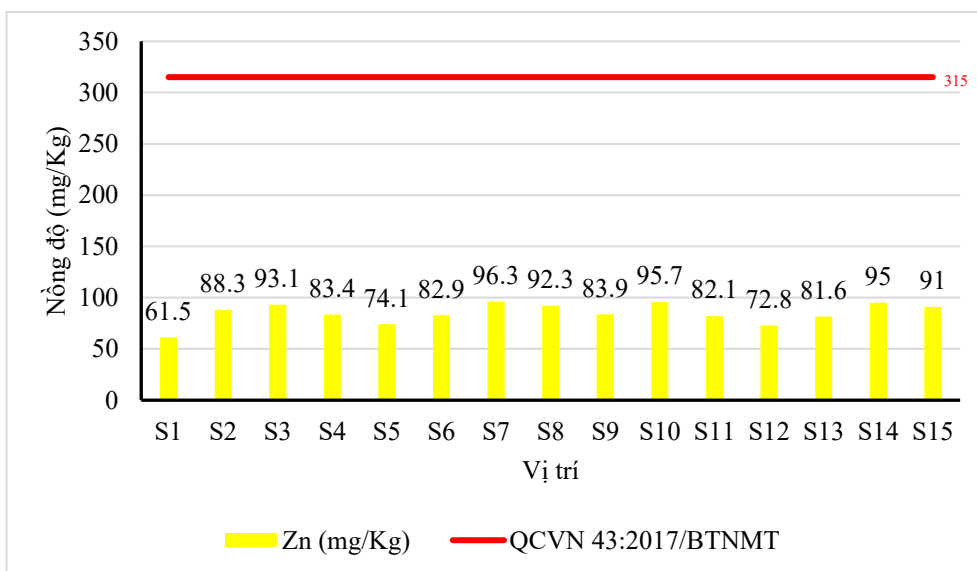
Hình 2.17 Kết quả phân tích chỉ tiêu Pb trong trầm tích



Hình 2.18 Kết quả phân tích chỉ tiêu Hg trong trầm tích



Hình 2.19 Kết quả phân tích chỉ tiêu Cd trong trầm tích



Hình 2.20 Kết quả phân tích chỉ tiêu Zn trong trầm tích

Nhận xét: Kết quả phân tích hàm lượng kim loại nặng tại 15 vị trí khảo sát (S1–S15) cho thấy mức độ biến động đáng kể giữa các điểm, đồng thời phản ánh tình trạng ô nhiễm kim loại nặng tại một số khu vực. Hàm lượng Asen (As) dao động từ 6,2 mg/kg đến 25,3 mg/kg, trong đó có tới 11/15 vị trí vượt giới hạn cho phép theo QCVN 43:2017/BTNMT (17 mg/Kg). Riêng vị trí S11 có giá trị thấp hơn ngưỡng, cho thấy mức độ nhiễm As tại vị trí này là không đáng kể. Hàm lượng Chì (Pb) tại tất cả các điểm đều vượt chuẩn, với giá trị dao động từ 77,3 mg/kg (S5) đến 365,1 mg/kg (S8), vượt từ 1 đến gần 4 lần so với ngưỡng cho phép (91,3 mg/kg), đặc biệt cao tại các vị trí S7, S8 và S10, cho thấy khả năng có nguồn thải Pb trong khu vực.

Đối với Thủy ngân (Hg), giá trị tại các điểm dao động từ 0,15 đến 0,34 mg/kg, đều nằm dưới giới hạn cho phép là 0,5 mg/kg, cho thấy mức độ ô nhiễm Hg tại khu vực là không đáng kể. Hàm lượng Cadimi (Cd) có giá trị trong khoảng 0,09–2,86 mg/kg, đều nằm trong ngưỡng giới hạn cho phép (3,5 mg/kg), tuy nhiên các vị trí S14 và S15 có giá trị tiệm cận mức giới hạn, cần được theo dõi bổ sung. Hàm lượng Kẽm (Zn) tại các vị trí dao động trong khoảng 61,5–96,3 mg/kg, thấp hơn nhiều so với giới hạn 315 mg/kg, cho thấy Zn chưa phải là tác nhân gây ô nhiễm đáng kể tại khu vực khảo sát.

Nhìn chung, trong số các kim loại nặng được phân tích, As và Pb là hai yếu tố có mức độ ô nhiễm vượt ngưỡng phổ biến và rõ rệt nhất, đặc biệt tại các vị trí S2, S7, S8 và S10. Điều này cho thấy cần có biện pháp kiểm soát và đánh giá nguồn phát thải, đồng thời tăng cường giám sát môi trường đất tại các khu vực có nồng độ cao để đảm bảo an toàn sinh thái và sức khỏe cộng đồng.

CHƯƠNG 3. ỨNG DỤNG MÔ HÌNH HỌC MÁY ĐỂ DỰ BÁO MỨC ĐỘ TÍCH LŨY KIM LOẠI NẶNG TRONG TRẦM TÍCH VÀ PHÂN TÍCH ĐỘ TIN CẬY CỦA MÔ HÌNH

3.1. Cơ sở dữ liệu đầu vào

3.1.1. Nguồn dữ liệu

Nguồn dữ liệu sử dụng trong nghiên cứu bao gồm cả dữ liệu do sinh viên thu thập trực tiếp và dữ liệu được cung cấp từ đơn vị phân tích chuyên môn.

Cụ thể, sinh viên nghiên cứu đã tiến hành lấy 15 mẫu trầm tích mặt tại các vị trí đại diện dọc theo khu vực cửa An Hòa – sông Trường Giang, tỉnh Quảng Nam. Các mẫu này được gửi đến Trung tâm Quan trắc và Phân tích môi trường Quảng Nam (84 Phan Bội Châu, phường Tân Thạnh, thành phố Tam Kỳ, tỉnh Quảng Nam) để phân tích các chỉ tiêu kim loại nặng gồm: Asen (As), Chì (Pb), Thủy ngân (Hg), Cadimi (Cd), và Kẽm (Zn). Ngoài việc phân tích 15 mẫu do sinh viên cung cấp, Trung tâm Quan trắc và Phân tích môi trường Quảng Nam còn hỗ trợ và cung cấp thêm 210 kết quả phân tích trầm tích mặt được thực hiện trước đó tại khu vực sông Trường Giang, góp phần làm phong phú và đa dạng hơn bộ dữ liệu đầu vào.

Tổng cộng, bộ dữ liệu đầu vào gồm 225 kết quả phân tích trầm tích mặt bao gồm các thông tin định lượng về các kim loại nặng (As, Pb, Hg, Cd, Zn) và các đặc tính môi trường liên quan, đảm bảo độ tin cậy cao, phục vụ cho quá trình xây dựng và hiệu chỉnh mô hình học máy một cách chính xác và khách quan.

3.1.2. Tiền xử lý dữ liệu

Trước khi đưa vào mô hình học máy, bộ dữ liệu được tiến hành các bước tiền xử lý nhằm đảm bảo chất lượng, tính chính xác và khả năng khái quát của mô hình:

Trong bước xử lý dữ liệu, việc phát hiện và loại bỏ những giá trị bất thường là rất cần thiết để mô hình hoạt động chính xác hơn. Đầu tiên, sinh viên sử dụng phương pháp IQR – tức là xem xét khoảng giữa của dữ liệu – để phát hiện các giá trị nằm quá xa so với phần lớn các giá trị còn lại, từ đó loại bỏ chúng. Sau đó, phương pháp z-score cũng được dùng để kiểm tra xem giá trị nào lệch quá nhiều so với mức trung bình. Những giá trị quá cao hoặc quá thấp sẽ được xem xét và xử lý lại cho phù hợp, nhằm đảm bảo dữ liệu đầu vào được sạch và ổn định hơn trước khi đưa vào mô hình.

3.1.3. Phân chia dữ liệu

Để huấn luyện và đánh giá mô hình học máy một cách hiệu quả và khách quan, bộ dữ liệu gồm 225 mẫu được phân chia thành hai tập: tập huấn luyện (training set) và tập kiểm tra (testing set). Tỷ lệ phân chia được điều chỉnh linh hoạt trong khoảng từ 70:30 đến 85:15 nhằm tìm kiếm cấu hình tối ưu cho từng mô hình cụ thể.

Tập huấn luyện (70% – 85%): được sử dụng để đào tạo mô hình, giúp máy học nhận diện các mối quan hệ giữa các biến đầu vào và biến mục tiêu.

Tập kiểm tra (30% – 15%): dùng để kiểm tra khả năng dự báo và đánh giá hiệu suất của mô hình trên dữ liệu chưa từng được học.

Việc thử nghiệm với nhiều tỷ lệ phân chia giúp:

- + Đánh giá ảnh hưởng của lượng dữ liệu huấn luyện đến hiệu suất mô hình.
- + Giảm thiểu rủi ro quá khớp (overfitting) hoặc không khớp (underfitting) trong quá trình huấn luyện.

Để tăng tính khách quan và độ tin cậy của kết quả, kỹ thuật k-fold Cross-Validation được áp dụng trong mỗi tỷ lệ phân chia ($k = 10$). Phương pháp này cho phép mô hình được huấn luyện và kiểm tra trên nhiều tổ hợp dữ liệu khác nhau, từ đó đánh giá mức độ ổn định và khả năng tổng quát hóa của mô hình.

Ngoài ra, trong quá trình chia dữ liệu, các đặc trưng thống kê quan trọng (như giá trị trung bình, độ lệch chuẩn, phân bố giá trị) của các biến đầu vào và đầu ra cũng được đảm bảo giữ ổn định giữa hai tập dữ liệu. Điều này giúp mô hình học máy có thể dự báo chính xác hơn và phản ánh đúng xu hướng thực tế của dữ liệu môi trường.

Sau khi dữ liệu đã được xử lý và phân chia hợp lý, một trong những biến mục tiêu quan trọng được mô hình học máy dự báo là **chỉ số rủi ro sinh thái (RI)** – đại lượng phản ánh mức độ ô nhiễm và nguy cơ sinh thái tiềm ẩn do sự hiện diện của các kim loại nặng trong trầm tích. Chỉ số này được đề xuất bởi Hakanson (1980), và được tính theo công thức:

$$RI = \sum_{i=1}^n T_r^i \times \frac{\bar{C}_{0-1}^i}{C_n^i}$$

Trong đó:

- RI: Tổng chỉ số rủi ro sinh thái
- T_r^i : Hệ số phản ứng độc tính của chất thứ i ($(Hg = 40, Cd = 30, As = 10, Pb = 5, Zn = 1)$)
- \bar{C}_{0-1}^i : Nồng độ trung bình của chất thứ i trong lớp trầm tích bề mặt (0–1 cm)

- C_n^i : Mức nền tham chiếu (Hg = 0.25, Cd = 1.0, As = 15, Pb = 70, Zn = 175)

Bảng 3.1 Chỉ số rủi ro Risk Index (RI)

STT	Chỉ số RI	Mức độ rủi ro sinh thái tổng thể
1	$RI < 150$	Rủi ro thấp
2	$150 \leq RI < 300$	Rủi ro trung bình
3	$300 \leq RI < 600$	Rủi ro cao
4	$RI \geq 600$	Rủi ro rất cao

3.2. Lựa chọn và xây dựng mô hình học máy

3.2.1. Tải và cài đặt phần mềm

R là một ngôn ngữ lập trình mã nguồn mở chuyên dùng trong lĩnh vực thống kê, phân tích dữ liệu và học máy. Để hỗ trợ quá trình lập trình và phân tích một cách hiệu quả hơn, phần mềm RStudio được phát triển như một môi trường làm việc tích hợp (IDE – Integrated Development Environment) dành riêng cho R. RStudio cung cấp giao diện trực quan, hỗ trợ viết mã, theo dõi biến, hiển thị đồ thị, và xuất báo cáo một cách thuận tiện.

Trước khi cài đặt RStudio, người dùng cần cài đặt phần mềm R phiên bản mới nhất. Hiện tại, phiên bản đang sử dụng là R 4.3.3, được phát hành ngày 12/03/2024. Người dùng có thể tải phần mềm R trực tiếp từ trang chính thức của "Comprehensive R Archive Network (CRAN)" tại địa chỉ: <https://cran.r-project.org/>

Tại trang này, người dùng chọn hệ điều hành phù hợp (Windows, macOS hoặc Linux) và tải tệp cài đặt tương ứng, ví dụ:

Với Windows: R-4.3.3-win.exe

Với macOS: R-4.3.3.pkg

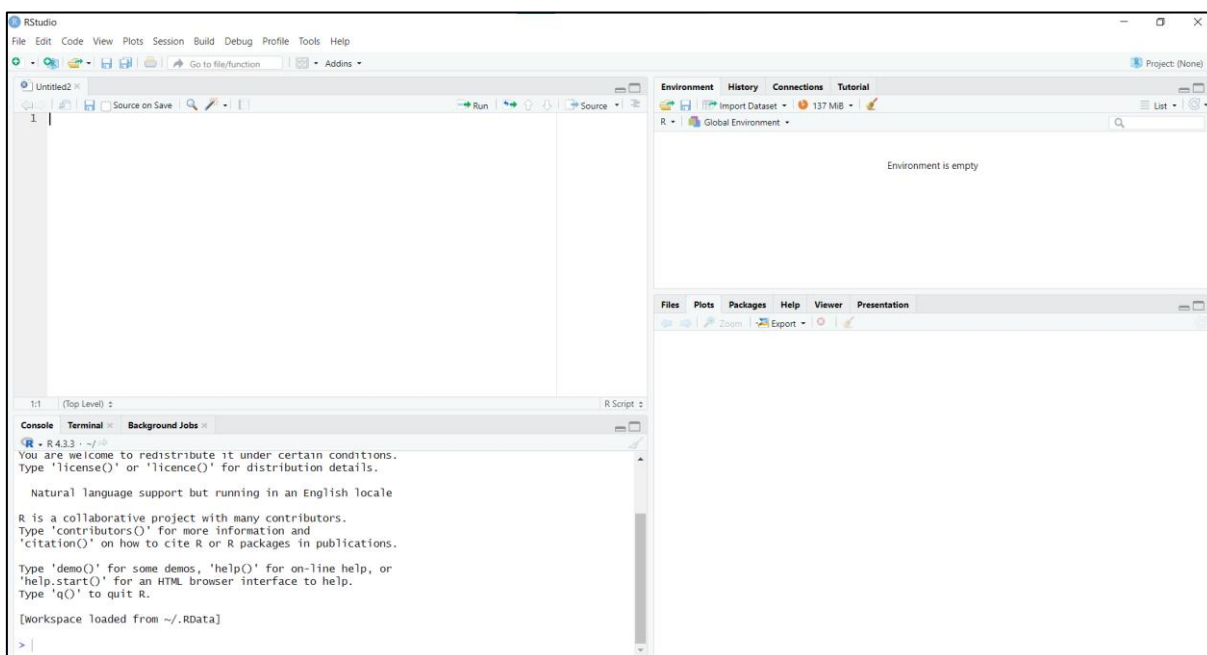
Sau khi cài đặt xong R, người dùng tiếp tục truy cập vào trang chính thức của RStudio để tải phần mềm RStudio Desktop phiên bản miễn phí tại địa chỉ: <https://posit.co/download/rstudio-desktop/>

RStudio có thể được cài đặt tương tự như các phần mềm thông thường. Sau khi cài đặt hoàn tất, người dùng có thể mở RStudio và bắt đầu viết mã R để phục vụ các mục tiêu học tập, nghiên cứu hoặc xây dựng mô hình học máy.



Hình 3.1 Biểu tượng R-Studio trên màn hình Desktop

Sau khi cài đặt hoàn tất, nhấn chuột vào biểu tượng ở trên Hình 3.1, chúng ta sẽ có một cửa sổ làm việc trên R-Studio như sau:



Hình 3.2 Màn hình làm việc của R-Studio

Màn hình cửa sổ làm việc của R-Studio gồm có 4 khung:

Bên trái phía trên là khung “nguồn”, khung này dùng để soạn các câu lệnh để xử lý dữ liệu và thực hiện mô hình;

Bên trái phía dưới là khung “tương tác”, khung này thể hiện các câu lệnh và kết quả thực hiện câu lệnh. Khung này giống với cửa sổ làm việc của R khi không sử dụng R-Studio.

Bên phải phía trên là khung "đối tượng" chứa ba thẻ, thẻ Environment dùng để thể hiện thông tin về các đối tượng sử dụng trong khi thực hiện các câu lệnh, thẻ History dùng để thể hiện tất cả các câu lệnh đã soạn kể từ một thời điểm nào đấy, thẻ Connection thể hiện các kết nối. Bên phải phía dưới là khung "đa chức năng" chứa 5 thẻ:

- + Files: chứa các thư mục và tập tin có liên quan;
- + Plots: thể hiện các biểu đồ là kết quả của những lệnh vẽ;

+ Packages: thể hiện các phụ kiện;

+ Help: thể hiện các trợ giúp;

+ Viewer: dùng để xem.

Bây giờ ta có thể bắt đầu thực hiện các mô hình học máy trên R-Studio

3.2.2. Các mô hình được sử dụng

a. RF- Random Forest (Rừng ngẫu nhiên) [10]

RF là 1 sự kết hợp của cây dự đoán, mỗi cây phụ thuộc vào giá trị của vector ngẫu nhiên được lấy mẫu độc lập (Independently) và với sự phân bố như nhau cho tất cả các cây có trong rừng. RF là một loại của thuật toán tổng hợp (Ensemble) được gọi là tổng hợp (aggregation) bootstrap và là một trong những phương pháp học máy phổ biến nhất. RF thực hiện (Perform) một dự đoán dựa vào biến đầu vào (x) xảy ra ở bước tiếp theo:

Lấy mẫu bootstrap và chọn ngẫu nhiên 1 tập hợp con (Subset) của tập dữ liệu học có giám sát;

Điều chỉnh 1 cây bằng cách sử dụng phân vùng đệ quy nhị phân (binary recursive partitioning);

Bắt đầu với tất cả quan sát trong 1 nút duy nhất và lặp lại đệ quy các bước cho mỗi nút ngắt (Unsplit) cho đến khi tiêu chí dừng được áp dụng. Trong bước này, 2 tham số turning, mtry (số lượng các mẫu biến ngẫu nhiên như các ứng cử viên ở mỗi phần) và ntree (số lượng cây) được sử dụng. Cuối cùng, đưa ra 1 dự đoán của 1 điểm mới x. Đối với hồi quy, giá trị trung bình của 1 số cây hồi quy là kết quả cuối cùng.

b. SVM – Support vector machine (Máy hỗ trợ vector) [10]

SVM là một trong những thuật toán phân lớp phổ biến và hiệu quả. SVM là một khái niệm trong thống kê và khoa học máy tính cho một tập hợp các phương pháp học có giám sát liên quan đến nhau để phân loại và phân tích hồi quy. SVM dạng chuẩn nhận dữ liệu vào và phân loại chúng vào hai lớp khác nhau. Do đó, SVM là một thuật toán phân loại nhị phân. Với một bộ các ví dụ luyện tập thuộc hai thể loại cho trước, thuật toán luyện tập SVM xây dựng một mô hình SVM để phân loại các ví dụ khác vào hai thể loại đó. Một mô hình SVM là một cách biểu diễn các điểm trong không gian và lựa chọn ranh giới giữa hai thể loại sao cho khoảng cách từ các ví dụ luyện tập tới ranh giới là xa nhất có thể. Các ví dụ mới cũng được biểu diễn trong cùng một không gian và được thuật toán dự đoán thuộc một trong hai thể loại tùy vào ví dụ đó nằm ở phía nào của ranh giới.

c. KNN – K Nearest neighbors (K- Điểm dữ liệu gần nhất) [10]

KNN là 1 loại thuật toán phân cụm (Clustering), 1 kỹ thuật học có giám sát sử dụng để phân loại (Classify) các điểm dữ liệu mới dựa trên vị trí (Position) của chúng trên điểm dữ liệu gần nhất. KNN dự đoán 1 mẫu mới sử dụng mẫu K- điểm dữ liệu gần nhất từ tập huấn luyện. KNN không có mô hình nào chỉ lưu trữ toàn bộ (Entire) tập dữ liệu, và hơn nữa, chúng không yêu cầu học cho thuật toán. Nó sử dụng cấu trúc dữ liệu phức tạp như các cây k-d quan sát và kết nối các mẫu mới trong quá trình dự đoán. Đối với mỗi trường hợp trong dữ liệu thử nghiệm, hàm sẽ nhận dạng KNN sử dụng khoảng cách Euclidean (khoảng cách giữa 2 điểm trên một phẳng), k là số do người dùng chỉ định. KNN có tỷ lệ lỗi nhỏ với dữ liệu lớn, nó tìm ra tối ưu lân cận gần nhất của 1 điểm và số lượng tính năng thấp (ít kích thước hơn).

KNN là một trong những thuật toán supervised learning đơn giản. Khi huấn luyện, thuật toán này không học một điều gì từ dữ liệu huấn luyện mà nhớ lại một cách máy móc toàn bộ dữ liệu đó. Đây cũng là lý do thuật toán này được sắp xếp vào loại lazy learning, mọi tính toán được thực hiện khi nó cần dự đoán đầu ra của dữ liệu mới. KNN có thể áp dụng được vào cả classification và regression. KNN còn được gọi là một thuật toán instance-based hay memory-based learning.

Tóm lại, KNN là thuật toán đi tìm đầu ra của một điểm dữ liệu mới bằng cách chỉ dựa trên thông tin của K điểm dữ liệu gần nhất trong tập huấn luyện.

Vì KNN ra quyết định dựa trên các điểm gần nhất nên có hai vấn đề cần lưu tâm. Thứ nhất, khoảng cách được định nghĩa như thế nào. Thứ hai, cần phải tính toán khoảng cách như thế nào cho hiệu quả.

Ưu điểm của KNN:

Độ phức tạp tính toán của quá trình huấn luyện bằng 0;

Việc dự đoán kết quả của dữ liệu mới rất đơn giản (sau khi đã xác định được các điểm lân cận);

Không cần giả sử về phân phối của các class.

Nhược điểm của KNN:

KNN rất nhạy cảm với nhiễu khi K nhỏ

KNN là một thuật toán mà mọi tính toán đều nằm ở khâu kiểm thử. Trong đó việc tính khoảng cách tới từng điểm dữ liệu trong tập huấn luyện tốn rất nhiều thời gian, đặc biệt là với các cơ sở dữ liệu có số chiều lớn và có nhiều dữ liệu. Với K càng lớn thì độ phức tạp cũng sẽ tăng lên. Ngoài ra, việc lưu toàn bộ dữ liệu trong bộ nhớ cũng ảnh hưởng tới hiệu năng của KNN.

d. Cubist (Mạng lập thể) [11]

Cubist là một thuật toán dựa trên các nguyên tắc (rules) được sử dụng để xây dựng các mô hình dự báo dựa trên việc phân tích dữ liệu đầu vào. Nó được phát triển dựa trên sự mở rộng của mô hình cây quyết định với khả năng xử lý lên tới hàng nghìn biến đầu vào.

Về mặt lý thuyết, thuật toán Cubist được thực hiện qua 4 bước:

Bước 1: Phân nhánh/chia tách dữ liệu để phát triển một cây quyết định hoàn chỉnh;

Bước 2: Phát triển mô hình hồi quy tại mỗi nút để dự báo dựa trên dữ liệu đã được phân nhánh/chia tách;

Bước 3: “Cắt tĩa” cây để tránh vấn đề “thiếu ăn khớp” (under-fitting) hoặc “thừa ăn khớp” (over-fitting) của các giá trị dự báo so với các giá trị thực tế. Trong bước này, một số biện pháp cắt tĩa được áp dụng để chuẩn hóa dữ liệu và làm cho mô hình “hiều” dữ liệu hơn trong quá trình học/huấn luyện/đào tạo;

Bước 4: “Làm mịn” cây để bù cho sự gián đoạn quá phân nhánh/chia tách dữ liệu.

Thuật toán Lập thể là 1 mô hình hồi quy hướng dự đoán được mô tả bởi Quinlan (Quinlan, 1992) [21] với sửa chữa bổ sung dựa vào lân cận gần nhất (nearest neighbors) trong tập huấn luyện, như đã mô tả ở Quinlan (Quinlan, 1993) [22]. Mô hình mạng lập thể đã được kết hợp bằng cách sử dụng kết hợp tuyến tính như sau:

$$\hat{y}_{par} = a \cdot x \cdot \hat{y}_k + (1 - a) \cdot x \cdot \hat{y}_p \quad (2)$$

Trong đó: \hat{y}_k là sự dự đoán từ thuật toán hiện hành, \hat{y}_p là mô hình mẹ ở trên nó trong cây, a là hệ số làm mịn.

Mạng lập thể là 1 mô hình cây, đồng thời (simultaneously) tạo thành (constitute) 1 tập hợp các quy tắc (a set of rules) và 1 số mô hình hồi quy tuyến tính. Nói cách khác, Lập thể sử dụng câu lệnh “Nếu” và “Sau đó”, “Nếu” là 1 điều kiện và “Sau đó” là 1 phương trình (equation) tuyến tính để tính toán ra kết quả. Hai tham số của Lập thể có thể mặc định (default) hoặc điều chỉnh được (tunable), bao gồm các số lượng các hoạt động thúc đẩy (boosting) và lân cận-số trường hợp được sử dụng để sửa dự đoán dựa trên quy tắc.

e. LR - Linear Regression (Hồi quy tuyến tính) [12]

Trong thống kê, hồi quy tuyến tính là một cách tiếp cận tuyến tính để mô hình hóa mối quan hệ giữa một phản ứng vô hướng và một hoặc nhiều biến giải thích (còn được gọi là các biến phụ thuộc và độc lập). Trường hợp của một biến giải thích được gọi là hồi quy tuyến tính đơn giản; đối với nhiều hơn một, quá trình này được gọi là hồi quy

nhiều tuyến tính. Thuật ngữ này khác với hồi quy tuyến tính đa biến, trong đó nhiều biến phụ thuộc tương quan được dự đoán, thay vì một biến vô hướng đơn lẻ.

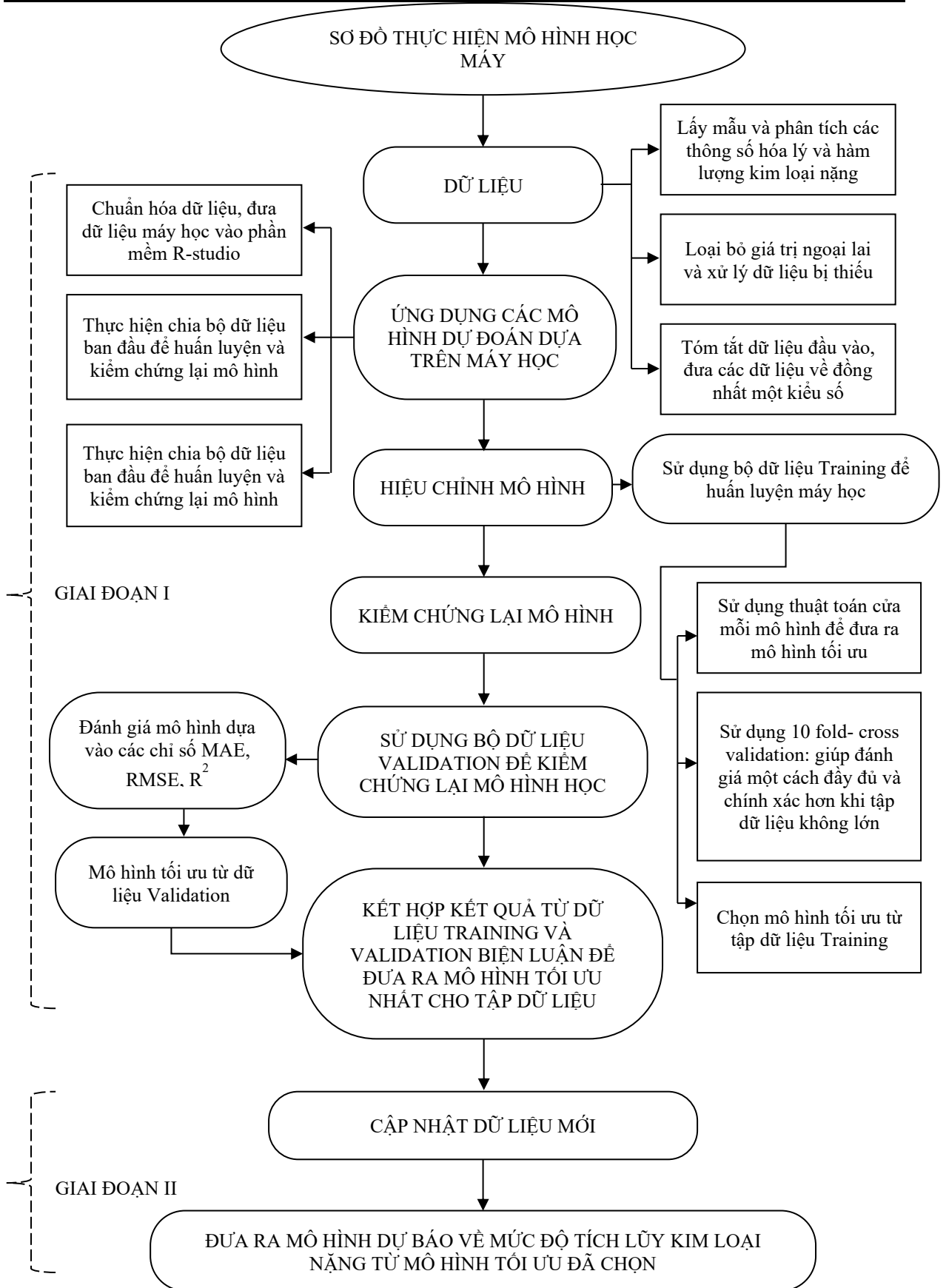
Trong hồi quy tuyến tính, các mối quan hệ được mô hình hóa bằng cách sử dụng các hàm dự báo tuyến tính mà các tham số mô hình chưa biết được ước tính từ dữ liệu. Các mô hình như vậy được gọi là mô hình tuyến tính. Thông thường nhất, giá trị trung bình có điều kiện của phản hồi cho các giá trị của các biến giải thích (hoặc các yếu tố dự đoán) được giả định là một hàm liên kết của các giá trị đó; ít phổ biến hơn, trung vị có điều kiện hoặc một số lượng tử khác được sử dụng. Giống như tất cả các hình thức phân tích hồi quy, hồi quy tuyến tính tập trung vào phân phối xác suất có điều kiện của phản hồi cho các giá trị của các yếu tố dự đoán, thay vì phân phối xác suất chung của tất cả các biến này, là lĩnh vực của phân tích đa biến.

Hạn chế của LR:

Rất nhạy cảm với nhiễu cho nên trước khi thực hiện linear regression thì nhiễu cần phải được loại bỏ (tiền xử lý)

Không biểu diễn được các mô hình phức tạp

3.2.3. Quy trình thực hiện mô hình



Hình 3.3 Sơ đồ quy trình thực hiện mô hình học máy

3.3. Xây dựng mô hình học máy

Trong nghiên cứu này, các mô hình học máy được triển khai nhằm dự báo chỉ số nguy cơ sinh thái (RI) dựa trên nồng độ các kim loại nặng như As, Pb, Hg, Cd và Zn trong trầm tích. Quá trình xây dựng mô hình được thực hiện với các bước cụ thể như sau:

3.3.1. Làm sạch môi trường và cài đặt thư viện

Trước tiên, để đảm bảo môi trường làm việc trong R không bị ảnh hưởng bởi các biến cũ, cần xóa toàn bộ các đối tượng và tắt các cửa sổ đồ họa đang mở. Đồng thời, cài đặt và gọi các thư viện cần thiết:

Làm sạch môi trường làm việc

```
rm(list=ls())
```

```
graphics.off()
```

Cài đặt và gọi thư viện

```
install.packages(c("readxl", "tidyverse", "caret", "randomForest", "e1071", "Cubist",  
"kknn", "gridExtra", "ggplot2", "Metrics"))
```

```
library(readxl)
```

```
library(tidyverse)
```

```
library(caret)
```

```
library(randomForest)
```

```
library(e1071)
```

```
library(Cubist)
```

```
library(Metrics)
```

```
library(ggplot2)
```

```
library(gridExtra)
```

```
library(kknn)
```

3.3.2. Đọc dữ liệu và kiểm tra

Dữ liệu được nhập từ tệp Excel chứa nồng độ kim loại nặng và chỉ số rủi ro RI tương ứng:

Đọc dữ liệu từ tệp Excel

```
data <- read_excel("E:/ĐỒ ÁN TỐT NGHIỆP/File  
thuyếtminh/DULIEUCHAYHOCMAYTG1.xlsx", sheet = "Sheet1")
```

```
# Kiểm tra dữ liệu
```

```
head(data)
```

```
str(data)
```

```
summary(data %>% select(As, Pb, Hg, Cd, Zn, RI))
```

3.3.3. Phân tích phân phối và mối tương quan

Để đánh giá sự phân bố và mối quan hệ giữa các biến đầu vào, các biểu đồ phân phối và ma trận tương quan được sử dụng:

```
# Vẽ biểu đồ
```

```
data %>%
```

```
select(As, Pb, Hg, Cd, Zn) %>%
```

```
pivot_longer(cols = everything(), names_to = "Kim_loai", values_to = "Gia_tri")  
%>%
```

```
ggplot(aes(x = Gia_tri)) +
```

```
geom_histogram(bins = 30, fill = "steelblue", color = "white") +
```

```
facet_wrap(~Kim_loai, scales = "free") +
```

```
labs(x = "Nồng độ (mg/kg)", y = "Tần suất")
```

3.3.4. Chia dữ liệu thành tập huấn luyện và kiểm tra

Dữ liệu được chia thành hai tập: 70% dùng để huấn luyện mô hình, 30% dùng để kiểm tra hiệu suất mô hình:

```
# Chia dữ liệu
```

```
set.seed(123)
```

```
train_index <- sample(nrow(data), size = 0.7 * nrow(data))
```

```
train_data <- data[train_index, ]
```

```
test_data <- data[-train_index, ]
```

```
# Loại bỏ cột không cần thiết
```

```
train_data <- train_data %>% select(-TT)
```

```
test_data <- test_data %>% select(-TT)
```

3.3.5. Thiết lập công thức mô hình và kiểm tra chéo

Các biến đầu vào được khai báo và mô hình được huấn luyện với kỹ thuật kiểm tra chéo 10-fold lặp lại 3 lần để đảm bảo tính ổn định:

```
predictors <- c("As", "Pb", "Hg", "Cd", "Zn")
```

```
formula_KQ <- as.formula(paste("KQ ~", paste(predictors, collapse = " + ")))  
ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3, verboseIter =  
TRUE)
```

3.3.6. Huấn luyện các mô hình

Năm thuật toán được huấn luyện trên cùng một tập dữ liệu để so sánh hiệu suất:

```
models <- list(  
"Random Forest" = train(formula_KQ, data = train_data, method = "rf", trControl =  
ctrl),  
"SVM" = train(formula_KQ, data = train_data, method = "svmRadial", trControl =  
ctrl),  
"KNN" = train(formula_KQ, data = train_data, method = "knn", trControl = ctrl),  
"Cubist" = train(formula_KQ, data = train_data, method = "cubist", trControl = ctrl),  
"Linear Regression" = train(formula_KQ, data = train_data, method = "lm", trControl  
= ctrl))
```

3.3.7. Đánh giá và so sánh mô hình

Sau khi huấn luyện, các mô hình được đánh giá dựa trên ba chỉ số phổ biến: R^2 , RMSE và MAE, trên cả hai tập huấn luyện và kiểm tra:

Dự đoán trên tập huấn luyện

```
pred_train <- predict(model, newdata = train_data)  
metric_train <- data.frame(  
Model = model_name,  
Dataset = "Training",  
R2 = R2(pred_train, train_data$KQ),  
RMSE = RMSE(pred_train, train_data$KQ),  
MAE = MAE(pred_train, train_data$KQ))
```

Dự đoán trên tập kiểm tra

```
pred_test <- predict(model, newdata = test_data)  
metric_test <- data.frame(  
Model = model_name,  
Dataset = "Test",  
R2 = R2(pred_test, test_data$KQ),
```

```
RMSE = RMSE(pred_test, test_data$KQ),
MAE = MAE(pred_test, test_data$KQ))
results_KQ <- rbind(results_KQ, metric_train, metric_test)}
```

3.3.8. Trực quan hóa kết quả mô hình

Để dễ dàng so sánh giữa các mô hình, các biểu đồ trực quan được sử dụng:

Biểu đồ so sánh R^2

```
ggplot(results_KQ, aes(x = Model, y = R2, fill = Dataset)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "So sánh R2 giữa các mô hình", y = "R2") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Dotplot so sánh kết quả từ kiểm tra chéo

```
resamples_obj <- resamples(models)
dotplot(resamples_obj, main = "So sánh các mô hình học máy")
```

3.3.9. Lựa chọn mô hình tối ưu

Mô hình có hiệu suất tốt nhất trên tập kiểm tra sẽ được lựa chọn làm mô hình dự đoán chính cho chỉ số RI:

```
best_model <- results_KQ %>%
  filter(Dataset == "Test") %>%
  arrange(desc(R2)) %>%
  slice(1)
cat("Mô hình tốt nhất cho chỉ số KQ:\n")
print(best_model)
```

3.4. Đánh giá hiệu suất mô hình

3.4.1. Các chỉ số đánh giá

Hiệu suất của mô hình được đánh giá thông qua ba chỉ số phổ biến gồm: Hệ số xác định (R-squared – R^2), Căn bậc hai của sai số bình phương trung bình (Root Mean Square Error – RMSE) và Sai số tuyệt đối trung bình (Mean Absolute Error – MAE).

R^2 : Chỉ số này phản ánh mức độ mà mô hình có thể giải thích được sự biến động của biến mục tiêu (nồng độ kim loại nặng). Giá trị R^2 dao động từ 0 đến 1, trong đó giá trị càng gần 1 thể hiện mô hình càng có khả năng dự báo tốt.

RMSE: Đây là chỉ số thể hiện độ lệch trung bình giữa giá trị dự báo và giá trị thực

tế, được tính bằng căn bậc hai trung bình của bình phương sai số. RMSE có đơn vị giống như biến mục tiêu và nhạy với các giá trị ngoại lai. Giá trị RMSE càng nhỏ cho thấy mô hình dự báo càng chính xác.

MAE: Là trung bình cộng của các sai số tuyệt đối giữa giá trị dự báo và giá trị thực tế. MAE cho biết sai số trung bình một cách trực quan, dễ hiểu hơn so với RMSE. Giá trị MAE càng nhỏ chứng tỏ mô hình có độ chính xác càng cao.

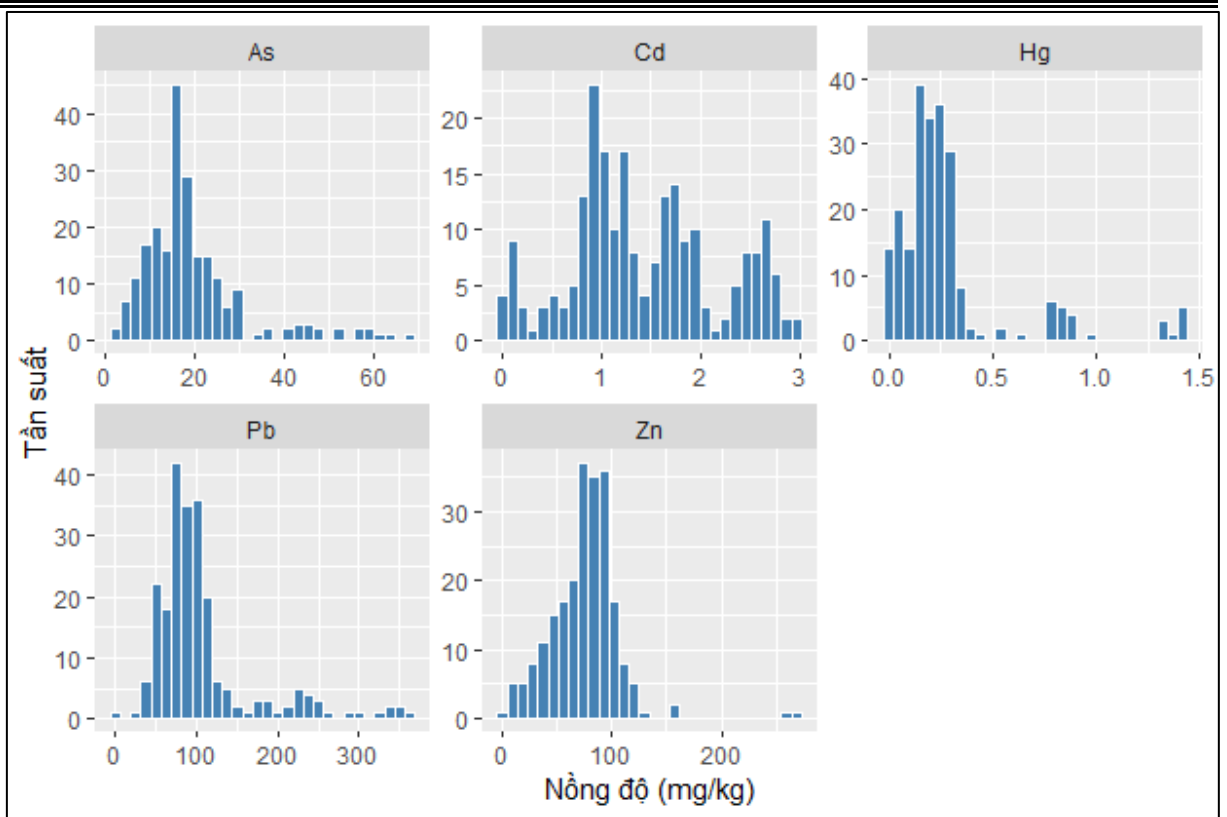
Việc kết hợp cả ba chỉ số giúp đánh giá mô hình một cách toàn diện, đảm bảo tính khách quan khi so sánh giữa các thuật toán hoặc các kịch bản huấn luyện khác nhau.

3.4.2. So sánh hiệu suất giữa các mô hình

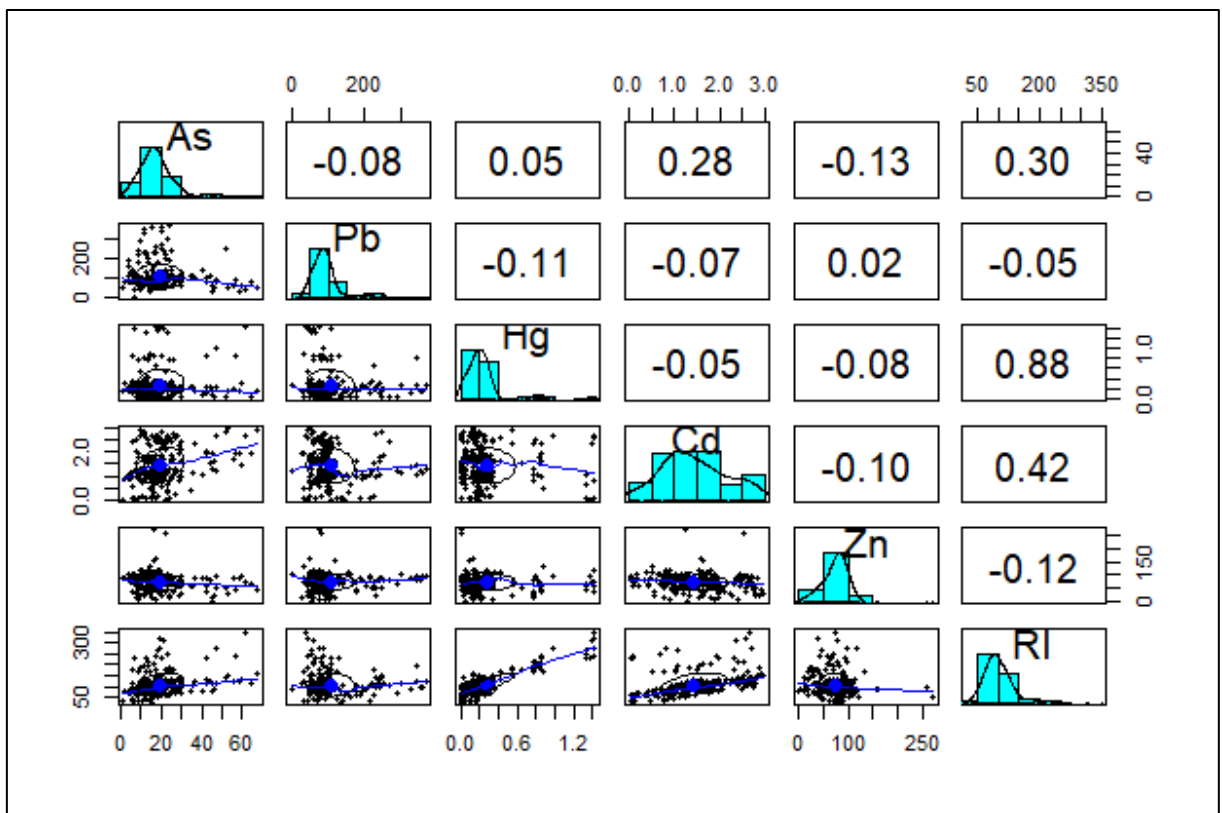
Sau khi tiến hành huấn luyện và kiểm tra các mô hình học máy khác nhau trên cùng tập dữ liệu, bước tiếp theo là thực hiện so sánh hiệu suất nhằm xác định mô hình nào có độ chính xác cao nhất và phù hợp nhất với mục tiêu dự báo.

```
> # Kiểm tra dữ liệu
> head(data)
# A tibble: 6 × 7
  TT     As     Pb     Hg     Cd     Zn     RI
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1     1  16.7  110.  0.343  1.10  61.5  107.
2     2  25.3  239.  0.299  1.60  88.3  130.
3     3  21.9  214.  0.286  1.06  93.1  108.
4     4  20.9  241.  0.293  1.02  83.4  109.
5     5  20.1  77.3  0.229  1.00  74.1  86.0
6     6  20.7  235.  0.263  1.31  82.9  112.
> str(data)
tibble [225 × 7] (S3: tbl_df/tbl/data.frame)
 $ TT: num [1:225] 1 2 3 4 5 6 7 8 9 10 ...
 $ As: num [1:225] 16.7 25.3 21.9 20.9 20.1 20.7 20 23.6 22.1 19.8 ...
 $ Pb: num [1:225] 110.1 239.2 214.5 241.1 77.3 ...
 $ Hg: num [1:225] 0.343 0.299 0.286 0.293 0.229 0.263 0.282 0.28 0.246 0.146 ...
 $ Cd: num [1:225] 1.1 1.6 1.06 1.02 1 ...
 $ Zn: num [1:225] 61.5 88.3 93.1 83.4 74.1 82.9 96.3 92.3 83.9 95.7 ...
 $ RI: num [1:225] 107 130 108 109 86 ...
> summary(data %>% select(As, Pb, Hg, Cd, Zn, RI))
   As           Pb           Hg           Cd           Zn           RI
Min.   : 1.20   Min.   : 2.4    Min.   :0.0040   Min.   :0.010   Min.   : 0.70   Min.   : 24.85
1st Qu.:12.40   1st Qu.: 72.8    1st Qu.:0.1320   1st Qu.:0.928   1st Qu.: 59.80   1st Qu.: 78.83
Median :16.90   Median : 88.1    Median :0.2100   Median :1.299   Median : 77.80   Median : 95.45
Mean   :19.42   Mean   :106.1    Mean   :0.2817   Mean   :1.433   Mean   : 75.11   Mean   :109.03
3rd Qu.:22.10   3rd Qu.:108.8    3rd Qu.:0.2820   3rd Qu.:1.906   3rd Qu.: 91.00   3rd Qu.:124.00
Max.   :67.10   Max.   :365.1    Max.   :1.4230   Max.   :2.979   Max.   :269.30   Max.   :350.76
```

Hình 3.4 Kết quả sau khi chạy kiểm tra dữ liệu



Hình 3.5 Biểu đồ phân phối của từng biến

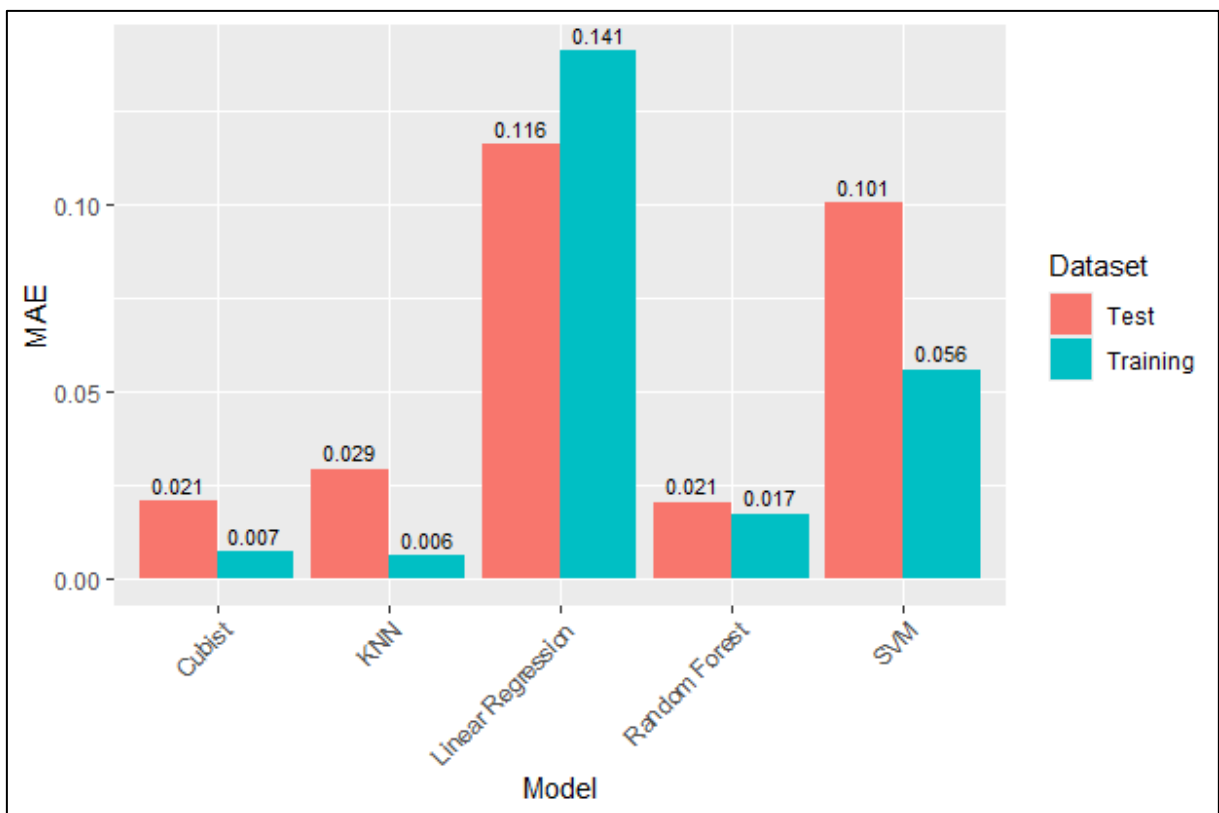


Hình 3.6 Biểu đồ tương quan giữa các biến

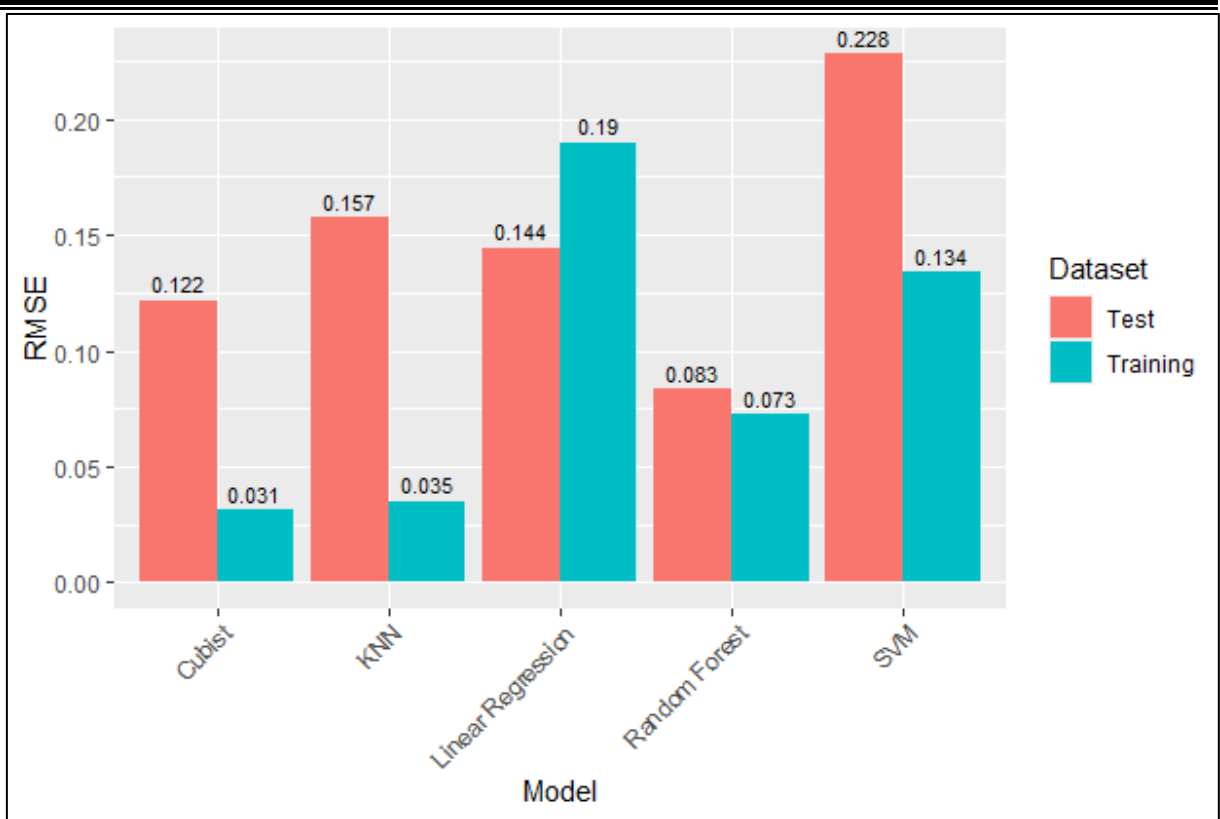
```
> print(results_KQ)
```

	Model	Dataset	R2	RMSE	MAE
1	Random Forest	Training	0.9678883	0.07256678	0.017492144
2	Random Forest	Test	0.9521200	0.08348982	0.020565686
3	SVM	Training	0.9162279	0.13375987	0.056067672
4	SVM	Test	0.6780586	0.22798299	0.100576733
5	KNN	Training	0.9933114	0.03521377	0.006284167
6	KNN	Test	0.8383741	0.15733193	0.029411765
7	Cubist	Training	0.9941469	0.03111103	0.007409026
8	Cubist	Test	0.8987918	0.12169482	0.020872025
9	Linear Regression	Training	0.7802264	0.18965141	0.141215712
10	Linear Regression	Test	0.8477674	0.14439515	0.116239566

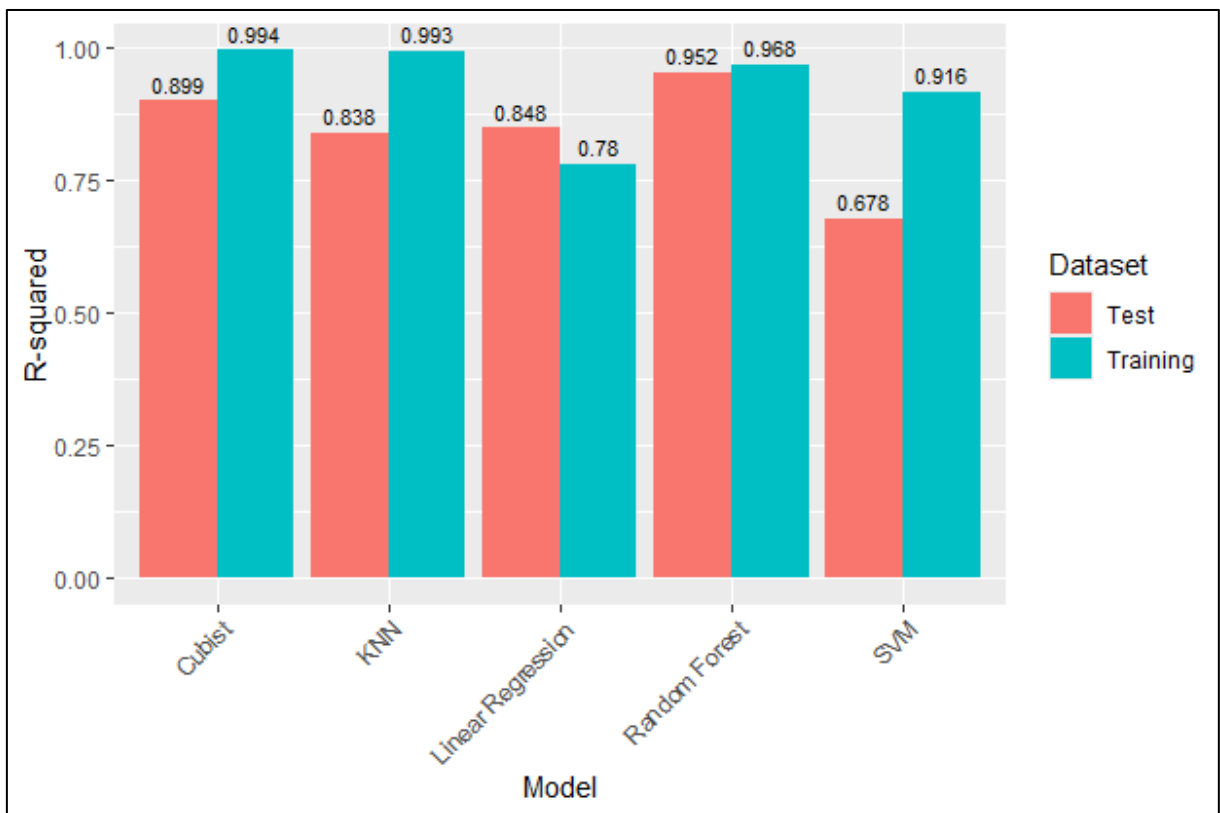
Hình 3.7 Kết quả sau khi chạy mô hình



Hình 3.8 So sánh MAE giữa các mô hình



Hình 3.9 So sánh RMSE giữa các mô hình



Hình 3.10 So sánh R^2 giữa các mô hình

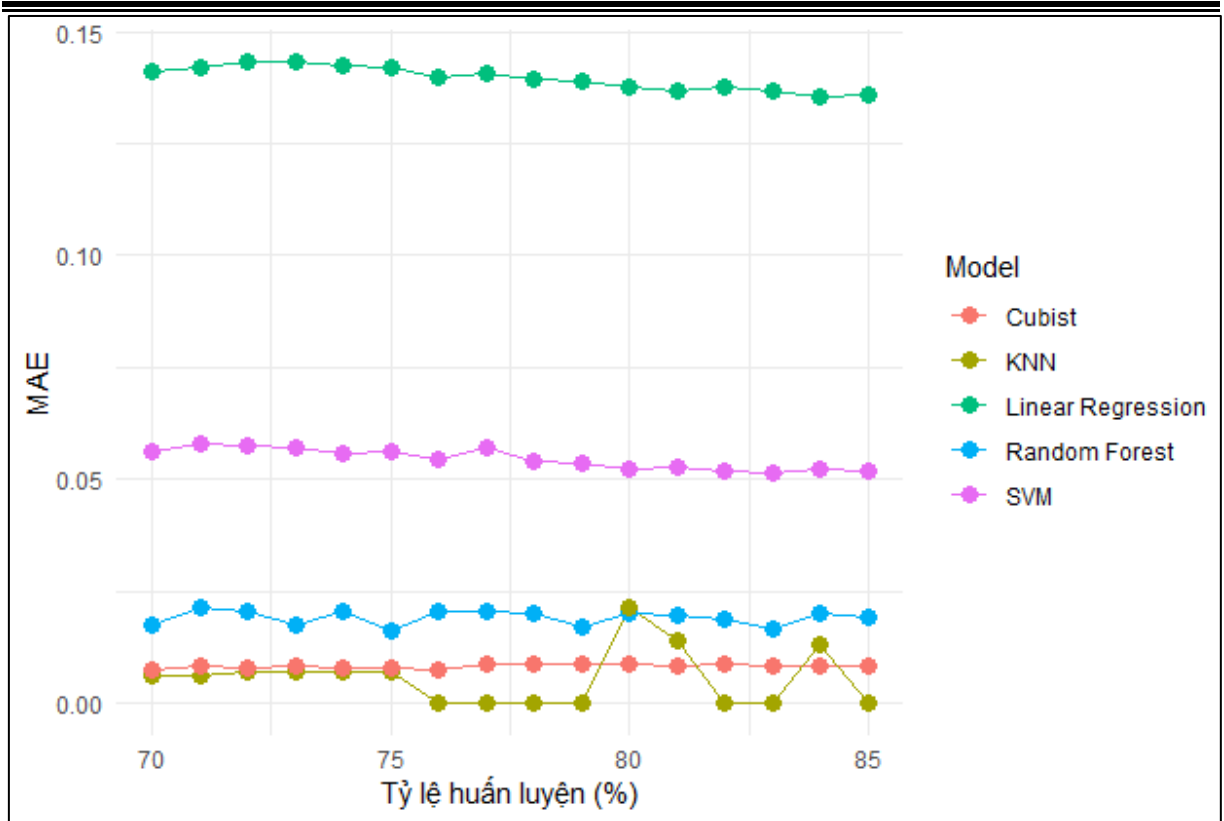
Sau khi có kết quả chạy các mô hình học máy sinh viên tiến hành chạy các mô hình này với các tỷ lệ training (70-85) và testing (30-15) để so sánh và đánh giá ảnh hưởng của lượng dữ liệu huấn luyện đến hiệu suất mô hình.

Bảng 3.2 Kết quả chạy các mô hình theo các tỷ lệ khác nhau từ tập dữ liệu Training

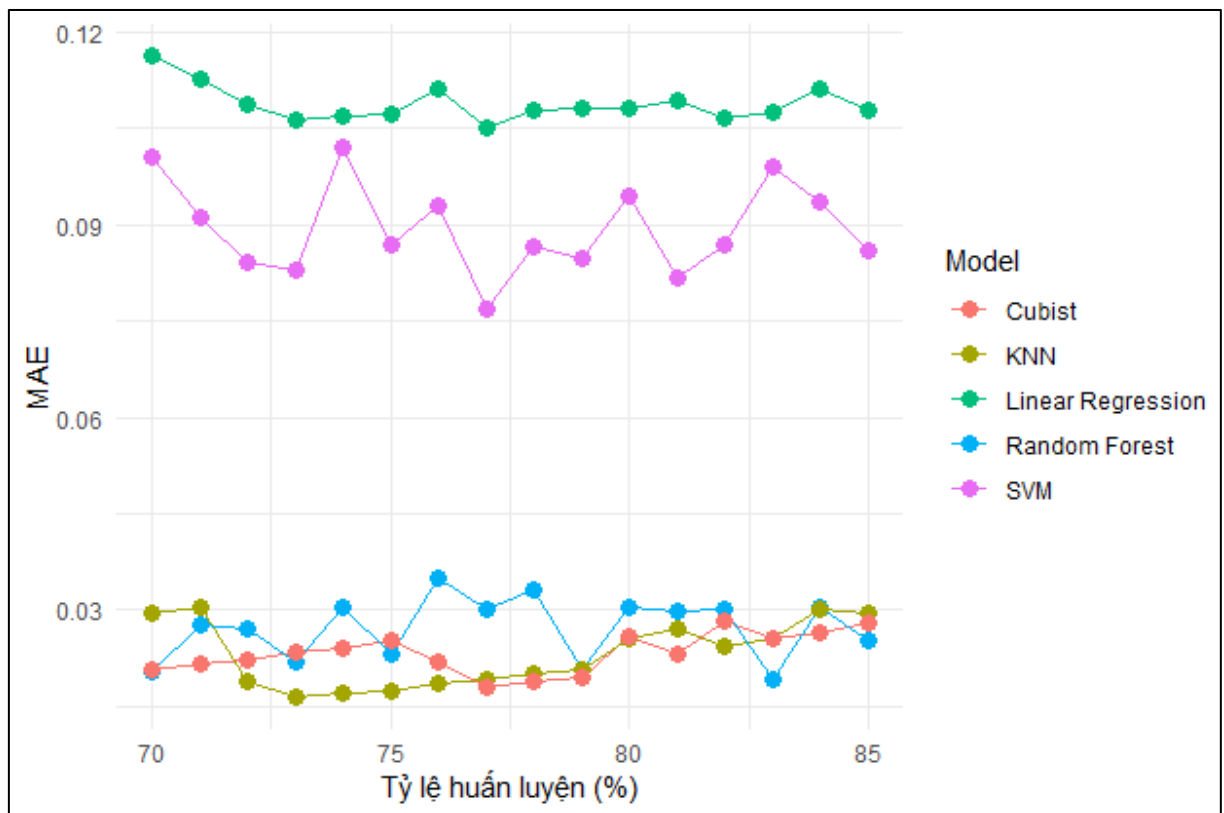
Split Ratio	MAE					RMSE					R ²				
	LR	SVM	KNN	RF	Cubist	LR	SVM	KNN	RF	Cubist	LR	SVM	KNN	RF	Cubist
70	0,141	0,056	0,006	0,018	0,007	0,190	0,134	0,035	0,074	0,031	0,780	0,916	0,993	0,967	0,994
71	0,142	0,056	0,006	0,017	0,008	0,190	0,134	0,035	0,073	0,035	0,777	0,915	0,993	0,967	0,993
72	0,143	0,060	0,007	0,016	0,008	0,190	0,137	0,038	0,070	0,035	0,778	0,912	0,992	0,970	0,993
73	0,143	0,056	0,007	0,020	0,008	0,190	0,134	0,038	0,069	0,034	0,777	0,916	0,992	0,972	0,993
74	0,142	0,056	0,007	0,020	0,008	0,189	0,134	0,038	0,070	0,035	0,777	0,915	0,992	0,972	0,993
75	0,142	0,056	0,007	0,017	0,008	0,188	0,134	0,037	0,070	0,034	0,776	0,915	0,992	0,970	0,993
76	0,140	0,055	0,000	0,016	0,007	0,187	0,130	0,000	0,069	0,033	0,782	0,918	1,000	0,970	0,993
77	0,141	0,054	0,000	0,020	0,009	0,188	0,127	0,000	0,069	0,037	0,782	0,922	1,000	0,972	0,992
78	0,139	0,055	0,000	0,021	0,009	0,187	0,129	0,000	0,071	0,038	0,783	0,919	1,000	0,970	0,991
79	0,139	0,053	0,000	0,020	0,009	0,186	0,130	0,000	0,072	0,038	0,782	0,919	1,000	0,969	0,991
80	0,138	0,053	0,022	0,020	0,009	0,185	0,124	0,092	0,068	0,037	0,786	0,925	0,956	0,973	0,992
81	0,137	0,052	0,014	0,020	0,009	0,185	0,128	0,066	0,071	0,036	0,786	0,922	0,976	0,970	0,992
82	0,137	0,052	0,000	0,016	0,009	0,185	0,122	0,000	0,070	0,037	0,789	0,927	1,000	0,970	0,992
83	0,137	0,055	0,000	0,017	0,008	0,184	0,127	0,000	0,071	0,037	0,789	0,919	1,000	0,969	0,992
84	0,135	0,052	0,013	0,020	0,008	0,183	0,123	0,065	0,069	0,036	0,789	0,925	0,977	0,972	0,992
85	0,136	0,052	0,000	0,017	0,008	0,183	0,123	0,000	0,071	0,036	0,792	0,925	1,000	0,968	0,992

Bảng 3.3 Kết quả chạy các mô hình theo các tỷ lệ khác nhau từ tập dữ liệu Testing

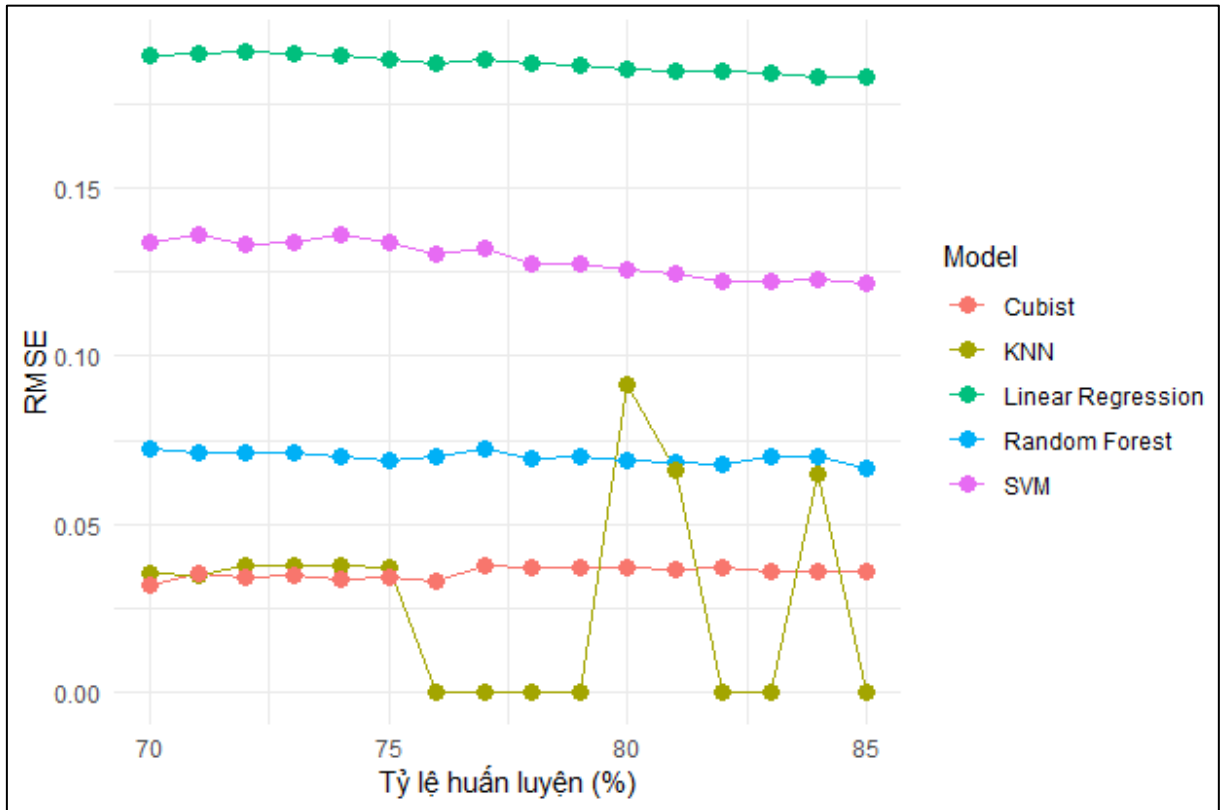
Split Ratio	MAE					RMSE					R ²				
	LR	SVM	KNN	RF	Cubist	LR	SVM	KNN	RF	Cubist	LR	SVM	KNN	RF	Cubist
70	0,116	0,105	0,029	0,020	0,021	0,144	0,233	0,157	0,081	0,122	0,848	0,659	0,838	0,954	0,899
71	0,113	0,099	0,030	0,021	0,022	0,140	0,230	0,160	0,084	0,121	0,859	0,687	0,838	0,952	0,902
72	0,109	0,077	0,019	0,021	0,022	0,134	0,205	0,107	0,084	0,118	0,868	0,735	0,936	0,949	0,903
73	0,106	0,090	0,016	0,028	0,023	0,131	0,229	0,106	0,083	0,121	0,876	0,672	0,942	0,951	0,901
74	0,107	0,097	0,017	0,033	0,024	0,133	0,239	0,108	0,089	0,122	0,877	0,653	0,942	0,946	0,902
75	0,107	0,097	0,018	0,023	0,025	0,133	0,241	0,109	0,088	0,130	0,879	0,659	0,941	0,950	0,893
76	0,111	0,084	0,019	0,024	0,022	0,136	0,220	0,136	0,091	0,099	0,867	0,697	0,891	0,944	0,935
77	0,105	0,090	0,019	0,029	0,019	0,128	0,232	0,139	0,086	0,091	0,871	0,623	0,885	0,948	0,945
78	0,108	0,084	0,020	0,033	0,023	0,130	0,223	0,141	0,095	0,112	0,871	0,679	0,885	0,938	0,921
79	0,108	0,101	0,021	0,029	0,024	0,131	0,254	0,144	0,091	0,114	0,877	0,569	0,884	0,945	0,921
80	0,108	0,082	0,026	0,030	0,022	0,132	0,222	0,144	0,092	0,098	0,870	0,668	0,888	0,948	0,938
81	0,109	0,102	0,027	0,030	0,023	0,134	0,257	0,154	0,097	0,101	0,873	0,532	0,877	0,948	0,938
82	0,107	0,088	0,024	0,018	0,028	0,129	0,232	0,156	0,079	0,124	0,861	0,646	0,895	0,968	0,907
83	0,108	0,079	0,026	0,018	0,030	0,131	0,217	0,160	0,078	0,127	0,864	0,768	0,895	0,971	0,907
84	0,111	0,090	0,030	0,027	0,026	0,135	0,245	0,167	0,093	0,109	0,867	0,685	0,892	0,960	0,933
85	0,108	0,082	0,029	0,015	0,028	0,132	0,241	0,171	0,081	0,113	0,850	0,809	1,000	0,999	0,928



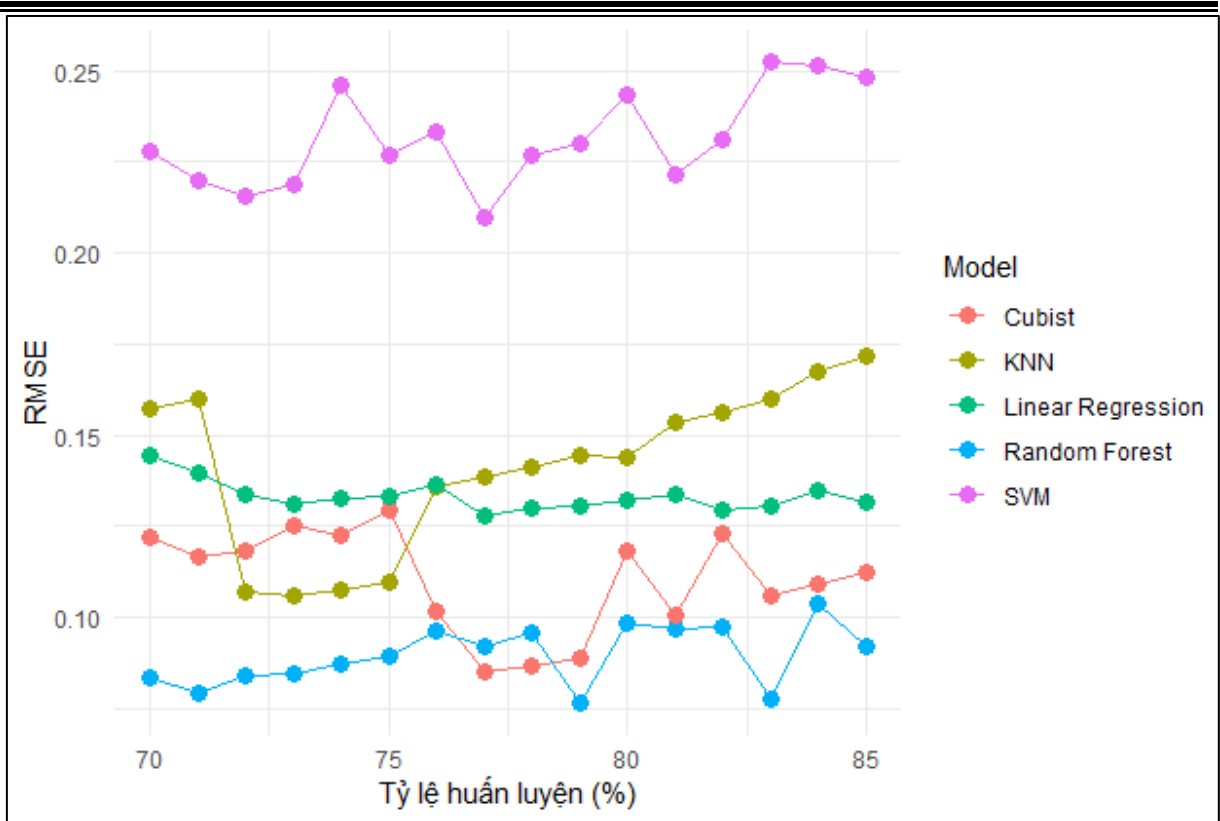
Hình 3.11 MAE trên tập Train theo các tỷ lệ



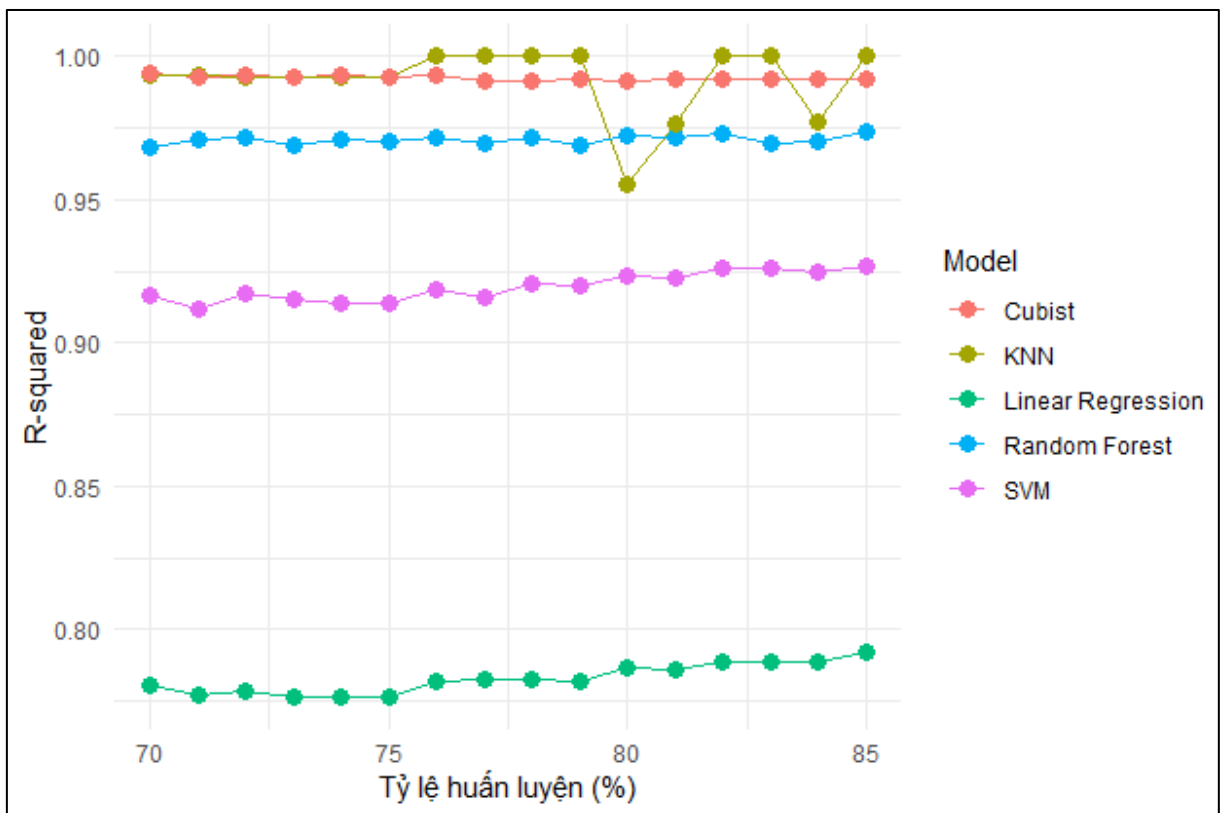
Hình 3.12 MAE trên tập Test theo các tỷ lệ



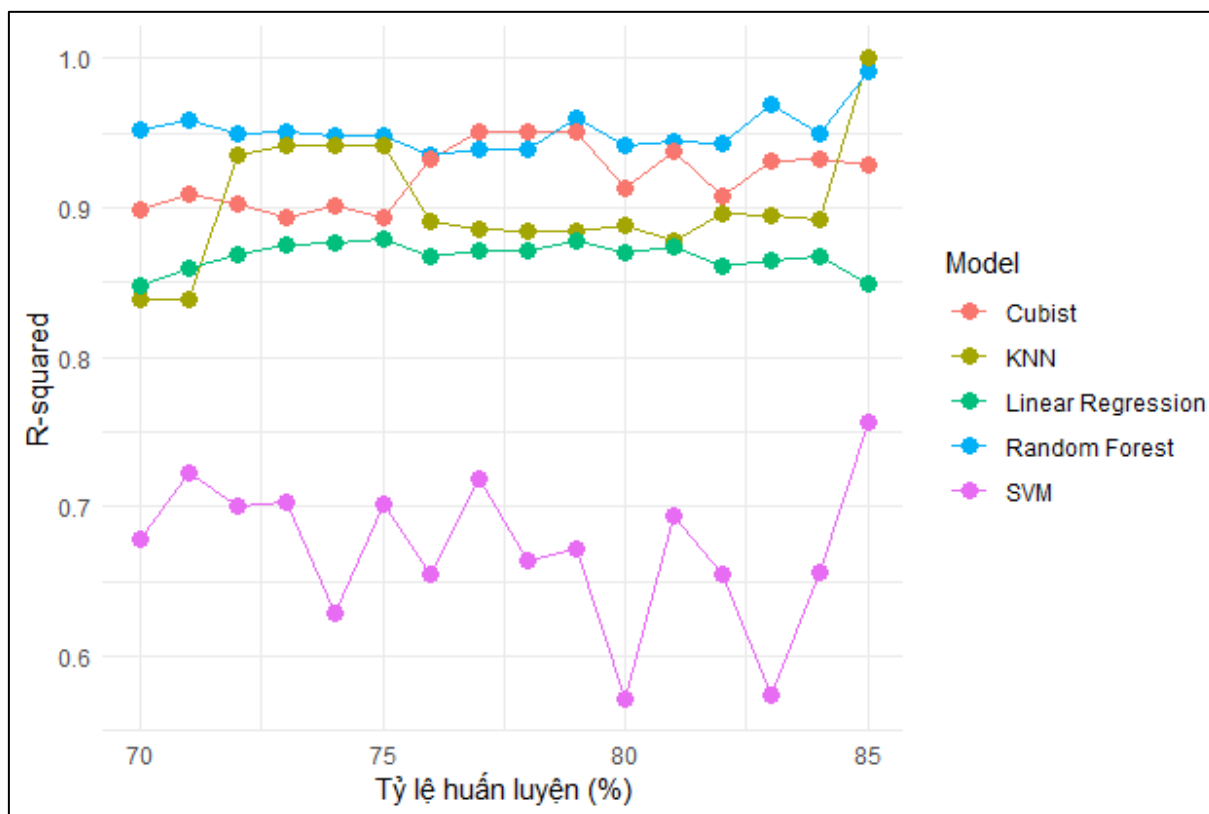
Hình 3.13 RMSE trên tập Train theo các tỷ lệ



Hình 3.14 RMSE trên tập Test theo các tỷ lệ



Hình 3.15 R^2 trên tập Train theo các tỷ lệ



Hình 3.16 R^2 trên tập Test theo các tỷ lệ

Trong nghiên cứu này, để xây dựng mô hình dự báo kim loại nặng và rủi ro sinh thái trong trầm tích mặt tại cửa An Hòa, sông Trường Giang. Sinh viên đã thử nghiệm và so sánh hiệu suất của năm thuật toán học máy phổ biến gồm: Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Cubist và Linear Regression (hồi quy tuyến tính). Việc đánh giá các mô hình được thực hiện dựa trên ba chỉ số quan trọng: hệ số xác định (R^2), lỗi trung bình tuyệt đối (MAE) và căn phương sai trung bình lỗi (RMSE) trên cả tập huấn luyện và tập kiểm tra.

Mô hình Linear Regression (LR) cho thấy hiệu suất không ổn định giữa hai tập dữ liệu. Trên tập huấn luyện, LR đạt R^2 trung bình khoảng 0,78–0,79, nhưng MAE và RMSE khá cao (MAE \approx 0,135–0,141; RMSE \approx 0,18–0,19). Trên tập kiểm tra, LR đạt R^2 trung bình khoảng 0,85–0,88; MAE \approx 0,105–0,116; RMSE \approx 0,128–0,136. Điều này phản ánh hạn chế của LR trong việc mô hình hóa các mối quan hệ phi tuyến phức tạp trong dữ liệu môi trường.

Mô hình SVM thể hiện hiệu suất tốt trên tập huấn luyện ($R^2 \approx$ 0,912–0,927; MAE \approx 0,052–0,056; RMSE \approx 0,123–0,137), trên tập kiểm tra ($R^2 \approx$ 0,532–0,809; MAE \approx

0,077–0,105; RMSE \approx 0,217–0,254). Sự suy giảm này cho thấy SVM có thể bị ảnh hưởng bởi hiện tượng overfitting hoặc thiếu tính tổng quát hóa khi áp dụng vào dữ liệu mới.

Mô hình KNN đạt kết quả gần như hoàn hảo trên tập huấn luyện ($R^2 \approx 0,99$ –1,0; MAE \approx 0,006–0,022; RMSE \approx 0,000–0,092), nhưng hiệu suất giảm mạnh trên tập kiểm tra ($R^2 \approx 0,84$ –1,0; MAE \approx 0,016–0,03; RMSE \approx 0,106–0,171). Sự chênh lệch lớn này là dấu hiệu rõ ràng của overfitting, khiến KNN không phù hợp để áp dụng trong thực tế.

Mô hình Random Forest (RF) và Cubist thể hiện hiệu suất vượt trội và ổn định trên cả hai tập dữ liệu. RF đạt $R^2 \approx 0,967$ –0,973; MAE \approx 0,016–0,020; RMSE \approx 0,068–0,074 trên tập huấn luyện và $R^2 \approx 0,938$ –0,999; MAE \approx 0,015–0,03 và RMSE \approx 0,078–0,095 trên tập kiểm tra. Trong khi đó, Cubist cũng cho kết quả tốt với $R^2 \approx 0,991$ –0,994; MAE \approx 0,007–0,009; RMSE \approx 0,031–0,038 trên tập huấn luyện và $R^2 \approx 0,893$ –0,945; MAE \approx 0,019–0,030; RMSE \approx 0,091–0,130 trên tập kiểm tra. Cả hai mô hình này đều duy trì được độ chính xác cao mà không có dấu hiệu overfitting.

Dựa trên phân tích toàn diện, mô hình Random Forest (RF) được lựa chọn là mô hình chính thức của nghiên cứu nhờ vào hiệu suất vượt trội, sai số thấp, khả năng khái quát hóa tốt và tính ổn định cao trên tập dữ liệu kiểm tra. Mô hình Cubist được xem là phương án dự phòng tin cậy, thích hợp khi cần thay thế hoặc so sánh mô hình trong các điều kiện cụ thể.

Các mô hình còn lại (LR, SVM, KNN) không được lựa chọn do hạn chế về độ tin cậy hoặc khả năng tổng quát hóa. Đặc biệt, KNN và SVM cho thấy sự biến động lớn giữa tập huấn luyện và kiểm tra, trong khi LR không đủ mạnh để xử lý các mối quan hệ phức tạp trong dữ liệu môi trường nên đã bị loại bỏ.

Kết quả này không chỉ khẳng định tiềm năng ứng dụng của Random Forest và Cubist trong dự báo ô nhiễm kim loại nặng, mà còn cung cấp cơ sở khoa học để phát triển các hệ thống giám sát môi trường tự động. Trong bối cảnh biến đổi khí hậu và gia tăng áp lực lên tài nguyên nước, việc sử dụng các mô hình học máy tiên tiến như Cubist có thể hỗ trợ hiệu quả công tác quản lý và giảm thiểu rủi ro sinh thái tại các khu vực cửa sông.

Nghiên cứu cũng mở ra hướng phát triển mới trong việc tích hợp đa dữ liệu để nâng cao độ chính xác của mô hình, đồng thời đề xuất ứng dụng các kỹ thuật tối ưu hóa tham số nhằm cải thiện hiệu suất dự báo trong tương lai.

Kết luận: Random Forest và Cubist là hai mô hình tối ưu nhất để dự báo tích lũy kim loại nặng trong trầm tích tại khu vực nghiên cứu, với ưu điểm vượt trội về độ chính xác, tính ổn định và khả năng tổng quát hóa.

3.5. Chạy dự báo nồng độ các kim loại nặng

Sau khi đánh giá và hiệu chỉnh mô hình, tiến hành dự báo nồng độ các kim loại As, Pb, Hg, Cd và Zn cho năm 2026 tại các vị trí lấy mẫu tương ứng. Mục tiêu nhằm ước lượng xu hướng biến động và nguy cơ tích lũy kim loại nặng trong tương lai gần, từ đó hỗ trợ công tác quản lý và kiểm soát ô nhiễm môi trường. Các bước thực hiện chạy:

1. Dọn sạch môi trường

```
rm(list = ls())  
graphics.off()
```

2. Cài đặt và gọi gói cần thiết

```
if (!require(pacman)) install.packages("pacman")  
pacman::p_load(readxl, dplyr, caret, randomForest, writexl)
```

3. Đọc dữ liệu từ file Excel

```
data <- read_excel("E:/ĐỒ ÁN TỐT NGHIỆP/File thuyết minh/File  
excel/DULIEUCHAYHOCMAYTG.xlsx", sheet = "Sheet1")
```

4. Tiền xử lý: chỉ lấy các cột cần thiết và lọc 2024–2025

```
data <- data %>%  
  rename(Vi_tri = `Vị trí lấy mẫu`) %>%  
  select(Vi_tri, Năm, As, Pb, Hg, Cd, Zn) %>%  
  filter(Năm %in% c(2024, 2025)) %>%  
  mutate(  
    Vi_tri = as.factor(Vi_tri),  
    Năm = as.numeric(Năm) # Chuyển sang numeric để tránh lỗi "new level"  
  ) %>%  
  na.omit()
```

5. Tạo dữ liệu dự báo cho năm 2026 (15 vị trí)

```
vi_tri_list <- levels(data$Vi_tri)
```

```
du_bao_2026 <- data.frame(  
  Vi_tri = factor(vi_tri_list, levels = vi_tri_list),  
  Năm = 2026 # Numeric (không phải factor))
```

6. Cấu hình huấn luyện mô hình

```
set.seed(123)  
ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
```

7. Huấn luyện và dự báo từng kim loại

```
kim_loai <- c("As", "Pb", "Hg", "Cd", "Zn")  
for (kl in kim_loai) {  
  cat("\n Huấn luyện mô hình cho:", kl, "\n")  
  df <- data %>%  
    select(Vi_tri, Năm, all_of(kl)) %>%  
    rename(GT = all_of(kl))  
  model <- train(  
    GT ~ Vi_tri + Năm,  
    data = df,  
    method = "rf",  
    trControl = ctrl,  
    preProcess = c("center", "scale") )  
  du_bao_2026[[kl]] <- predict(model, newdata = du_bao_2026)  
  cat("Dự báo xong cho", kl, "\n")  
  du_bao_2026 <- du_bao_2026 %>%  
    mutate(Vi_tri_so = as.numeric(gsub("S", "", Vi_tri))) %>%  
    arrange(Vi_tri_so) %>%  
    select(-Vi_tri_so)
```

8. Hiển thị kết quả

```
cat("\n KẾT QUẢ DỰ BÁO NĂM 2026:\n")  
print(du_bao_2026)
```

Bảng 3.4 Nồng độ các kim loại dự báo

Vị trí	Năm	As (mg/Kg)	Pb(mg/Kg)	Hg(mg/Kg)	Cd(mg/Kg)	Zn(mg/Kg)
S1	2026	19.1	198.5	0.270	1.2	80.3
S2	2026	23.9	217.7	0.333	1.8	81.0
S3	2026	20.9	202.3	0.244	1.4	77.8
S4	2026	18.7	217.8	0.248	1.1	79.8
S5	2026	17.8	125.7	0.202	0.8	79.1
S6	2026	18.7	213.1	0.254	1.1	78.1
S7	2026	18.6	285.1	0.221	1.3	86.1
S8	2026	21.2	311.1	0.219	1.3	86.5
S9	2026	18.5	193.9	0.216	1.0	76.3
S10	2026	18.0	298.4	0.320	1.1	79.8
S11	2026	15.1	156.2	0.457	0.3	79.4
S12	2026	19.1	207.2	0.307	0.8	75.9
S13	2026	22.7	229.3	0.382	2.0	79.7
S14	2026	19.8	130.0	0.198	2.4	79.0
S15	2026	18.7	220.7	0.211	2.2	79.5

CHƯƠNG 4. ĐỀ XUẤT GIẢI PHÁP QUẢN LÝ Ô NHIỄM VÀ GIẢM THIỂU RỦI RO SINH THÁI

4.1. Biện pháp quản lý

4.1.1. Kiểm soát nguồn ô nhiễm

Các cơ sở công nghiệp, xí nghiệp và khu dân cư cần lắp đặt hệ thống xử lý nước thải hiện đại trước khi xả thải ra môi trường. Một số công nghệ có thể áp dụng bao gồm keo tụ – lắng, trao đổi ion, hấp phụ bằng than hoạt tính và màng nano để loại bỏ kim loại nặng. Việc kiểm tra định kỳ chất lượng nước thải nên được thực hiện theo chu kỳ 3 đến 6 tháng/lần, nhằm đảm bảo nước thải đầu ra đạt tiêu chuẩn QCVN 40:2011/BTNMT.

Chất thải rắn công nghiệp, đặc biệt là bùn thải, cần được thu gom và xử lý bằng các phương pháp như đốt, chôn lấp an toàn hoặc ổn định hóa bằng xi măng và vôi. Bên cạnh đó, cần tăng cường hoạt động thanh tra và xử phạt nghiêm khắc các hành vi vi phạm. Các cơ sở vi phạm nên bị công khai thông tin và buộc bồi thường thiệt hại môi trường gây ra.

Cần thu gom chất thải phát sinh từ tàu thuyền, đặc biệt là dầu thải và nước dằn tàu. Việc trang bị hệ thống thu gom tại các cảng biển sẽ góp phần hạn chế tình trạng xả trực tiếp ra sông. Khuyến khích chuyển đổi sang sử dụng nhiên liệu sạch, ít lưu huỳnh, đồng thời lắp đặt bộ lọc khí thải để giảm phát tán kim loại nặng.

Thiết lập vùng đệm cấm xả thải cách bờ tối thiểu 500m là cần thiết nhằm ngăn chặn việc xả rác, dầu mỡ và chất thải sinh hoạt ra sông.

Việc sử dụng thức ăn và hóa chất trong nuôi trồng thủy sản cần được kiểm soát chặt chẽ. Nên khuyến cáo người dân sử dụng thức ăn công nghiệp an toàn và hạn chế dùng các hóa chất có chứa CuSO_4 , Pb, As.

Quy hoạch vùng nuôi cần hợp lý, có thể áp dụng các mô hình như nuôi tôm kết hợp rừng ngập mặn hoặc mô hình tuần hoàn sinh thái (biofloc) nhằm giảm thiểu áp lực ô nhiễm môi trường.

Tổ chức tập huấn kỹ thuật cho người dân giúp nâng cao nhận thức và kỹ năng sử dụng hóa chất cũng như xử lý nước ao nuôi hiệu quả hơn.

Giám sát và đánh giá rủi ro

4.1.2. Giám sát và đánh giá rủi ro

Cần tiến hành lấy mẫu nước và trầm tích tại các điểm nóng ít nhất 6 tháng/lần. Việc phân tích nên sử dụng thiết bị hiện đại như ICP-MS để đảm bảo độ chính xác.

Toàn bộ dữ liệu quan trắc nên được tích hợp vào hệ thống GIS, giúp thuận tiện cho việc truy cập, phân tích không gian và theo dõi biến động ô nhiễm theo thời gian.

Đánh giá rủi ro sinh thái có thể thực hiện thông qua các chỉ số như Igeo, PLI, và RI để phân loại mức độ ô nhiễm và khoanh vùng cảnh báo.

4.1.3. Cảnh báo sớm và truyền thông cộng đồng

Cần xây dựng và công bố bản đồ rủi ro ô nhiễm trực tuyến, giúp người dân dễ dàng tiếp cận thông tin môi trường.

Tổ chức các hội thảo, phát tờ rơi nhằm tuyên truyền về tác hại của kim loại nặng, đồng thời hướng dẫn ngư dân cách phòng tránh và phản hồi thông tin nhanh chóng.

4.1.4. Chính sách và quy hoạch vùng

Cần ban hành quy định cấm xả thải chưa qua xử lý vào sông Trường Giang, đồng thời thành lập Ban điều phối bảo vệ môi trường lưu vực sông để điều hành hiệu quả.

Việc quy hoạch không gian sử dụng biển – sông cần có sự phân vùng rõ ràng giữa khu bảo tồn, khu sản xuất và khu dân cư để tránh xung đột chức năng và ô nhiễm chéo.

Cần có chính sách thu hút đầu tư xanh, ưu đãi cho các doanh nghiệp áp dụng công nghệ sạch và xử lý chất thải hiệu quả.

4.2. Biện pháp kỹ thuật

4.2.1. Xử lý ô nhiễm hiện có

Xử lý bằng thực vật (phytoremediation) là giải pháp thân thiện với môi trường. Một số loài có thể được trồng như cỏ vetiver, bèo tây, sậy hoặc rong nhằm hấp thu kim loại nặng từ nước và trầm tích.

Ngoài ra, có thể bổ sung các chủng vi sinh vật như Pseudomonas, Bacillus... giúp chuyển hóa kim loại nặng sang dạng ít độc hoặc không độc.

Có thể cố định kim loại nặng bằng cách sử dụng các vật liệu hấp phụ như biochar, zeolit hoặc apatit, giúp giảm độc tính của chúng trong môi trường.

Nạo vét trầm tích ô nhiễm cũng là một giải pháp hiệu quả, sử dụng công nghệ hút bùn kín nhằm hạn chế phát tán, sau đó tiến hành xử lý bằng phương pháp rửa trôi hoặc

đóng rắn. Cải tạo đáy sông bằng cách phủ một lớp đất sạch hoặc vật liệu cách ly cũng giúp ngăn chặn sự lan truyền của kim loại nặng trong trầm tích.

4.2.2. Công nghệ giảm thiểu ô nhiễm

Hệ thống này có thể được thiết kế dưới dạng bãi lọc sinh học, trồng các loại cây như lau, sậy, dương xỉ để lọc nước thải đầu ra.

Lắp đặt các rào chắn hoặc đập chắn bùn tại các cửa sông và nhánh rạch giúp hạn chế sự lan truyền của trầm tích có chứa kim loại nặng ra khu vực rộng lớn hơn.

4.2.3. Công nghệ tiên tiến (áp dụng khi có ngân sách)

Công nghệ màng lọc nano (nanofiltration) có thể được sử dụng để loại bỏ kim loại nặng trong nước thải tại các nhà máy.

Xử lý điện hóa (electrokinetic remediation) sử dụng dòng điện một chiều để di chuyển các ion kim loại ra khỏi trầm tích, là phương pháp hiện đại và hiệu quả cao.

Có thể tái sử dụng chất thải bằng cách xử lý bùn chứa kim loại nặng, sau đó biến chúng thành vật liệu xây dựng thông qua các công nghệ ổn định hóa.

KẾT LUẬN

Nghiên cứu đã tập trung vào việc ứng dụng các mô hình học máy hiện đại nhằm dự báo mức độ tích lũy và rủi ro sinh thái của một số kim loại nặng (As, Pb, Cd, Hg) trong trầm tích mặt tại khu vực cửa An Hòa, sông Trường Giang, tỉnh Quảng Nam – một khu vực đặc thù chịu ảnh hưởng mạnh từ các hoạt động công nghiệp, giao thông thủy và nuôi trồng thủy sản. Thông qua việc xây dựng và so sánh năm mô hình học máy gồm Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Cubist và Linear Regression, luận văn đã đánh giá hiệu suất của từng mô hình dựa trên các chỉ tiêu thống kê gồm hệ số xác định (R^2), sai số trung bình tuyệt đối (MAE) và căn bậc hai của sai số trung bình bình phương (RMSE).

Kết quả phân tích cho thấy mô hình Random Forest (RF) thể hiện hiệu quả dự báo vượt trội, với giá trị R^2 dao động từ 0,938 đến 0,999 trên tập kiểm tra, MAE trong khoảng 0,015 đến 0,030 và RMSE từ 0,078 đến 0,095. Trên tập huấn luyện, RF cũng duy trì hiệu suất cao với R^2 từ 0,967 đến 0,973, MAE từ 0,016 đến 0,020, và RMSE từ 0,068 đến 0,074. Đây là mô hình có độ chính xác và độ ổn định cao nhất trong số các mô hình được nghiên cứu, nhờ khả năng xử lý hiệu quả các quan hệ phi tuyến và giảm thiểu nhiễu trong dữ liệu môi trường.

Mô hình Cubist cũng đạt hiệu suất tốt, với R^2 từ 0,893 đến 0,945, MAE từ 0,019 đến 0,030 và RMSE từ 0,091 đến 0,130 trên tập kiểm tra. Trên tập huấn luyện, Cubist đạt R^2 từ 0,991 đến 0,994, MAE từ 0,007 đến 0,009, và RMSE từ 0,031 đến 0,038. Với cấu trúc kết hợp giữa cây quyết định và hồi quy tuyến tính, Cubist tỏ ra linh hoạt và hiệu quả trong xử lý dữ liệu phi tuyến. Tuy nhiên, so với RF, mô hình này có độ dao động sai số lớn hơn, đặc biệt là ở tập kiểm tra, cho thấy tính ổn định thấp hơn một chút. Các mô hình còn lại như SVM, KNN và hồi quy tuyến tính cho kết quả không ổn định và có sai số cao, do đó không phù hợp với yêu cầu thực tiễn của bài toán dự báo ô nhiễm kim loại nặng.

Kết quả mô hình hóa không chỉ giúp nhận diện không gian phân bố ô nhiễm mà còn cung cấp cơ sở dữ liệu đầu vào cho các chương trình giám sát môi trường trong tương lai. Mô hình Random Forest (RF) thể hiện tiềm năng nổi bật nhờ khả năng dự báo nhanh, chính xác và ổn định trên nhiều tỷ lệ dữ liệu khác nhau. Đây là công cụ có thể

giúp cơ quan quản lý thực hiện giám sát định kỳ chất lượng trầm tích, đưa ra cảnh báo sớm tại các khu vực có nguy cơ vượt ngưỡng an toàn và ưu tiên xử lý các nguồn ô nhiễm xuất phát từ hoạt động sản xuất công nghiệp hoặc nuôi trồng thủy sản.

Mặc dù đạt được những kết quả tích cực, nghiên cứu vẫn còn tồn tại một số hạn chế. Dữ liệu thu thập trong khuôn khổ luận văn mới chỉ giới hạn tại khu vực cửa An Hòa và số lượng chưa đủ nhiều do đó khi chạy các mô hình kết quả vẫn chưa đạt được sự tối ưu nhất. Ngoài ra, mô hình hiện tại chưa tích hợp các yếu tố thủy văn như dòng chảy, thủy triều hay lượng mưa, vốn có thể ảnh hưởng lớn đến động học và tích lũy của kim loại nặng trong môi trường trầm tích.

Về mặt khoa học, luận văn là một minh chứng rõ ràng cho hiệu quả ứng dụng mô hình Random Forest và Cubist trong dự báo ô nhiễm kim loại nặng tại các hệ sinh thái ven biển – cửa sông, đồng thời xây dựng được một quy trình phân tích và đánh giá có thể áp dụng cho các nghiên cứu tương tự ở các khu vực khác. Về mặt xã hội, kết quả nghiên cứu góp phần quan trọng trong việc hỗ trợ cơ quan quản lý môi trường đưa ra quyết định dựa trên dữ liệu, qua đó bảo vệ hệ sinh thái thủy sinh và sức khỏe cộng đồng. Luận văn đồng thời đưa ra các định hướng và giải pháp bền vững nhằm góp phần cân bằng giữa phát triển kinh tế địa phương và bảo tồn tài nguyên môi trường.

KIẾN NGHỊ

❖ Đối với khu vực nghiên cứu

Kết quả nghiên cứu đã cho thấy khu vực cửa An Hòa đang đối mặt với rủi ro sinh thái ở mức từ trung bình đến cao từ các kim loại như As, Pb, Cd và Hg. Do đó, cần thiết lập hệ thống giám sát môi trường thường xuyên và đồng bộ. Cụ thể, việc triển khai hệ thống quan trắc tự động tại các điểm nóng như khu công nghiệp, cảng biển và vùng nuôi trồng thủy sản sẽ giúp theo dõi định kỳ nồng độ kim loại nặng, cung cấp dữ liệu kịp thời phục vụ công tác quản lý.

Ngoài ra, cần xây dựng bản đồ rủi ro sinh thái dựa trên kết quả của mô hình học máy để xác định những khu vực ưu tiên xử lý. Các biện pháp xử lý tại chỗ như trồng cỏ Vetiver hoặc sử dụng bèo tây – những loài có khả năng hấp thụ kim loại nặng – nên được triển khai nhằm giảm thiểu sự tích tụ trong trầm tích.

Về mặt quản lý nguồn phát thải, cần siết chặt các quy định về xả thải đối với các nhà máy công nghiệp, yêu cầu xử lý nước thải đạt chuẩn trước khi xả ra sông. Đồng thời, cần có sự điều chỉnh trong quy hoạch vùng nuôi trồng thủy sản theo hướng bền vững, giảm sử dụng hóa chất và thức ăn công nghiệp chứa kim loại nặng. Các hoạt động giao thông thủy cũng cần được giám sát chặt chẽ hơn, đặc biệt là việc xả dầu thải và rác từ tàu thuyền.

Bên cạnh đó, công tác nâng cao nhận thức cộng đồng là yếu tố then chốt. Việc tổ chức các chương trình tập huấn cho ngư dân và doanh nghiệp địa phương sẽ giúp họ hiểu rõ hơn về tác hại của kim loại nặng và các biện pháp phòng tránh. Dữ liệu ô nhiễm cũng nên được công khai trên các nền tảng trực tuyến để người dân có thể chủ động theo dõi, giám sát và cùng tham gia vào quá trình bảo vệ môi trường.

❖ Đối với mô hình học máy

Nghiên cứu đã chứng minh hiệu quả của mô hình Random Forest và Cubist trong việc dự báo mức độ tích lũy kim loại nặng và đánh giá rủi ro sinh thái trong trầm tích. Để cải thiện độ chính xác và ổn định, cần tăng số lượng của bộ dữ liệu và tích hợp thêm các biến môi trường như lưu lượng dòng chảy, chế độ thủy triều và độ mặn – những yếu tố có thể ảnh hưởng lớn đến quá trình phát tán và tích lũy kim loại trong môi trường nước. Ngoài ra, nên xem xét thử nghiệm các mô hình lai (hybrid models) kết hợp Cubist

với các thuật toán học sâu như LSTM (Long Short-Term Memory) hoặc GRU (Gated Recurrent Unit) nhằm nâng cao khả năng dự báo xu hướng tích lũy dài hạn.

Việc hiệu chỉnh và tối ưu hóa tham số cho Random Forest và Cubist cũng cần được thực hiện nhằm giảm thiểu sai số trên tập kiểm tra và tăng tính tổng quát của mô hình. Đồng thời, để đánh giá tính ứng dụng rộng hơn, các mô hình này nên được triển khai tại các khu vực cửa sông khác trong lưu vực Trường Giang và các vùng chịu tác động mạnh từ con người.

Để mô hình dự báo trở nên dễ hiểu và đáng tin cậy hơn, cần áp dụng các phương pháp phân tích giúp xác định rõ yếu tố nào trong dữ liệu có ảnh hưởng lớn đến kết quả dự báo. Việc này sẽ giúp chúng ta hiểu rõ nguyên nhân gây ô nhiễm và đưa ra giải pháp kiểm soát hiệu quả hơn. Bên cạnh đó, cũng nên sử dụng các phương pháp kiểm tra tính chắc chắn của mô hình, nhằm đảm bảo kết quả dự báo vẫn đáng tin cậy trong điều kiện môi trường thay đổi phức tạp, như ảnh hưởng của thời tiết, thủy triều hay hoạt động con người.

TÀI LIỆU THAM KHẢO

- [1] Cổng thông tin điện tử thành phố Tam Kỳ. (2019). Điều kiện tự nhiên.
- [2] Cổng thông tin điện tử tỉnh Quảng Nam. (2020). Phát huy nội lực- yếu tố quan trọng để Kinh tế - xã hội huyện Núi Thành giai đoạn 2016-2020 phát triển.
- [3] Ban quản lý các khu kinh tế và khu công nghiệp tỉnh Quảng Nam. (2022). Tình hình bảo vệ môi trường trên địa bàn KKTM Chu Lai và các KCN trên địa bàn tỉnh Quảng Nam.
- [4] Sở Tài nguyên và Môi trường, "Báo cáo hiện trạng Môi trường tỉnh Quảng Nam giai đoạn 2016 -2020 ".
- [5] Available: <https://machinelearningcoban.com/2016/12/27/categories/>.
- [6] Nguyễn Đức Phong and Hà Hải Dương, "Nghiên cứu ứng dụng các mô hình học máy để dự báo chỉ số chất lượng nước mặt vùng bán đảo Cà Mau."
- [7] Ngô Viết Thắng Lê Phước Cường, "Ứng dụng mô hình học máy dự báo chất lượng nước dưới đất: Điển hình tại khu vực thành phố Hội An, tỉnh Quảng Nam," Tạp chí Khoa học và Công nghệ-Đại học Đà Nẵng, pp. 106-110, 2022.
- [8] Bộ Lao động – Thương binh và Xã hội, Thông tư số 03/2025/TT-BLĐTBXH: Quy định tiêu chuẩn phân loại lao động theo điều kiện lao động, Hà Nội, 2025.
- [9] TCVN 6663-15:2004 (ISO 5667-15:1999) – Chất lượng nước – Lấy mẫu – Phần 15: Hướng dẫn bảo quản và xử lý mẫu bùn và trầm tích, 2004.
- [10] Xuan Cuong Nguyen et al., "Developing a new approach for design support of subsurface constructed wetland using machine learning algorithms," vol. 301, p. 113868, 2022.
- [11] Hoang Nguyen, Nam Xuan Bui, Hieu Quang Tran, and Giang Huong Thi Le, "A novel soft computing model for predicting blast-induced ground vibration in open-pit mines using gene expression programming," Journal of Mining Earth Sciences, vol. 61, no. 5, pp. 107-116, 2020.
- [12] Vũ Hữu Tiệp, Machine learning cơ bản. Nhà xuất bản Khoa học và Kỹ thuật, 2018.
- [13] Bộ Tài nguyên và Môi trường, QCVN 08:2023/BTNMT - Quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt, Hà Nội: Nhà xuất bản Tài nguyên - Môi trường và Bản đồ Việt Nam, 2023.

- [14] Bộ Khoa học và Công nghệ, TCVN 6663-1:2011 (ISO 5667-2:2006) - Chất lượng nước – Lấy mẫu – Phần 1: Hướng dẫn kỹ thuật lấy mẫu, Hà Nội: NXB Khoa học và Kỹ thuật, 2011.
- [15] Bộ Khoa học và Công nghệ, TCVN 6663-6:2018 - Chất lượng nước – Lấy mẫu – Phần 6: Hướng dẫn lấy mẫu nước sông và suối, Hà Nội: NXB Khoa học và Kỹ thuật, 2018.
- [16] Bộ Khoa học và Công nghệ, TCVN 6663-3:2003 (ISO 5667-3:1985) - Chất lượng nước – Lấy mẫu – Phần 3: Hướng dẫn bảo quản và xử lý mẫu, Hà Nội: NXB Khoa học và Kỹ thuật, 2003.
- [17] Bộ Khoa học và Công nghệ, TCVN 6663-13:2000 (ISO 5667-13:1997) - Chất lượng nước – Lấy mẫu – Phần 13: Hướng dẫn lấy mẫu bùn nước, bùn nước thải và bùn liên quan, Hà Nội: NXB Khoa học và Kỹ thuật, 2000.
- [18] Bộ Khoa học và Công nghệ, TCVN 6663-15:2004 (ISO 5667-15:1999) - Chất lượng nước – Lấy mẫu – Phần 15: Hướng dẫn bảo quản và xử lý mẫu bùn và trầm tích, Hà Nội: NXB Khoa học và Kỹ thuật, 2004.
- [19] Bộ Tài nguyên và Môi trường, QCVN 43:2017/BTNMT - Quy chuẩn kỹ thuật quốc gia về trầm tích, Hà Nội: NXB Tài nguyên - Môi trường và Bản đồ Việt Nam, 2017.
- [20] Bộ Khoa học và Công nghệ, TCVN 6179-1:1996 - Chất lượng nước – Xác định coliforms tổng số – Phần 1: Phương pháp xác định số có xác suất cao nhất (MPN), Hà Nội: NXB Khoa học và Kỹ thuật, 1996.
- [21] Bộ Khoa học và Công nghệ, TCVN 6194:1996 - Chất lượng nước – Xác định hàm lượng nitrat (NO_3^-), Hà Nội: NXB Khoa học và Kỹ thuật, 1996.
- [22] Bộ Khoa học và Công nghệ, TCVN 6625:2000 - Chất lượng nước – Xác định Amoni bằng phương pháp trắc quang, Hà Nội: NXB Khoa học và Kỹ thuật, 2000.
- [23] Bộ Khoa học và Công nghệ, TCVN 6178:1996 - Chất lượng nước – Xác định tổng chất rắn lơ lửng, Hà Nội: NXB Khoa học và Kỹ thuật, 1996.
- [24] Bộ Khoa học và Công nghệ, TCVN 6491:1999 - Chất lượng nước – Xác định hàm lượng photphat, Hà Nội: NXB Khoa học và Kỹ thuật, 1999.
- [25] Bộ Khoa học và Công nghệ, TCVN 6001-1:2008 - Chất lượng nước – Xác định nhu cầu oxy sinh hóa (BOD) – Phần 1: Phương pháp pha loãng và cấy, Hà Nội: NXB Khoa học và Kỹ thuật, 2008.

[26] Bộ Tài nguyên và Môi trường, QCVN 43:2017/BTNMT – Quy chuẩn kỹ thuật quốc gia về chất lượng trầm tích, Hà Nội: Bộ Tài nguyên và Môi trường, 2017.

PHỤ LỤC 1



PHIẾU KHẢO SÁT

**Chất lượng môi trường, tình hình sức khỏe người dân tại
cửa An Hòa, sông Trường Giang, Quảng Nam**

*Những thông tin mà quý vị cung cấp nhằm phục vụ cho đề tài nghiên cứu của sinh viên
Khoa Môi trường - Trường Đại học Bách khoa, Đại học Đà Nẵng,
mọi thông tin sẽ đảm bảo được hoàn toàn giữ bí mật.*

A. Thông tin người được phỏng vấn

Họ và tên:.....Nam/Nữ

Tuổi:.....

Địa chỉ:.....

Điện thoại:.....

Ngày phỏng vấn:.....

1. Tình trạng hôn nhân hiện tại:

A. Độc thân

B. Kết hôn

C. Khác

2. Dân tộc:

A. Kinh

B. Dân tộc khác

3. Nghề nghiệp hiện tại:

A. Nông dân

B. Công nhân

C. Cán bộ viên chức

D. Khác

4. Thu nhập bình quân gia đình:

A. Dưới 1 triệu đồng/tháng

B. Từ 1-2 triệu đồng/tháng

C. Từ 2-5 triệu đồng/tháng

D. Trên 5 triệu đồng/tháng

Hoạt động kinh tế của gia đình ông (bà) là gì? Sự đóng góp này quan trọng thế nào? Nguồn thu nhập	Mức độ quan trọng của sự đóng góp			
	Rất quan trọng	Quan trọng	Không quan trọng lắm	Không quan trọng chút nào
Trồng lúa nước				
Trồng rừng				
Trồng rau/cây ăn trái				
Chăn nuôi gia súc				
Chăn nuôi gia cầm				
Nuôi trồng và đánh bắt thủy hải sản				
Chế biến lâm sản				
Chế biến nước mắm				
Cơ khí				
Xây dựng				
Gia công mộc				
Buôn bán				
Làm thuê				
Khác (ghi rõ)				

Ghi chú mã số mức độ quan trọng:

1. Rất quan trọng

2. Quan trọng

3. Không quan trọng lắm

4. Không quan trọng chút nào

6. Thời gian sống tại khu vực này

A. Chưa được 1 năm

B. Ít hơn 5 năm

C. Hơn 5 năm

B. Đánh giá hiện trạng môi trường và rủi ro sinh thái

I. Môi trường không khí

1. Theo Ông (Bà) môi trường không khí hiện nay tại khu vực là như thế nào?

A. Bình thường

B. Ô nhiễm ít

C. Ô nhiễm

D. Ô nhiễm nghiêm trọng

2. Mô tả chi tiết nguyên nhân gây ô nhiễm:

Nguồn gây ô nhiễm	Các hoạt động gây ô nhiễm	Mức độ ảnh hưởng	Thời gian ảnh hưởng
Hoạt động giao thông thủy			
Hoạt động nuôi trồng thủy sản			
Sản xuất công nghiệp gần khu vực			
Hoạt động sinh hoạt dân cư			
Nguồn khác (nếu có)			

Ghi chú mã số mức độ ảnh hưởng:

1. Không đáng kể
2. Trung bình
3. Khá nghiêm trọng
4. Nghiêm trọng
5. Rất nghiêm trọng

Ghi chú mã số thời gian ảnh hưởng:

1. Theo mùa
2. Theo ngày
3. Tùy từng thời điểm

3. Theo ông/bà rừng nói chung và cây xanh nói riêng có giúp giảm ô nhiễm không khí ở khu vực này không ?

A. Có

B. Không

4. Theo ông/bà hiện trạng tài nguyên rừng của địa phương hiện nay so với 5 năm trước như thế nào?

A. Tốt hơn trước

B. Kém hơn trước

C. Không biết

II. Chất lượng môi trường nước và trầm tích

1. Theo Ông (Bà), chất lượng nước và trầm tích tại khu vực Cửa An Hòa, sông Trường Giang hiện nay như thế nào?

- A. Bình thường
B. Ô nhiễm ít
C. Ô nhiễm
D. Ô nhiễm nghiêm trọng

2. Mô tả chi tiết nguyên nhân gây ô nhiễm:

Nguồn gây ô nhiễm	Các hoạt động gây ô nhiễm	Mức độ ảnh hưởng	Thời gian ảnh hưởng
Hoạt động nuôi trồng thủy sản			
Hoạt động khai thác cát, nạo vét sông			
Nước thải từ khu dân cư, nhà máy			
Nguồn khác (ghi rõ):			

- Ghi chú mã số mức độ ảnh hưởng:**
1. Không đáng kể
 2. Trung bình
 3. Khá nghiêm trọng
 4. Nghiêm trọng
 5. Rất nghiêm trọng

- Ghi chú mã số thời gian ảnh hưởng:**
1. Theo mùa
 2. Theo ngày
 3. Tùy từng thời điểm

3. Ông (Bà) có sử dụng nước giếng khoan phục vụ sinh hoạt hàng ngày không?

A. Có
B. Không

Hãy giải thích lý do

.....

.....

.....

4. Nguồn nước phục vụ nông nghiệp của gia đình lấy từ đâu?

.....
.....
Nguồn nước đó có ổn định hay không?

A. Có

B. Không

III. Môi trường đất

1. Theo Ông (Bà) môi trường đất hiện nay tại khu vực là như thế nào?

A. Bình thường

B. Ô nhiễm ít

C. Ô nhiễm

D. Ô nhiễm nghiêm trọng

2. Mô tả chi tiết nguyên nhân gây ô nhiễm

Nguồn gây ô nhiễm	Các hoạt động gây ô nhiễm	Mức độ ảnh hưởng	Thời gian ảnh hưởng
Hoạt động nuôi trồng thủy sản			
Xả rác sinh hoạt & công nghiệp			
Hoạt động khai thác cát & phù sa			
Nguồn khác (nếu có)			

Ghi chú mã số mức độ ảnh hưởng:

1. Không đáng kể

2. Trung bình

3. Khá nghiêm trọng

4. Nghiêm trọng

5. Rất nghiêm trọng

Ghi chú mã số thời gian ảnh hưởng:

1. Theo mùa

2. Theo ngày

3. Tùy từng thời điểm

3. Ông (Bà) có sử dụng đất cho mục đích trồng trọt hay không?

A. Có

B. Không

4. Chất lượng cây trồng hiện nay so với trước đây như thế nào?

.....
.....
.....

IV. Ảnh hưởng đến sức khỏe cộng đồng

1. Trong 1 năm qua, Ông/Bà hoặc người thân có mắc các bệnh nào sau đây không?

- Bệnh về da (mụn nhọt, viêm da...)
- Bệnh đường hô hấp (viêm phổi, ho kéo dài...)
- Bệnh đường tiêu hóa (tiêu chảy, rối loạn tiêu hóa...)
- Bệnh gan, thận
- Không có bệnh đáng kể

2. Ông/Bà có cho rằng các vấn đề sức khỏe trên có liên quan đến ô nhiễm môi trường không?

- Có
- Không
- Không rõ

3. Ông/Bà có thường xuyên tập thể dục hoặc tham gia hoạt động thể chất không?

- Hàng ngày
- 2-3 lần/tuần
- Ít hơn 1 lần/tuần
- Hầu như không bao giờ

4. Số lần khám sức khỏe định kỳ của ông/bà?.....

5. Theo Ông (Bà) thì chất lượng dịch vụ y tế tuyến Phường (Xã) tại địa phương là như thế nào?

- A. Tốt
- B. Khá
- C. Trung bình
- D. Kém

6. Ông/Bà có gặp khó khăn trong việc tiếp cận các dịch vụ y tế khi cần không?

- Không gặp khó khăn
- Khó khăn về tài chính
- Khó khăn do khoảng cách, giao thông
- Khó khăn do quá tải tại cơ sở y tế

V. Ý kiến cộng đồng về giải pháp bảo vệ môi trường và sức khỏe

1. Ông/Bà có tham gia vào các hoạt động bảo vệ môi trường không?
 A. Có B. Không
2. Nếu có, đó là hoạt động gì?
 Dọn dẹp vệ sinh bờ sông
 Tuyên truyền nâng cao nhận thức
 Đóng góp ý kiến về quản lý môi trường
 Hoạt động khác (Ghi rõ):
3. Theo Ông/Bà, chính quyền địa phương cần thực hiện biện pháp nào để giảm thiểu ô nhiễm nước và trầm tích?
 Kiểm soát chặt chẽ nguồn xả thải từ doanh nghiệp
 Cải thiện hệ thống xử lý nước thải sinh hoạt
 Tăng cường giám sát môi trường
 Trồng thêm rừng ngập mặn
 Hỗ trợ người dân chuyển đổi sinh kế ít ảnh hưởng đến môi trường
 Ý kiến khác:
4. Theo Ông/Bà thì doanh nghiệp tại khu vực cửa An Hòa, sông Trường Giang có phải chịu trách nhiệm chi trả các thiệt hại về môi trường và sức khỏe cho người dân tại địa phương không?
 A. Có B. Không
5. Ông/bà có đề xuất, kiến nghị gì với trung tâm y tế cấp Phường (xã) để nâng cao chất lượng dịch vụ chăm sóc sức khỏe cộng đồng?

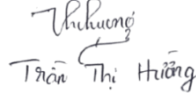
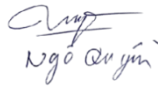



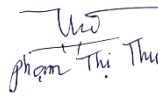
Xin chân thành cảm ơn sự giúp đỡ của quý vị!



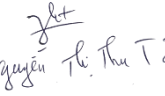

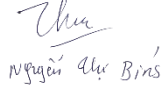
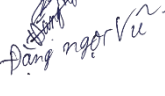
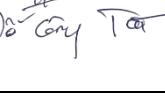
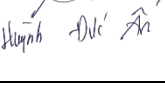
Người điều tra

Ngày ... tháng ... năm 2025



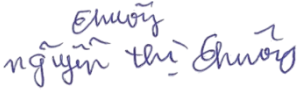

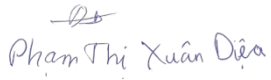

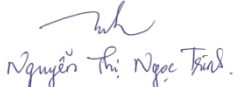
Người cung cấp thông tin


PHỤ LỤC 2

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
1	Trần Thị Hương	Nữ	32	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0356585978		
2	Ngô Quyền	Nam	68	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0894525217		
3	Lê Quang Trang	Nam	32	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0362823243		
4	Nguyễn Thị Huệ	Nữ	27	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0344800879		
5	Trần Thị Cúc	Nữ	30	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0943472927		
6	Phạm Thị Thu	Nữ	70	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	-		

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
7	Ngô Hữu Thạnh	Nam	50	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0352396269		
8	Phạm Thảo	Nam	50	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0326890122		
9	Nguyễn Thị Thu Tâm	Nữ	43	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0326240568		
10	Phan Thị Tiến	Nữ	42	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0357809939		
11	Nguyễn Thị Bích	Nữ	40	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0989461968		
12	Đặng Ngọc Vũ	Nam	17	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0327272849		
13	Đỗ Công Tài	Nam	30	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0347610476		
14	Huỳnh Đức Ân	Nam	33	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0917188535		

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
15	Phạm Văn Tuấn	Nam	53	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0974353927		
16	Đình Công Quốc	Nam	29	Đông Thạnh, Tam Hòa, Núi Thành, Quảng Nam	0869003852		
17	Hồ Thị Kim Vân	Nữ	22	KCN Cảng Tam Hiệp, Núi Thành, Quảng Nam	0865671603		
18	Phạm Viết Cảnh	Nam	29	KCN Cảng Tam Hiệp, Núi Thành, Quảng Nam	0387603544		
19	Nguyễn Thị Thùy	Nữ	45	Đông Tuấn, Tam Hải, Núi Thành, Quảng Nam	-		
20	Phạm Thị Lan	Nữ	50	Đông Tuấn, Tam Hải, Núi Thành, Quảng Nam	-		
21	Huỳnh Thị Ngọc Liễu	Nữ	55	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0987165484		
22	Trần Thị Tới	Nữ	69	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0388447978		

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
23	Nguyễn Minh Thuận	Nam	45	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0935617606		
24	Lê Văn Huy	Nam	62	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0913418730		
25	Nguyễn Thị Chuông	Nữ	61	Thôn 7, Tam Hải, Núi Thành, Quảng Nam	-		
26	Bùi Văn Thành	Nam	58	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0333011965		
27	Phạm Thị Xuân Diệu	Nữ	18	Thôn Long Thạch Đông, Tam Hải, Núi Thành, Quảng Nam	0799312395		
28	Phạm Thùy Tiên	Nữ	17	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0987647960		
29	Nguyễn Thị Thi	Nữ	40	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	-		
30	Nguyễn Thị Ngọc Trinh	Nữ	29	Thôn Đông Tuấn, xã Tam Hải, Núi Thành, Quảng Nam	0348808695		

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
31	Bùi Thị Thôi	Nữ	31	Thôn Đông Tuần, xã Tam Hải, Núi Thành, Quảng Nam	0984241288		
32	Hồ Thanh Thi	Nữ	49	Thôn Đông Tuần, xã Tam Hải, Núi Thành, Quảng Nam	0704145939		
33	Huỳnh Văn Vương	Nam	18	Thôn Long Thạnh Đông, xã Tam Hải, Núi Thành, Quảng Nam	0389567313		
34	Lê Thanh Hiếu	Nam	18	Thôn Tân Lập, xã Tam Hải, Núi Thành, Quảng Nam	0342869754		
35	Nguyễn Thị Lộc	Nữ	60	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-		
36	Vũ Ngọc Bình	Nữ	35	Thôn Đông Tuần, xã Tam Giang, Núi Thành, Quảng Nam	-		
37	Dương Thị Thảo	Nữ	58	Thôn Tiên Xuân 1, xã Tam Anh Nam, Núi Thành, Quảng Nam	0944513042		
38	Lê Thị Minh Tuyết	Nữ	38	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-		

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
39	Nguyễn Văn Hưng	Nam	45	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-	 Nguyễn Văn Hưng	
40	Võ Minh Quân	Nam	33	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-	 Võ Minh Quân	
41	Lê Văn Ngọc	Nam	38	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	0987377641	 Lê Văn Ngọc	
42	Phạm Văn Long	Nam	50	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-	 Phạm Văn Long	
43	Trần Thị Kim Ngân	Nữ	27	Thôn Đông Xuân, xã Tam Giang, Núi Thành, Quảng Nam	-	 Trần Thị Kim Ngân	
44	Lê Thị Như Quỳnh	Nữ	27	Thôn An Hải, xã Tam Quang, Núi Thành, Quảng Nam	0368547592	 Lê Thị Như Quỳnh	
45	Bùi Văn Tịnh	Nam	22	Tam Anh Nam, Núi Thành, Quảng Nam	0384980914	 Bùi Văn Tịnh	
46	Đỗ Lương Trọng	Nam	23	Tam Giang, Núi Thành, Quảng Nam	0336595983	 Đỗ Lương Trọng	

STT	Họ và tên	Giới tính	Tuổi	Địa chỉ	Số điện thoại	Chữ ký	Ghi chú
47	Nguyễn Thị Quỳnh Hoa	Nữ	29	thôn An Hải, xã Tam Quang, Núi Thành, Quảng Nam	-	 Nguyễn Thị Quỳnh Hoa	
48	Võ Thị Thanh Trúc	Nữ	32	Thôn An Hải, xã Tam Quang, Núi Thành, Quảng Nam	-	 Võ Thị Thanh Trúc	
49	Đinh Văn Toàn	Nam	48	Thôn An Hải, xã Tam Quang, Núi Thành, Quảng Nam	-	 Đinh Văn Toàn	
50	Huỳnh Văn Đạt	Nam	38	Thôn An Hải, xã Tam Quang, Núi Thành, Quảng Nam	-	 Huỳnh Văn Đạt	

PHỤ LỤC 3

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
1	16.7	110.1	0.343	1.099	61.5	107.2	Rủi ro thấp
2	25.3	239.2	0.299	1.599	88.3	130.3	Rủi ro thấp
3	21.9	214.5	0.286	1.056	93.1	107.9	Rủi ro thấp
4	20.9	241.1	0.293	1.024	83.4	109.2	Rủi ro thấp
5	20.1	77.3	0.229	1.002	74.1	86.0	Rủi ro thấp
6	20.7	234.6	0.263	1.309	82.9	112.4	Rủi ro thấp
7	20	338.6	0.282	1.429	96.3	126.1	Rủi ro thấp
8	23.6	365.1	0.28	1.299	92.3	126.1	Rủi ro thấp
9	22.1	192.3	0.246	1.067	83.9	100.3	Rủi ro thấp
10	19.8	356.2	0.146	1.159	95.7	97.3	Rủi ro thấp
11	6.2	122.6	0.18	0.089	82.1	44.8	Rủi ro thấp
12	16.7	226.6	0.199	0.125	72.8	63.3	Rủi ro thấp
13	17.3	253.1	0.197	1.859	81.6	117.4	Rủi ro thấp
14	16.6	80.3	0.163	2.589	95	121.1	Rủi ro thấp
15	20.2	244.2	0.191	2.859	91	147.8	Rủi ro thấp
16	8.2	96.4	0.767	1.021	28.4	165.9	Rủi ro trung bình
17	16.8	225.5	0.645	1.521	55.2	176.5	Rủi ro trung bình
18	13.4	200.8	0.238	0.978	60	91.0	Rủi ro thấp
19	12.4	227.4	0.243	0.946	50.3	92.1	Rủi ro thấp
20	11.6	63.6	0.158	0.924	41	65.5	Rủi ro thấp
21	12.2	220.9	0.338	1.231	49.8	115.2	Rủi ro thấp
22	11.5	324.9	0.123	1.351	90.1	91.6	Rủi ro thấp
23	15.1	351.4	0.139	1.221	80.4	94.5	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
24	13.6	178.6	0.154	0.989	71.1	76.5	Rủi ro thấp
25	11.3	342.5	0.823	1.081	79.9	196.6	Rủi ro trung bình
26	6.7	108.9	1.377	0.011	93.3	233.4	Rủi ro trung bình
27	8.2	212.9	0.094	0.047	89.3	37.6	Rủi ro thấp
28	8.8	239.4	0.099	1.781	80.9	92.7	Rủi ro thấp
29	8.1	66.6	0.014	2.511	92.7	88.3	Rủi ro thấp
30	11.7	230.5	0.194	2.781	23.8	138.9	Rủi ro thấp
31	25.1	137.4	0.337	0.957	69.4	109.6	Rủi ro thấp
32	59.2	54.1	0.293	1.421	96.2	133.4	Rủi ro thấp
33	9.2	102.6	0.28	0.914	101	86.3	Rủi ro thấp
34	2.3	82.4	0.287	0.882	91.3	80.3	Rủi ro thấp
35	4.3	190.8	0.223	0.86	82	78.4	Rủi ro thấp
36	4.7	130.9	0.257	1.167	90.8	89.1	Rủi ro thấp
37	6.6	2.4	0.276	1.222	104.2	86.0	Rủi ro thấp
38	22.6	300.7	0.274	1.152	100.2	115.5	Rủi ro thấp
39	3.9	110.2	0.24	0.925	91.8	77.1	Rủi ro thấp
40	4.3	152.7	0.14	1.017	103.6	67.3	Rủi ro thấp
41	24	187.2	0.174	0.053	90	59.3	Rủi ro thấp
42	22	142.3	0.193	0.011	80.7	56.5	Rủi ro thấp
43	56.1	59	0.191	1.716	89.5	124.2	Rủi ro thấp
44	6.1	107.5	0.157	2.442	102.9	110.7	Rủi ro thấp
45	9.4	87.3	0.185	2.711	98.9	124.0	Rủi ro thấp
46	42.6	82.8	0.866	1.975	60.2	232.5	Rủi ro trung bình
47	47.2	88.2	0.989	2.702	97.9	277.6	Rủi ro trung bình
48	53.2	48	0.137	2.979	30.4	150.4	Rủi ro trung bình
49	44.7	27.8	0.142	1.975	11.6	113.8	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
50	40.4	136.2	0.057	1.038	88.4	77.4	Rủi ro thấp
51	46.9	76.3	0.237	1.221	48.9	111.5	Rủi ro thấp
52	46.3	52.2	0.022	1.618	20.6	86.8	Rủi ro thấp
53	52.2	246.1	0.038	2.805	47.9	142.9	Rủi ro thấp
54	45	55.6	0.053	1.274	0.7	80.7	Rủi ro thấp
55	41.9	98.1	0.112	2.035	36.2	114.1	Rủi ro thấp
56	15	132.6	0.018	0.068	79.4	24.8	Rủi ro thấp
57	56.5	87.7	0.863	1.936	42.9	240.3	Rủi ro trung bình
58	61.1	38.2	1.417	2.672	73.9	350.8	Rủi ro cao
59	67.1	52.9	0.134	2.942	43.8	158.5	Rủi ro trung bình
60	58.6	32.7	0.139	1.942	85.2	122.4	Rủi ro thấp
61	15.4	97.5	0.267	1.263	49.4	98.1	Rủi ro thấp
62	42.6	102.1	0.223	1.991	87.1	131.6	Rủi ro thấp
63	28.5	108.1	0.21	2.266	19.6	128.4	Rủi ro thấp
64	27.1	99.6	0.217	1.268	69.3	98.3	Rủi ro thấp
65	23.4	95.3	0.153	0.326	77.6	57.1	Rủi ro thấp
66	15.6	101.8	0.187	0.543	38.1	64.1	Rủi ro thấp
67	17.6	101.2	0.206	0.903	9.8	79.1	Rủi ro thấp
68	22	107.1	0.204	2.092	37.1	117.9	Rủi ro thấp
69	19.8	99.9	0.17	0.568	19.5	64.7	Rủi ro thấp
70	24.4	96.8	0.769	1.323	25.4	186.1	Rủi ro trung bình
71	23	56	1.323	0.678	68.6	251.7	Rủi ro trung bình
72	30	75.2	0.04	1.215	32.1	68.4	Rủi ro thấp
73	64.1	79.8	0.045	1.941	63.1	114.2	Rủi ro thấp
74	14.1	85.8	0.17	2.232	33	109.9	Rủi ro thấp
75	7.2	77.3	0.14	1.136	59.8	67.1	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
76	24.8	98.9	0.28	1.866	64.4	124.7	Rủi ro thấp
77	29.4	103.5	0.236	2.136	70.4	129.2	Rủi ro thấp
78	35.4	109.5	0.223	1.136	61.9	101.5	Rủi ro thấp
79	26.9	101	0.23	0.199	57.6	68.2	Rủi ro thấp
80	22.6	96.7	0.166	0.416	64.1	61.4	Rủi ro thấp
81	29.1	103.2	0.2	0.775	63.5	82.4	Rủi ro thấp
82	28.5	102.6	0.219	1.962	69.4	120.6	Rủi ro thấp
83	34.4	108.5	0.217	0.436	62.2	78.8	Rủi ro thấp
84	27.2	101.3	0.183	1.182	59.1	90.4	Rủi ro thấp
85	24.1	98.2	0.782	0.771	17	171.4	Rủi ro trung bình
86	18	56.1	1.336	1.073	45.7	262.2	Rủi ro trung bình
87	10.7	84.8	0.053	1.803	54.3	76.1	Rủi ro thấp
88	19.3	93.4	0.058	2.073	50.9	91.3	Rủi ro thấp
89	15.9	90	0.183	1.045	49.9	77.9	Rủi ro thấp
90	14.9	89	0.153	0.852	111	67.0	Rủi ro thấp
91	11.7	54.8	0.367	1.608	107.6	119.3	Rủi ro thấp
92	20.3	51.4	0.323	1.871	106.6	125.6	Rủi ro thấp
93	16.9	50.4	0.31	0.834	105.8	90.1	Rủi ro thấp
94	15.9	49.6	0.317	0.698	106.4	86.4	Rủi ro thấp
95	15.1	50.2	0.253	0.152	105.7	59.3	Rủi ro thấp
96	15.7	49.5	0.287	0.517	109.3	76.1	Rủi ro thấp
97	15	53.1	0.306	1.708	107.8	114.6	Rủi ro thấp
98	18.6	51.6	0.304	0.178	105.5	70.7	Rủi ro thấp
99	17.1	49.3	0.27	0.938	91.9	86.8	Rủi ro thấp
100	14.8	35.7	0.869	0.508	133.4	167.5	Rủi ro trung bình
101	19	77.2	1.423	0.815	117.9	271.0	Rủi ro trung bình

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
102	13.9	61.7	0.14	1.543	116.9	83.0	Rủi ro thấp
103	10.5	60.7	0.145	1.811	116.1	89.5	Rủi ro thấp
104	9.5	59.9	0.27	0.823	116.7	79.2	Rủi ro thấp
105	8.7	60.5	0.24	1.906	53.5	106.0	Rủi ro thấp
106	12.9	125.5	0.418	2.627	91.2	163.8	Rủi ro trung bình
107	9.5	42.2	0.541	2.903	23.7	183.1	Rủi ro trung bình
108	8.5	90.7	0.128	1.906	4.9	89.8	Rủi ro thấp
109	7.7	70.5	0.133	0.969	81.7	61.0	Rủi ro thấp
110	8.3	178.9	0.048	1.186	42.2	61.8	Rủi ro thấp
111	7.6	119	0.228	1.546	13.9	96.5	Rủi ro thấp
112	11.2	88	0.013	2.733	41.2	98.1	Rủi ro thấp
113	9.7	288.8	0.029	1.211	67.1	68.4	Rủi ro thấp
114	7.4	98.3	0.044	1.989	29.5	78.8	Rủi ro thấp
115	6.2	140.8	0.103	0.056	72.7	32.8	Rủi ro thấp
116	18	175.3	0.182	1.592	36.2	101.6	Rủi ro thấp
117	19.8	130.4	0.854	2.313	67.2	228.9	Rủi ro trung bình
118	18.8	47.1	1.408	2.592	37.1	319.1	Rủi ro cao
119	18	95.6	0.125	1.556	78.5	86.0	Rủi ro thấp
120	18.6	75.4	0.13	0.753	106	61.8	Rủi ro thấp
121	13	104.9	0.413	1.59	111.4	130.6	Rủi ro thấp
122	21.6	110.3	0.536	2.32	71.2	178.0	Rủi ro trung bình
123	18.2	70.1	0.123	2.59	51	114.8	Rủi ro thấp
124	17.2	49.9	0.128	1.59	159.4	84.1	Rủi ro thấp
125	16.4	158.3	0.043	0.653	99.5	49.3	Rủi ro thấp
126	17	98.4	0.223	0.87	75.4	80.6	Rủi ro thấp
127	16.3	74.3	0.008	1.23	269.3	55.9	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
128	19.9	268.2	0.024	2.42	78.8	109.3	Rủi ro thấp
129	18.4	77.7	0.039	0.89	121.3	51.4	Rủi ro thấp
130	16.1	120.2	0.098	1.65	155.8	85.4	Rủi ro thấp
131	15.9	154.7	0.177	0.36	110.9	61.4	Rủi ro thấp
132	16.2	109.8	0.849	1.63	61.4	203.7	Rủi ro trung bình
133	20.8	60.3	1.403	2.36	76.1	313.9	Rủi ro cao
134	26.8	75	0.12	2.63	55.9	121.6	Rủi ro thấp
135	18.3	54.8	0.125	1.63	88.3	85.5	Rủi ro thấp
136	13.1	70.4	0.309	1.68	92.9	114.1	Rủi ro thấp
137	21.7	75	0.265	2.41	98.9	135.1	Rủi ro thấp
138	18.3	81	0.252	2.68	90.4	139.2	Rủi ro thấp
139	17.3	72.5	0.259	1.61	86.1	106.9	Rủi ro thấp
140	16.5	68.2	0.195	0.743	92.6	69.9	Rủi ro thấp
141	17.1	74.7	0.229	0.96	92	82.7	Rủi ro thấp
142	16.4	74.1	0.248	1.32	97.9	96.1	Rủi ro thấp
143	20	80	0.246	2.51	90.7	134.2	Rủi ro thấp
144	18.5	72.8	0.212	0.98	87.6	81.4	Rủi ro thấp
145	16.2	69.7	0.112	1.74	51.7	86.2	Rủi ro thấp
146	6.7	33.8	0.146	0.11	70.1	33.9	Rủi ro thấp
147	21	52.2	0.165	1.72	78.7	96.2	Rủi ro thấp
148	29.6	60.8	0.163	2.45	75.3	124.1	Rủi ro thấp
149	26.2	57.4	0.129	2.72	74.3	124.2	Rủi ro thấp
150	25.2	56.4	0.157	1.72	73.3	98.0	Rủi ro thấp
151	16.9	106.7	0.276	1.006	77.9	93.7	Rủi ro thấp
152	44.1	111.3	0.232	1.506	83.9	120.1	Rủi ro thấp
153	30	117.3	0.219	0.963	75.4	92.7	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
154	28.6	108.8	0.226	0.931	71.1	91.3	Rủi ro thấp
155	24.9	104.5	0.162	0.909	77.6	77.7	Rủi ro thấp
156	17.1	111	0.196	1.216	77	87.6	Rủi ro thấp
157	19.1	110.4	0.215	1.332	82.9	95.5	Rủi ro thấp
158	23.5	116.3	0.213	1.206	75.7	94.7	Rủi ro thấp
159	21.3	109.1	0.179	0.928	72.6	78.9	Rủi ro thấp
160	25.9	106	0.778	1.02	34.4	180.1	Rủi ro trung bình
161	12.3	67.8	1.332	0.08	61.1	228.9	Rủi ro trung bình
162	8.2	94.5	0.049	1.78	60.1	73.8	Rủi ro thấp
163	12.8	93.5	0.054	1.72	59.3	75.8	Rủi ro thấp
164	18.8	92.7	0.179	2.45	59.9	121.6	Rủi ro thấp
165	10.3	93.3	0.149	2.72	75.9	119.4	Rủi ro thấp
166	16.9	76.7	0.33	0.936	80.5	98.1	Rủi ro thấp
167	15.9	81.3	0.286	1.436	86.5	105.7	Rủi ro thấp
168	15.1	87.3	0.273	0.893	78	87.2	Rủi ro thấp
169	15.7	78.8	0.28	0.861	73.7	87.1	Rủi ro thấp
170	15	74.5	0.216	0.839	80.2	75.5	Rủi ro thấp
171	18.6	81	0.25	1.146	79.6	93.0	Rủi ro thấp
172	17.1	80.4	0.269	1.266	85.5	98.7	Rủi ro thấp
173	14.8	86.3	0.267	1.136	78.3	93.3	Rủi ro thấp
174	4.5	79.1	0.233	0.858	75.2	72.1	Rủi ro thấp
175	12.3	76	0.133	0.95	44.1	63.7	Rủi ro thấp
176	1.2	44.9	0.167	0.01	84.7	31.5	Rủi ro thấp
177	25.5	85.5	0.186	1.71	89.3	104.7	Rủi ro thấp
178	30.1	90.1	0.184	1.65	95.3	106.0	Rủi ro thấp
179	36.1	96.1	0.15	2.38	86.8	126.8	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
180	27.6	87.6	0.178	2.65	75.7	133.1	Rủi ro thấp
181	6.2	78.5	0.359	0.57	80.3	84.7	Rủi ro thấp
182	10.8	83.1	0.315	0.169	86.3	69.1	Rủi ro thấp
183	16.8	89.1	0.302	0.946	77.8	94.7	Rủi ro thấp
184	8.3	80.6	0.309	0.914	73.5	88.6	Rủi ro thấp
185	4	76.3	0.245	0.892	80	74.5	Rủi ro thấp
186	10.5	82.8	0.279	1.199	79.4	94.0	Rủi ro thấp
187	9.9	82.2	0.298	1.319	85.3	100.2	Rủi ro thấp
188	15.8	88.1	0.296	1.189	78.1	100.3	Rủi ro thấp
189	8.6	80.9	0.262	0.918	75	81.4	Rủi ro thấp
190	5.5	77.8	0.861	1.01	39.1	177.5	Rủi ro trung bình
191	17	46.7	1.415	0.07	57.5	243.5	Rủi ro trung bình
192	15	87.3	0.132	1.77	66.1	90.8	Rủi ro thấp
193	19.6	91.9	0.137	1.71	62.7	93.2	Rủi ro thấp
194	18.2	97.9	0.262	2.44	61.7	134.6	Rủi ro thấp
195	17.1	89.4	0.232	2.71	88.2	136.7	Rủi ro thấp
196	12.4	71.8	0.245	1.69	92.6	103.8	Rủi ro thấp
197	17	76.4	0.348	2.42	9.3	145.1	Rủi ro thấp
198	23	82.4	0.017	2.69	57.8	105.0	Rủi ro thấp
199	14.5	73.9	0.093	1.69	37.6	80.7	Rủi ro thấp
200	10.2	69.6	0.008	0.753	61.8	36.0	Rủi ro thấp
201	16.7	76.1	0.115	0.97	75.7	64.5	Rủi ro thấp
202	16.1	75.5	0.027	1.33	42.4	60.6	Rủi ro thấp
203	22	81.4	0.011	2.52	255.9	99.3	Rủi ro thấp
204	14.8	74.2	0.004	0.99	46.8	45.8	Rủi ro thấp
205	11.7	71.1	0.063	1.75	107.9	76.1	Rủi ro thấp

TT	As (mg/Kg)	Pb (mg/Kg)	Hg (mg/Kg)	Cd (mg/Kg)	Zn (mg/Kg)	RI	KQ
206	14.1	32.9	0.049	0.47	86.9	34.2	Rủi ro thấp
207	15.9	59.6	0.842	1.8	59.9	203.9	Rủi ro trung bình
208	13.2	58.6	0.789	2.53	97.6	215.7	Rủi ro trung bình
209	11.3	57.8	0.113	2.8	30.1	113.9	Rủi ro thấp
210	15.7	58.4	0.118	1.8	11.3	87.6	Rủi ro thấp
211	14.2	96.8	0.23	1.803	87.3	107.8	Rủi ro thấp
212	18.8	101.4	0.333	2.532	91.9	149.5	Rủi ro thấp
213	24.8	107.4	0.073	2.453	97.9	110.0	Rủi ro thấp
214	16.3	98.9	0.078	1.542	89.4	77.2	Rủi ro thấp
215	12	94.6	0.007	0.864	85.1	42.3	Rủi ro thấp
216	18.5	101.1	0.173	1.045	91.6	79.1	Rủi ro thấp
217	17.9	100.5	0.012	1.443	91	64.8	Rủi ro thấp
218	23.8	106.4	0.004	2.631	96.9	103.6	Rủi ro thấp
219	16.6	99.2	0.011	1.101	89.7	53.5	Rủi ro thấp
220	13.5	96.1	0.048	1.861	86.6	79.9	Rủi ro thấp
221	11.4	57.9	0.034	0.78	50.7	40.9	Rủi ro thấp
222	13.4	84.6	0.827	1.529	69.1	193.6	Rủi ro trung bình
223	13.8	83.6	0.774	2.392	77.7	211.2	Rủi ro trung bình
224	15.7	82.8	0.098	2.643	74.3	111.8	Rủi ro thấp
225	13.4	83.4	0.103	1.642	73.3	81.0	Rủi ro thấp