ĐỊNH LƯỢNG Ô NHIỄM NƯỚC MẶT TẠI THÀNH PHỐ HỘI AN DỰA TRÊN DỮ LIỆU VIỄN THÁM VÀ MÔ HÌNH HỌC MÁY

ĐỖ THỊ NHUNG[(1)], NGUYỄN THỊ DIỄM MY[(1)], TRẦN QUỐC TUẤN[(2)]
NGHIÊM VĂN TUẤN[(3)], PHẠM VĂN MẠNH[(1)]

[(1)]Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội
[(2)]Viện Bảo tồn Di tích, Bộ Văn hóa, Thể thao và Du lịch
[(3)]Cục Viễn thám Quốc gia, Bộ Tài nguyên và Môi trường

Tóm tắt:

Tình trạng nước mặt tại thành phố Hội An đang phải đối diện với những thách thức về mức độ ô nhiễm trong hơn một thập kỷ qua. Nghiên cứu này giới thiệu tích hợp mô hình học máy với hệ thống suy luận mờ dựa trên mạng thích ứng (ANFIS), kết hợp với dữ liệu viễn thám quang học và radar để ước tính ba thông số chất lượng nước như TSS, COD và BOD. Hiệu suất của mô hình được đánh giá bằng cách sử dụng các tham số bao gồm RMSE, R^2, MAE. Các chỉ số dự đoán cung cấp kết quả đáng kể, với RMSE dao động từ 3,52 mg/l đến 4,59 mg/l, R^2 dao động từ 0,69-0,82 và MAE dao động từ 2,39 mg/l đến 3,16 mg/l. Kết quả cho thấy, nồng độ của ba thông số đánh giá chất lượng nước cao phân bố ở khu vực Đô thị cổ Hội An, sông Hoài. Phương pháp này có thể sử dụng để đánh giá các thông số chất lượng nước mặt, nhằm giám sát nhanh tình trạng của môi trường nước, cung cấp một phương pháp giám sát chất lượng nước mặt, đây có thể là nền tảng để đưa ra giải pháp bảo vệ và sử dụng nguồn nước sạch ở các thành phố Di sản.

Từ khóa: Ô nhiễm nước mặt, Viễn thám, Chất lượng nước, Mô hình ANFIS, Thành phố Hội An.

1. Tính cấp thiết

Môi trường nước tự nhiên là cơ sở cho sự phát triển bền vững và nó cũng là nguồn lực trực tiếp của việc phát triển và sử dụng du lịch [1]. Trong thời kì hội nhập, đô thị hóa được nhìn nhận như một quá trình, xu hướng và đã trở thành một làn sóng không ngừng gia tăng mạnh mẽ [2]. Sự suy giảm chất lượng nước ở các sông, hồ mang theo nhiều chất thải gây ô nhiễm môi trường nước mặt. Nước mặt là một trong những tài nguyên cảnh quan quan trọng trong các địa điểm du lịch [3]. Do đô thị hóa và các hoạt động du lịch, chất lượng nước mặt ở các danh lam thắng cảnh thường bị tổn hại. Giám sát chất lượng nước mặt và giảm thiểu tác động tiêu cực đến tài nguyên nước đòi hỏi dữ liệu chính xác, đáng tin cậy phải được cập nhật thường xuyên [4], [5]. Hiện nay, với việc

Ngày nhận bài: 1/5/2022, ngày chuyển phản biện: 5/5/2022, ngày chấp nhận phản biện: 9/5/2022, ngày chấp nhận đăng: 20/5/2022

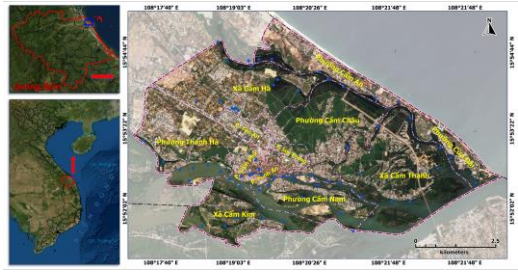
ứng dụng rộng rãi công nghệ viễn thám (RS) và hệ thống thông tin địa lý (GIS), các nghiên cứu đã giải quyết được từ góc độ không gian và thời gian, chẳng hạn như lập bản đồ phân bố tài nguyên nước và giám sát mức độ ô nhiễm nước [6], [7]. Các phương pháp phát hiện từ xa dựa vào việc đo những thay đổi này mà tín hiệu phản xạ hoặc tán xạ ngược từ môi trường nước và liên hệ với những thay đổi có chủ ý này bằng các mô hình phân tích hoặc thực nghiệm với các thông số chất lượng nước. Các mô hình điều tra chất lượng nước là nguồn lực hữu hiệu để đánh giá và xác định mức độ ô nhiễm, xu hướng của các chất ô nhiễm, ghi nhận sự phân bố theo không gian và thời gian của các chất ô nhiễm trong nước [8]. Tiềm năng của việc áp dụng các mô hình học máy kết hợp với dữ liệu viễn thám quang học và radar có thể giúp ước tính chất lượng nước mặt, và dữ liệu SAR có thể cải thiện giá trị gần đúng cho các chỉ số ở vùng nước bề mặt [9].

Trong những năm gần đây, sự mở rộng của khu vực đô thị đã được ghi nhận trên toàn thành phố Hội An. Sự phát triển mạnh mẽ của ngành du lịch-dịch vụ-thương mại và công nghiệp hóa, hiện đại hóa ở Hội An đã mang lại nhiều giá trị kinh tế và nâng cao mức sống của người dân. Tuy nhiên, du lịch phát triển và thiếu sự kiểm soát chặt chẽ ở Hội An có thể dẫn đến nguy cơ phát triển theo hướng thiếu bền vững và gây ra nhiều hậu quả nguy hại đến môi trường, mỹ quan thành phố, đặc biệt là vấn đề ô nhiễm nguồn nước mặt [10]. Do đó, nghiên cứu này thực hiện một phân tích định lượng dựa trên tích hợp mô hình học máy với hệ thống suy luận mờ dựa trên mạng thích ứng (ANFIS), kết hợp với dữ liệu viễn thám quang học và radar để đánh giá nồng độ của các thông số chất lượng nước mặt tại khu vực.

2. Khu vực nghiên cứu và cách tiếp cận

2.1. Khu vực nghiên cứu

Thành phố Hội An (phần đất liền) thuộc tỉnh Quảng Nam có tổng diện tích 6.171,25 ha nằm ở vùng cửa sông ven biển, cuối tả ngạn của sông Thu Bồn, ôm trọn bờ bắc Cửa Đại, được giới hạn từ $15^0$15’26” đến $15^0$55’15” vĩ độ Bắc và từ $108^0$17’08” đến $108^0$23’10” kinh độ Đông (Hình 1).



Hình 1: Khu vực nghiên cứu thành phố Hội An phần đất liền và vị trí các mẫu nước

Thành phố Hội An mang đặc điểm chung của khí hậu nhiệt đới ẩm, gió mùa. Do nằm ở hạ lưu của sông Thu Bồn-hệ thống sông lớn bậc nhất miền Trung, vùng thượng nguồn của hệ thống sông có lượng mưa rất lớn (3.000-4.000 mm/năm). Quá trình đô thị hóa các hoạt động du lịch, dịch vụ khiến môi trường tự nhiên tại các khu dân cư ở Hội An thêm bức bí. Nồng độ một số thông số chất lượng nước đo được tại một số điểm quanh khu vực Chùa Cầu, sông Hoài và các kênh rạch tại Cẩm Châu, Cẩm Thanh,... đã vượt giới hạn tiêu chuẩn gấp nhiều lần [10]. Khắc phục tình trạng ô nhiễm môi trường Hội An dù cấp thiết nhưng không hề đơn giản.

2.2. Dữ liệu và cách tiếp cận

2.2.1. Bộ dữ liệu và xử lý ảnh vệ tinh

Trong khi các phương pháp tại chỗ cung cấp dữ liệu chính xác tại thời điểm lấy mẫu, nhưng thường tốn nhiều công sức và do đó không hiệu quả về thời gian và chi phí [11]. Hơn nữa, giám sát toàn bộ vùng nước mặt

(sông, hồ) từ tập dữ liệu một lần/duy nhất là khá khác biệt [12]. Do đó, bộ dữ liệu/kỹ thuật không gian địa lý là công cụ có thể định lượng trong các nghiên cứu chất lượng nước [13], được sử dụng để đo các thông số như tổng lượng cặn lơ lửng (TSS), Nhu cầu oxy sinh hóa (BOD), nhu cầu oxy hóa học (COD),… [7], [14], [15], [16], [17], [18]. Những nghiên cứu này cung cấp cơ sở khoa học và phương pháp tính toán hiện đại trong giám sát chất lượng nước mặt. Do đó, trong nghiên cứu này, dữ liệu viễn thám được sử dụng để tham gia vào mô hình tính toán một số thông số chất lượng nước (BOD, COD, TSS) tại khu vực nghiên cứu. Các thông số được lựa chọn ở đây là những thông số được sử dụng phổ biến nhất để phân tích chất lượng nước và xác định mức độ phù hợp sử dụng nước theo các mức độ chất lượng nước được chỉ định khác nhau [19].

Hai nguồn dữ liệu ảnh vệ tinh được sử dụng: (i) Dữ liệu đa phổ Sentinel-2 và (ii) dữ liệu Sentinel-1 SAR, được thu nhận vào 26/4/2022 trong khoảng thời gian gần với các phép đo lấy mẫu ngoài thực địa được tiến hành. Mặc dù có sự khác biệt nhỏ giữa ngày của các thông số chất lượng nước đo được và hình ảnh vệ tinh (vì các dữ liệu này được thu thập từ hai nguồn khác nhau), ngày lấy mẫu nước khá gần với ngày của thu nhận ảnh vệ tinh và như vậy, các thông số này sau đó có thể được so sánh khả thi. Hơn nữa, không có bộ dữ liệu tương tự nào khác cho bất kỳ ngày thay thế nào trong khoảng thời gian đã lấy mẫu, nên phân tích này được thực hiện dựa trên dữ liệu có sẵn.

Dữ liệu Sentinel-2 đa phổ (Blue, Green, Red, NIR với độ phân giải không gian 10m; Red edge và SWIR với độ phân giải không gian 20m) được sử dụng trong nghiên cứu này. Khu vực thành phố Hội An có độ phủ mây dưới 3%. Quy trình xử lý ảnh bao gồm ba phần: (i) hiệu chỉnh đỉnh phản xạ ở tầng khí quyển; (ii) loại bỏ mây mù trên ảnh; (iii) chuyển đổi giá trị bức xạ về phản xạ bề mặt. Đối với dữ liệu ảnh vệ tinh Sentinel-1 SAR, các giai đoạn xử lý ảnh bao gồm, (1) hiệu chuẩn đo tán xạ để chuyển đổi các giá trị của hình ảnh thô (DN) sang hệ số tán xạ ngược của Radar (sigma naught); (2) sử dụng bộ lọc thích ứng Lee với kích thước cửa sổ (7×7) được áp dụng để giảm nhiễu hạt và làm mịn hệ số tán xạ ngược của dữ liệu; (3) hiệu chỉnh địa hình bằng cách sử dụng dữ liệu mô hình độ cao ở độ phân giải không gian 10m để hiệu chỉnh các biến dạng hình học. Cuối cùng, cả hai dữ liệu được tham chiếu về hệ tọa độ VN2000 múi 49.

Bảng 1: Các chỉ số trích xuất từ dữ liệu viễn thám sử dụng trong nghiên cứu

Biến dự đoán		Chú thích	Công thức
Sentinel-2A	BLUE	Xanh lam (458 nm-523 nm)	
	GREEN	Xanh lá (543-578 nm)	
	RED	Đỏ (650-680 nm)	
	NIR	Cận hồng ngoại (785-899 nm)	
Các chỉ số quang	NDWI	Chỉ số Nước chênh lệch chuẩn hóa	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
	MNDWI	Chỉ số Nước chênh lệch chuẩn hóa sửa đổi	$MNDWI = \frac{GREEN - SWIR1}{GREEN + SWIR1}$
	WRI	Chỉ số Tỷ lệ nước	$WRI = \frac{GREEN + RED}{NIR + SWIR1}$
	$AWEI_{nsh}$	Chỉ số Khai thác nước tự động	$AWEI_{nsh} = 4 \times (GREEN - SWIR1) - (0.25 \times NIR + 2.75 \times SWIR2)$
	$AWEI_{sh}$	Chỉ số Khai thác nước tự động	$AWEI_{sh} = BLUE + 2.5 * GREEN - 1.5 * (NIR + SWIR1) - 0.25 * SWIR2$

Biến dự đoán		Chú thích	Công thức
	SWI	Chỉ số Nước Sentinel-2	$SWI = \frac{VRE1 - SWIR2}{VRE1 + SWIR2}$
	MBWI	Chỉ số Nước đa kênh phổ	$MBWI = 2 \times (GREEN) - RED - NIR - SWIR1 - SWIR2$
Sentinel-1A	VH	Phân cực truyền dọc-nhận ngang (dB)	$Sigma\theta_VH$
	VV	Phân cực truyền dọc-nhận dọc (dB)	$Sigma\theta_VV$
	$ADDITION_{vhvv}$	Chỉ số phân cực tổng (dB)	VH+VV
	$AVERAGE_{vhvv}$	Chỉ số phân cực trung bình (dB)	(VH+VV)/2
	$MULT_{vhvv}$	Chỉ số phân cực tích (dB)	VH*VV

2.2.2. Nguồn và phân tích dữ liệu chất lượng nước

Tổng cộng có 65 vị trí được lấy mẫu nước mặt dọc theo các sông chính (35 mẫu) và hồ (30 mẫu) của khu vực nghiên cứu (Hình 1). Các mẫu này được thu thập vào khoảng thời gian từ ngày 25-27 tháng 4 năm 2022, các mẫu nước được thu thập dưới mặt nước 0,5 m và cách bờ sông ít nhất 10 m bằng cách sử dụng bình nhựa có dung tích 500 ml và các mẫu được bảo quản trong thùng đá lạnh cho đến khi tiến hành phân tích.

Hình 2: Minh họa vị trí lấy mẫu nước mặt và bảo quản mẫu tại Hội An

Các giá trị của các thông số đánh giá chất lượng nước (BOD, COD và TSS) là kết quả phân tích trong phòng thí nghiệm từ các chuyên gia thuộc Trung Tâm Quan Trắc Môi Trường Hội An cung cấp. Giá trị các chỉ số đánh giá chất lượng trong nước (BOD, COD và TSS) trong các mẫu nước mặt, đối với BOD có giá trị dao động trong khoảng (20,69-43,53 mg/l), COD (25,51-54,53 mg/l) và TSS (40,48-87,62 mg/l). Các mẫu được chia thành hai phần ngẫu nhiên, 70% số mẫu được sử dụng tham gia vào mô hình tính toán và 30% số mẫu còn lại được sử dụng để kiểm chứng kết quả của mô hình dự đoán. Các mẫu (BOD, COD và TSS) được sử dụng là các biến phụ thuộc, trong khi đó các biến độc lập được chiết tách từ dữ liệu ảnh viễn thám.

2.3.2 Mô hình ước tính các thông số chất lượng nước từ dữ liệu viễn thám

Hệ thống suy luận mờ dựa trên mạng thích ứng (ANFIS) được sử dụng phổ biến để giải quyết nhiều vấn đề trong các ứng dụng hồi quy đa biến [20]. ANFIS là một kỹ thuật học tập dữ liệu đơn giản sử dụng Logic mờ để chuyển đổi các đầu vào đã cho thành một đầu ra mong muốn thông qua mạng thần kinh được kết nối thông tin và tính toán trọng số để ánh xạ các đầu vào thành một đầu ra. Trong ANFIS, số lượng các nút ẩn trong mạng nơ-ron cũng như trong một hệ thống suy luận mờ bao gồm năm lớp cố định (Hình 3). Kiến trúc ANFIS cho mô hình đa biến với đầu vào gồn 16 biến độc lập (Bảng 1), và một đầu ra *y* bằng cách giả định bộ quy tắc với các quy tắc "if-then" mờ được hiển thị như phương trình (1).

$$If\ Z_k \text{ is } A_{kj}\ then\ y^{(j)} = \theta_{j0} + \sum_{k=1}^{p} \theta_{jk} Z_k \theta_{jk} \quad (1)$$

Trong đó Z_k là đầu vào, và $y^{(j)} = \theta_{j0} + \sum_{k=1}^{p} \theta_{jk} Z_k \theta_{jk}$ như đầu ra được chỉ định

bởi quy tắc mờ; θ_{j0}, θ_{jk} là các tham số tuyến tính; A_{kj} là các tham số phi tuyến; j=1,2…m; k=1,2…*p*; và có thể được ước tính thông qua một số bước như được mô tả trong năm lớp sau:

Lớp-1: Lớp này bao gồm dữ liệu đào tạo với mười sáu biến dự đoán liên quan (như được trình bày trong Bảng 1). Bước sơ bộ, một phương pháp phân cụm đã được sử dụng để xác định các cụm tối ưu cho tập dữ liệu đào tạo này. Các đầu ra của Lớp-1 là cấp độ thành viên mờ của các đầu vào có thể được viết dưới dạng hàm thành viên theo phương trình (2).

$$\mu_{A_{kj}}(Z_k) = \exp\left(-\frac{1}{2}\left(\frac{Z_k - c_{jk}}{a_{jk}}\right)^2\right), j = 1,2...m; k = 1,2...p \quad (2)$$

Trong đó Z là đầu vào của nút k, c_{jk} là tham số vị trí và a_{jk} là tham số tỷ lệ của hàm thành viên. Đây là các tham số tiền đề và được đưa vào lớp tiếp theo.

Lớp-2: Trong lớp 2, các nút là các nút cố định và được gắn nhãn với $\prod$, lớp này sử dụng các toán tử mờ để làm mờ các đầu vào và được thực hiện như một cấp số nhân đơn giản. Đầu ra của lớp này có thể được biểu diễn dưới dạng phương trình (3).

$$W_j = \prod_{k=1}^{p} \mu_{A_{kj}}, j = 1,2...m \quad (3)$$

Lớp-3: Mọi nút trong lớp thứ ba cũng là một nút cố định, tính toán tỷ lệ độ chặt và được chuẩn hóa trong điều này được gắn nhãn N. Kết quả đầu ra của lớp này được gọi là độ chặt chuẩn hóa. Đầu ra có thể được biểu diễn theo phương trình (4).

$$\overline{W}_j = \frac{W_j}{\sum_{j=1}^{m} W_j} \quad (4)$$

Lớp-4: Trong lớp 4, các nút là lớp thích ứng và được lấy tổng các hàm tuyến tính nhân với trọng số chuẩn hóa ở bước trước. Đầu ra của mỗi nút được tính theo phương trình (5).

$$\overline{W}_j y^{(j)} = \overline{W}_j\left(\theta_{j1}Z_1 + \theta_{j2}Z_2 + ... + \theta_{jp}Z_p + \theta_{j0}\right) \quad (5)$$

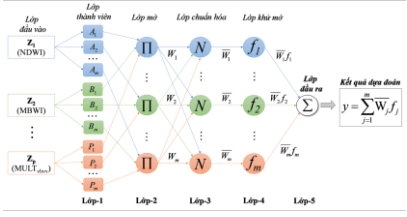
Lớp-5: Mỗi nút trong lớp thứ năm là một nút cố định bổ sung tất cả các tín hiệu đến đều được gắn nhãn $\sum$. Đầu ra của toàn mạng có thể được tính bằng (6).

$$y = \sum_{j=1}^{m} \overline{W}_j\left(\theta_{j1}Z_1 + \theta_{j2}Z_2 + ... + \theta_{jp}Z_p + \theta_{j0}\right) \quad (6)$$

Khi mô hình ANFIS theo phương trình (7), thì việc ước lượng các tham số hệ quả của mô hình ANFIS (θ_{jk} ; j=1,2…m; và k=1,2…p) có thể sử dụng thuật toán lọc thích ứng nhằm mục đích giảm bớt sai số.

$$\hat{\theta}_{11}(\overline{w}_1 Z_{t-1}) + \hat{\theta}_{12}(\overline{w}_1 Z_{t-2}) + ... + \hat{\theta}_{1p}(\overline{w}_1 Z_{t-p}) + \hat{\theta}_{10}\overline{w}_1 + \hat{\theta}_{21}(\overline{w}_2 Z_{t-1}) + \hat{\theta}_{22}(\overline{w}_2 Z_{t-2}) + ... + \hat{\theta}_{2p}(\overline{w}_2 Z_{t-p})$$
$$+\hat{\theta}_{20}\overline{w}_2 + ... + \hat{\theta}_{m1}(\overline{w}_m Z_{t-1}) + \hat{\theta}_{m2}(\overline{w}_m Z_{t-2}) + ... + \hat{\theta}_{mp}(\overline{w}_m Z_{t-p}) + \hat{\theta}_{m0}\overline{w}_m = Z_t \quad (7)$$

Khi các tham số dự đoán đã được ước tính, thì các tham số độc lập có thể được cập nhật bằng cách sử dụng phương pháp truyền ngược lỗi. Tất cả dữ liệu đào tạo đều đi qua mạng nơ-ron, để điều chỉnh các tham số đầu vào để tìm mối quan hệ giữa đầu vào/ đầu ra và giảm thiểu lỗi.



Hình 3: Hệ thống suy luận mờ dựa trên mạng thích ứng (ANFIS)

Nghiên cứu đã sử dụng 70% dữ liệu mẫu để tham gia quá trình đào tạo mô hình và lựa chọn các chỉ số tối ưu ước tính ba thông số chất lượng nước (BOD, COD và TSS); 30% dữ liệu mẫu còn lại được sử dụng để đánh giá hiệu suất của mô hình tính toán và kiểm chứng kết quả dựa đoán. Các thông số của mô hình học máy đã được tối ưu hóa bằng cách kiểm tra theo chu kỳ với phạm vi của các tham số và kích thước nhất định. Phương trình ước tính các thông số chất lượng nước (BOD, COD và TSS), được tính toán từ mô hình ANFIS và được mô tả trong phương trình 8 (TSS), phương trình 9 (COD) và phương trình 10 (BOD).

$$TSS = 99.3 - 32.3 \times MNDWI - 43 \times AWEI_{sh} + 24.5 \times AWEI_{nsh} - 11 \times NDWI + 0.65 \times VH \quad (8)$$

$$COD = 31.9 + 42.3 \times \mathrm{AWE}I_{nsh} - 8.8 \times \mathrm{WRI} - 63 \times \mathrm{AWE}I_{sh} - 94 \times MBWI + 27.9 \times NDWI - 1.78 \times VV - 19.1 \times MNDWI + 10.6 \times SWI - 0.25 \times VH \quad (9)$$

$$BOD = 80.65 + 0.12 \times MULT_{vhvv} + 2.59 \times VV + 2.26 \times VH - 25.4 \times NDWI + 4.38 \times \mathrm{WRI} - 4.9 \times MNDWI - 13.3 \times \mathrm{AWE}I_{sh} \quad (10)$$

Các thông số chất lượng nước (BOD, COD và TSS) được tính toán từ mô hình ANFIS theo các phương trình (8), (9) và (10), được kết hợp với phương pháp Fuzzy- AHP để lập bản đồ chỉ số ô nhiễm nước mặt tiềm năng liên quan như đã thảo luận trong phần 2.4.

2.3. Đánh giá độ chính xác

Trong mô hình ANFIS, nghiên cứu sử dụng phương pháp xác nhận chéo k-fold được sử dụng để đánh giá hiệu suất của mô hình dự đoán, trong đó dữ liệu được chia ngẫu nhiên 10-fold như đã được chỉ ra rằng 10-fold là con số tối ưu để xác nhận chéo tối ưu [11]. Các chỉ số hiệu suất của mô hình để hiệu chuẩn và xác nhận được đánh giá thông qua ước tính Hệ số xác định (R^2), Sai số toàn phương trung bình (RMSE) và Sai số tuyệt đối trung bình (MAE). Những giá trị này xác định mô hình học máy hiệu quả nhất với dữ liệu và vị trí nhất định. Hệ số xác định (R^2) được sử dụng để xác định tính độc lập của các biến bằng phương trình 8.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{ob} - y^{pre})^2}{\sum_{i=1}^{n}(y_i^{ob} - \bar{y}^{ob})^2} \quad (11)$$

Trong đó, n là một số mẫu quan sát được; y_i^{ob} là các giá trị quan sát được, là các giá trị dự đoán là y_i^{pre}; $\bar{y}^{ob}$ mức trung bình của giá trị quan sát được.

Sai số toàn phương trung bình (RMSE) phản ánh độ chính xác của mô hình ước tính bằng cách so sánh các giá trị dự đoán với dữ liệu quan sát thực tế. RMSE được tính toán bằng phương trình 9.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{ob} - \bar{y}^{pre})^2}{n}} \quad (12)$$

Sai số tuyệt đối trung bình (MAE) cung cấp một sai số dự đoán trung bình với điểm số với định hướng tiêu cực. MAE được thể hiện bằng phương trình 10.

$$MAE = \frac{\sum_{i=1}^{n} | y_i^{ob} - y_i^{pre} |}{n} \quad (13)$$

Để tối ưu hóa việc ước tính các thông số chất lượng nước (BOD, COD, và TSS), các giá trị được điều chỉnh trong học máy với mỗi lần lặp lại và được tách ngẫu nhiên (70% dữ liệu cho đào tạo và 30% dữ liệu còn lại để xác nhận) từ tập dữ liệu gốc. Dữ liệu đào tạo (45 mẫu) và dữ liệu xác nhận (20 mẫu) được tách

ngẫu nhiên từ 65 mẫu để đánh giá chất lượng nước từ các thông số BOD, COD và TSS.

2.4. Lập bản đồ chỉ số ô nhiễm nước mặt tại thành phố Hội An

Phương pháp tích hợp quá trình phân tích mờ (Fuzzy-AHP) được sử dụng để xác định trọng số của ba thông số chất lượng nước sử dụng trong nghiên cứu. Phương pháp Fuzzy-AHP dựa trên việc đưa trực tiếp lý thuyết tập mờ vào ma trận so sánh theo cặp để đảm bảo tính khách quan và chính xác của ba thông số chất lượng nước. Giá trị của ba thông số đánh giá chất lượng nước TSS, COD và BOD được tính toán từ phương trình (8), (9) và (10). Chỉ số mức độ ô nhiễm nước mặt (SWP) được tính toán theo phương trình (14) kết hợp trọng số từ Fuzzy-AHP. Chỉ số này được sử dụng để đánh giá sự khác biệt trong phân bố không gian của mức độ ô nhiễm của ba thông số.

$$SWP = W_A \times (BOD) + W_B \times (COD) + W_C \times (TSS) \quad (14)$$

Trong đó: W_A, W_B và W_C lần lượt là trọng số của BOD, COD và TSS. Phạm vi không gian của SWP được trình bày trong phần 3. Các chỉ số này được phân loại thành năm mức độ giá trị để đánh giá mức độ chất lượng nước dựa trên thuật toán "Natural Break" trong ArcGIS. Bảng 2 phân ngưỡng giá trị ô nhiễm liên quan đến các chỉ số chất lượng nước và được chia thành năm loại: (i) Rất thấp, (ii) Thấp, (iii) Vừa phải, (iv) Cao và (v) Rất cao.

Bảng 2: Mức độ ô nhiễm nước tại khu vực nghiên cứu

	Phân ngưỡng mức độ ô nhiễm				
	Rất thấp	Thấp	Vừa phải	Cao	Rất cao
TSS	< 55	55 - 60	60 - 70	70 -85	> 85
COD	< 35	35 - 45	45 - 60	60 - 80	> 80
BOD	< 30	30 - 35	35 - 45	45 - 55	> 55

3. Kết quả và thảo luận

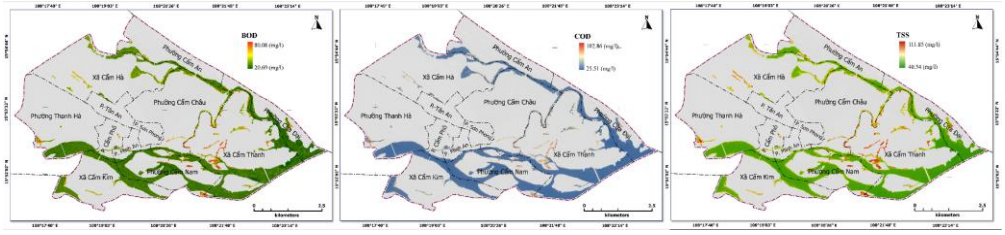
3.1. Đánh giá mô hình hiệu suất mô hình tính toán các thông số chất lượng nước

Nghiên cứu sử dụng tổng số 45 mẫu hiện trường để đào tạo trong mô hình ANFIS để tính toán các thông số BOD, COD và TSS trong các vùng nước mặt, và 20 mẫu còn lại để đánh giá hiệu suất và kết quả dự đoán. Kết quả của mô hình học máy đã được tối ưu hóa với quy trình điều chỉnh bằng cách xác nhận chéo 10 lần theo chu kỳ. Dựa trên kết quả của các mô hình được tối ưu hóa, CV-RMSE dao động từ 3,45 mg/l đến 3,61 mg/l, CV-R^2 dao động từ 0,73-0,83, CV-MAE dao động từ 2,32 mg/l đến 2,74 mg/l (Bảng 3). Trong khi đó, giá trị của các thông số chất lượng nước (BOD, COD và TSS) được ước tính từ dữ liệu ảnh viễn thám có giá trị lớn hơn không đáng kể so với hiệu suất của mô hình tính toán. Độ chính xác ước tính bằng cách sử dụng mô hình học máy với các số liệu kết hợp khác nhau là tương đối tốt. Nhìn chung, các chỉ số (BOD, COD và TSS) đều có giá trị CV-RMSE, CV-R^2 và CV-MAE chấp nhận được. Như vậy, phương trình ước tính TSS (8), phương trình ước tính COD (9) và phương trình ước tính BOD (10) có thể được áp dụng để ước tính các thông số chất lượng nước từ dữ liệu ảnh viễn thám tại thành phố Hội An.

Bảng 3: Hiệu suất của mô hình ước tính thông số chất lượng nước

Thông số chất lượng nước	Chỉ số đã chọn	Hiệu suất mô hình			Kiểm chứng		
		CV-RMSE	CV-R^2	CV-MAE	RMSE	R^2	MAE
TSS (Tổng chất rắn lơ lửng)	MNDWI, $AWEI_{sh}$, $AWEI_{nsh}$, NDWI, VH	3.45	0.83	2.32	3.52	0.82	2.39
COD (Nhu cầu oxy hóa học)	$AWEI_{nsh}$, WRI, $AWEI_{sh}$, MBWI, NDWI, VV, MNDWI, SWI, VH	3.57	0.81	2.59	4.11	0.79	3.05
BOD (Nhu cầu oxy sinh học)	$MULT_{vhvv}$, VV, VH, NDWI, WRI, MNDWI, $AWEI_{sh}$	3.61	0.73	2.74	4.59	0.69	3.16

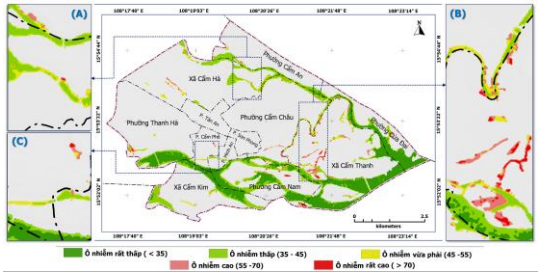
Hình 4 minh họa sự phân bố không gian nồng độ Nhu cầu oxy sinh hóa (BOD) cao và rất cao chủ yếu ở một số hồ nhỏ ở trung tâm thành phố, phân bố dọc theo sông Hoài, trong một số khu vực nuôi trồng thủy sản và khu canh tác nông nghiệp. Lượng BOD cần thiết để phân hủy các chất hữu cơ trong nước dao động tại khu vực nghiên cứu dao động từ 20,69-80,08 mg/l. Sự phân bố không gian của nồng độ COD cao và rất cao được phân bố ở các khu vực nước mặt, nồng độ tập trung ở các sông, kênh nước và ven sông lớn nơi có mô hình nuôi trồng thủy sản (sông Đế Vọng, sông Thu Bồn, sông Hoài) với giá trị dao động từ 25,51-102,86 mg/l. Trong khi đó, phân bố không gian nồng độ TSS cao và rất cao dọc theo hai bên bờ các sông lớn và các hồ, ao nước trong thành phố với tổng lượng cặn lơ lửng từ 40,54-111,85 mg/l.



Hình 4: Phân bố theo không gian của BOD (bên trái), COD (trung tâm), TSS (bên trái)

3.2. Bản đồ chỉ số ô nhiễm nước mặt tại thành phố Hội An

Từ các thông số đánh giá chất lượng nước được tính toán từ mô hình, bản đồ chỉ số ô nhiễm nước mặt được xác định kịp thời. Quy trình phân tích thứ bậc mờ (Fuzzy-AHP) được sử dụng để xác định trọng số của các chỉ số chất lượng nước. Chỉ số ô nhiễm nước mặt (SWP) được ước tính bằng giá trị của SWP theo phương trình (14). Trọng số chỉ số được tính cho BOD, COD và TSS lần lượt là 0,37; 0,47; 0,15 với λ_{max} = 3,05, CI = 0,03 và CR = 0,05 <0,10 (thỏa mãn điều kiện). Kết quả thể hiện trong Hình 5 mô tả sự phân bố theo không gian của chỉ số SWP tại thành phố Hội An với năm mức độ: (i) Ô nhiễm rất thấp (< 35), (ii) Ô nhiễm thấp (35-45), (iii) Ô nhiễm vừa phải (45-55), (iv) Ô nhiễm cao (55-70) và (v) Ô nhiễm rất cao (> 70). Sự phân bố theo không gian của SWP có mức độ ô nhiễm cao và rất cao ở các khu vực đông dân cư với các hoạt động kinh doanh và các khu vực trung tâm thành phố. Vùng ô nhiễm rất cao 35,52 ha (xấp xỉ 2,5%), vùng ô nhiễm cao 59,38 ha (xấp xỉ 4,18%), vùng ô nhiễm vừa phải 92,05 ha (xấp xỉ 6,48%), vùng ô nhiễm thấp 514 ha (xấp xỉ 36,2%), và vùng ô nhiễm rất thấp là 719,05 ha (khoảng 50,64%).



Hình 5: Phân bố theo không gian ô nhiễm nước mặt tại thành phố Hội An

3.3. Thảo luận

Kết quả phân tích ô nhiễm nước mặt cho thấy, ô nhiễm tại các khu vực ven sông Hoài, sông Đế Vọng và một số hồ nước trong trung tâm thành phố với giá trị BOD, COD, TSS cao và rất cao. Việc xả nước thải sinh hoạt và nguồn thải từ hoạt động nuôi trồng thủy sản là nguyên nhân chính gây ô nhiễm nước ở các khu vực này [10]. Kết quả mô hình cho thấy mức độ ô nhiễm nguồn nước cao tại các khu vực nuôi trồng thủy sản ven khu vực Trà Quế nằm bên nhánh sông Cổ Cò (Hình 5-A). Nuôi tôm công nghiệp với mật độ cao cùng với quá trình nuôi phải sử dụng các loại thuốc và hóa chất để xử lý nên nguồn nước thải ra môi trường có nguy cơ gây ô nhiễm rất cao. Tại các khu vực có mật độ hoạt động phát triển du lịch-thương mại ở đô thị cổ Hội An, Sông Hoài ghi nhận nguy cơ ô nhiễm nguồn nước cao (Hình 5-C). Một hồ điều tiết nhỏ trong nội khu của phố cổ và ao nước tù cho thấy mức độ ô nhiễm cao. Kết quả mô hình cho thấy nguy cơ ô nhiễm nguồn nước cao tại các khu vực nuôi trồng thủy sản trên sông Đế Vọng, và một phần của hệ sinh thái rừng dừa (các phường Cẩm Phô, Cẩm Thanh, Cẩm Hà) (Hình 5-B). Một lượng lớn chất thải hữu cơ và hóa học từ hoạt động chăn nuôi, nuôi trồng thủy sản và du lịch đã thải trực tiếp ra môi trường nước như vậy dẫn đến tác động tiêu cực đến cuộc sống hiện tại và cho tương lai.

Trong nghiên cứu này, chúng tôi đã đề xuất một phương pháp để tích hợp hai loại hình ảnh viễn thám (quang học và radar), trong đó tích hợp hình ảnh SAR bổ sung và cải thiện việc phát hiện các vùng nước mặt và các chỉ số chất lượng nước. Tuy nhiên, dữ liệu SAR với tán xạ ngược từ mặt nước bị ảnh hưởng bởi một số yếu tố do tương tác tín hiệu với các vật thể trên mặt nước hoặc độ nhám bề mặt, đặc biệt là dọc theo bờ sông (dừa, ao sen...). Đối với các khu vực có giá trị tán xạ rất cao từ mặt nước (ví dụ: rừng dừa, lồng nuôi trồng thủy sản), thuật toán có thể cho đánh giá quá cao các giá trị tham số, đặc biệt là ở các khu vực có độ sáng bề mặt cao. Việc tích hợp dữ liệu với mô hình học máy vẫn còn trong giai đoạn thử nghiệm cơ bản. Tuy nhiên, kết quả là khả quan và hứa hẹn để mở ra hướng nghiên cứu mới trong việc tích hợp radar và hình ảnh quang học để phân tích và ước tính các thông số chất lượng nước mặt ở các thành phố.

4. Kết luận

Sự kết hợp của nhiều dữ liệu viễn thám là rất quan trọng trong quá trình loại bỏ nhiễu và cải thiện độ chính xác của ước tính chất lượng nước mặt tại Hội An. Các kết quả ước tính từ phương pháp được đề xuất phù hợp với các mẫu tại chỗ, với mối tương quan đáng kể giữa các giá trị dự đoán và quan sát được. Kỹ thuật viễn thám và GIS kết hợp với lấy mẫu tại chỗ truyền thống, là những công cụ hiệu quả nhất, rẻ và đáng tin cậy hơn để theo dõi các thông số chất lượng nước trong các nguồn nước đô thị (hồ, sông). Dựa trên kết quả từ nghiên cứu này, các nhà hoạch định chính sách và quy hoạch đô thị có thể dễ dàng đưa ra các chiến lược can thiệp để kiểm soát các nguồn ô nhiễm nước mặt để đạt được mục tiêu phát triển bền vững thành phố du lịch xanh. Bên cạnh đó,

cộng đồng dân cư, doanh nghiệp, cơ quan, tổ chức khác phải hành động và nhận thức trách nhiệm của mình trong việc bảo vệ môi trường nước nói chung và tài nguyên nước của di sản nói riêng.❍

Lời cảm ơn

Nghiên cứu này được tài trợ bởi đề tài "Nghiên cứu xây dựng hệ thống hỗ trợ quản lý di sản thế giới của Việt Nam dựa trên ứng dụng công nghệ viễn thám và GIS".

Tài liệu tham khảo

[1]. Santos-Roldán L., Castillo Canalejo A.M., Berbel-Pineda J.M., et al. (2020). Sustainable Tourism as a Source of Healthy Tourism. *Int J Environ Res Public Health*, 17(15), 5353.

[2]. Nijskens R., Lohuis M., Hilbers P., et al., eds. (2019), *Hot Property: The Housing Market in Major Cities*, Springer International Publishing, Cham.

[3]. Chen P., Zhao Y., Zuo D., et al. (2021). Tourism, Water Pollution, and Waterway Landscape Changes in a Traditional Village in the Huizhou Region, China. *Land*, 10(8), 795.

[4]. Wolfram J., Stehle S., Bub S., et al. (2021). Water quality and ecological risks in European surface waters – Monitoring improves while water quality decreases. *Environ Int*, 152, 106479.

[5]. Uddin Md.G., Nash S., and Olbert A.I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecol Indic*, 122, 107218.

[6]. Said S. and Khan S.A. (2021). Remote sensing-based water quality index estimation using data-driven approaches: a case study of the Kali River in Uttar Pradesh, India. *Environ Dev Sustain*.

[7]. Patel P.P., Mondal S., and Ghosh K.G. (2020). Some respite for India's dirtiest river? Examining the Yamuna's water quality at Delhi during the COVID-19 lockdown period. *Sci Total Environ*, 744, 140851.

[8]. Do Thi Nhung, Nguyen Thi Diem My, Nguyen Thi Hong, Bui Quang Thanh, Lui Thi Phuong Mai, & Pham Van Manh (2021). Estimation risk of surface water pollution based on optical remote sensing data and multi-criteria decision analysis method. *Ho Chi Minh City University of Education Journal of Science*, 18(12), 2283-2296.

[9]. Hafeez S., Sing Wong M., Abbas S., et al. (2019). Detection and Monitoring of Marine Pollution Using Remote Sensing Technologies. *Monitoring of Marine Pollution*. IntechOpen.

[10]. Trung tâm Quản lý Bảo tồn Di sản Văn hóa Hội An (2021). Báo động ô nhiễm môi trường tại Hội An, tỉnh Quảng Nam. .

[11]. Hassan Omer N. (2020). Water Quality Parameters. *Water Quality - Science, Assessments and Policy*. IntechOpen.

[12]. Gholizadeh M., Melesse A., and Reddi L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, 16(8), 1298.

[13]. Elhag M., Gitas I., Othman A., et al. (2019). Assessment of Water Quality Parameters Using Temporal Remote Sensing Spectral Reflectance in Arid Environments, Saudi Arabia. *Water*, 11(3), 556.

[14]. Chen W., Wang J., Cao X., et al. (2021). Possibility of using multiscale normalized difference vegetation index data for the assessment of total suspended solids (TSS) concentrations in surface water: A

specific case of scale issues in remote sensing. *Environ Res*, 194, 110636.

[15]. El-Zeiny A. and El-Kafrawy S. (2017). Assessment of water pollution induced by human activities in Burullus Lake using Landsat 8 operational land imager and GIS. *Egypt J Remote Sens Space Sci*, 20, S49–S56.

[16]. He B., Oki K., Wang Y., et al. (2009). Using remotely sensed imagery to estimate potential annual pollutant loads in river basins. *Water Sci Technol*, 60(8), 2009–2015.

[17]. Shareef M.A., Toumi A., and Khenchaf A. (2014). Prediction of water quality parameters from SAR images by using multivariate and texture analysis models. Amsterdam, Netherlands, 924319, 924319.

[18]. Wang C., Li W., Chen S., et al. (2018). The spatial and temporal variation of total suspended solid concentration in Pearl River Estuary during 1987–2015 based on remote sensing. *Sci Total Environ*, 618, 1125–1138.

[19]. Abbasi T. and Abbasi S.A. (2012), *Water Quality Indices*, Elsevier, Oxford.

[20]. Pham M.H., Do T.H., Pham V.-M., et al. (2020). Mangrove forest classification and aboveground biomass estimation using an atom search algorithm and adaptive neuro-fuzzy inference system. *PLOS ONE*, 15(5), e0233110.❍

Summary

Quantitative of surface water pollution in hoi an city using remote sensing data and machine learning model

Do Thi Nhung, Nguyen Thi Diem My, Pham Van Manh - University of Science, Vietnam National University, Hanoi

Tran Quoc Tuan - Monument conservation institute, Ministry of Culture, Sports and Tourism of Viet Nam

Nghiem Van Tuan - Department of National Remote Sensing, Ministry of Natural Resources and Environment

The situation of surface water in Hoi An city has been facing challenges in terms of pollution levels for more than a decade. In order to estimate three water quality parameters, such as TSS, COD, and BOD, this study introduces the integration of a machine learning model with an adaptive network-based fuzzy inference system (ANFIS), paired with optical and radar remote sensing data. The RMSE, R^2, and MAE parameters were used to assess the model's performance. With RMSE ranging from 3.52 mg/l to 4.59 mg/l, R^2 ranging from 0.69-0.82, and MAE ranging from 2.39 mg/l to 3.16 mg/l, the predictive indices produced significant findings. The results show that the concentrations of three high water quality assessment parameters are distributed in the area of Hoi An Ancient Town, Hoai River. This method can be used to assess surface water quality parameters, to quickly monitor the status of the water environment, to provide a method of monitoring water surface quality, which can be the foundation for providing solutions to protect and use clean water sources in Heritage cities.❍

Keywords: Surface water pollution, Remote sensing, Water quality, ANFIS model, Hoi An city.