

# TỐI ƯU HÓA QUÁ TRÌNH HẤP PHỤ XANH METHYLENE BẰNG PHƯƠNG PHÁP MẠNG THẦN KINH NHÂN TẠO VÀ PHƯƠNG PHÁP MÁY VECTƠ HỖ TRỢ BÌNH PHƯƠNG NHỎ NHẤT

Hoàng Lương Cường<sup>(1)</sup>, Nguyễn Đức Hào<sup>(1)</sup>, Vũ Đức Lân<sup>(1)</sup>, Võ Hoa Sơn<sup>(1)</sup>

(1) Trường Đại học Khoa học Tự nhiên, VNU-HCM

Ngày nhận bài 2/5/2024; Chấp nhận đăng 30/5/2024

Liên hệ email: hoasonvo@gmail.com

<https://doi.org/10.37550/tdmu.VJS/2024.03.570>

## Tóm tắt

Vấn đề ô nhiễm nguồn nước đang trở nên ngày càng nghiêm trọng. Tuy nhiên, quá trình xử lý nước thải chưa được thực hiện một cách hiệu quả, dẫn đến việc một lượng lớn các chất độc hại bị thải trực tiếp ra môi trường nước mà không qua xử lý. Methylene Blue (MB) là một chất nhuộm hữu cơ, được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau như y học, sinh học, hóa học và công nghiệp, việc sử dụng MB với nồng độ lớn có thể gây tác động tiêu cực đối với sức khỏe, bao gồm các vấn đề như tổn thương tim, nôn mửa, sốc, và tê liệt cơ tứ chi. Trong bài báo này, chúng tôi đề xuất việc tối ưu hóa quá trình hấp phụ chất màu hữu cơ độc hại Methylene Blue (MB) nhờ vật liệu tổ hợp nano Sunfat Kẽm/Than hoạt tính ( ZnS NPs/AC ) nhằm xác định các điều kiện thực nghiệm để tối ưu hóa hiệu suất hấp phụ MB bằng mô hình trí tuệ nhân tạo (ANN) và mô hình Least Squares Support Vector Machine (LS-SVM). Kết quả thu được hệ số xác định ( $R^2$ ), sai số căn quân phương (RMSE) lần lượt là 0.98, 0.74 với mô hình ANN và 0.99, 0.24 với mô hình LS-SVM. Điều này cho thấy, mô hình LS-SVM cho khả năng dự báo chính xác hơn so với mô hình ANN và nhận được các điều kiện thực nghiệm tối ưu với độ pH là 6.6, nồng độ MB 8.8mg/L, khối lượng chất hấp phụ là 0.015g, thời gian siêu âm 4.9 phút và hiệu suất hấp phụ MB trên 97%.

**Từ khóa:** hấp phụ, LS-SVM, mạng nơron nhân tạo (ANN), MB, methylene blue

## Abstract

### OPTIMIZATION OF METHYLENE BLUE ADSORPTION PROCESS USING ARTIFICIAL NEURAL NETWORKS AND LEAST SQUARES SUPPORT VECTOR MACHINE METHOD

The water pollution problem is becoming increasingly serious. However, the wastewater treatment process has not been carried out effectively causing a large amount of toxic substances being discharged directly into the water environment without treatment. Methylene Blue (MB) is an organic dye, widely used in many different fields such as medicine, biology, chemistry and industry. Using MB in large concentrations can cause significant health issues such as heart damage, vomiting, shock, and limb paralysis. This research aims to optimize the adsorption process for MB using a composite of Zinc Sulfate Nanoparticles and Activated Carbon (ZnS NPs/AC). We utilized Artificial Neural Networks (ANN) and Least Squares Support Vector Machine (LS-SVM) models to identify the optimal conditions for MB adsorption. The performance of the models was assessed by their determination coefficients ( $R^2$ ) and root mean square errors (RMSE). Results revealed that the LS-SVM model, with an  $R^2$  of 0.99 and RMSE of 0.24, outperformed the ANN model, which had an  $R^2$  of 0.98 and an RMSE of 0.74. The optimal adsorption conditions were achieved at a pH of 6.6, MB concentration of 8.8mg/L, adsorbent mass of 0.015g, and ultrasonication time of 4.9 minutes, yielding an adsorption efficiency exceeding 97%.

## 1. Giới thiệu

Bài toán tối ưu hóa quá trình hấp phụ Methylene Blue (MB) bằng phương pháp trí tuệ nhân tạo và phương pháp Least Squares Support Vector Machine (LS-SVM) là một nghiên cứu hứa hẹn trong lĩnh vực xử lý nước và môi trường (Asfaram và cs., 2016). Mạng nơron nhân tạo ANN (Artificial Neural Networks) là một trong những kỹ thuật quan trọng trong lĩnh vực Trí tuệ nhân tạo (AI) (Braspenning và cs., 1995; Asadollahfardi, 2015), nó được thiết kế để mô phỏng cấu trúc và chức năng của não người để giải quyết các vấn đề phức tạp và thực hiện các nhiệm vụ học máy. Phương pháp ANN đã chứng minh độ hiệu quả của mình trong nhiều lĩnh vực khác nhau và đang ngày càng trở nên quan trọng trong cả lĩnh vực nghiên cứu và ứng dụng thực tế (Palani và cs., 2008). LS-SVM (Least Squares Support Vector Machine) là một biến thể của phương pháp SVM (Support Vector Machine) được thiết kế để giải quyết vấn đề hồi quy (Suykens và cs., 1999; Valyon & Horváth, 2007). LS-SVM giải quyết bài toán tối ưu hóa bằng cách tối thiểu hóa tổng bình phương sai số giữa giá trị thực tế và giá trị dự đoán của mô hình, đồng thời kiểm soát sự phức tạp của mô hình thông qua các tham số kiểm soát (Moayeri & Hemami, 2003).

Kết quả của nghiên cứu này cung cấp một quy trình tối ưu hóa quá trình thực nghiệm, với việc tối ưu các điều kiện hấp phụ MB như độ pH, nồng độ MB, khối lượng chất hấp phụ và thời gian siêu âm (Asfaram và cs., 2016). Tối ưu hóa các điều kiện này góp phần làm giảm quá trình thực nghiệm, giảm thời gian nghiên cứu cũng như tối ưu hóa chi phí đồng thời tăng hiệu suất xử lý nước thải công nghiệp.

## 2. Phương pháp nghiên cứu

Dựa vào số liệu thực nghiệm, chúng tôi xây dựng mô hình toán học gồm các biến độc lập ở lớp đầu vào và biến lệ thuộc ở lớp đầu ra của mô hình. Từ đó, trước tiên chúng tôi dùng phương pháp mạng nơron nhân tạo (ANN) để tối ưu hóa mô hình và dự đoán kết quả hấp phụ MB. Sau đó chúng tôi cải thiện mô hình ANN bằng phương pháp LS-SVM (Asfaram và cs., 2016; Wang & Hu, 2005), lập bảng so sánh với các phương pháp ANN và LS-SVM) và đưa ra phương trình hồi quy tối ưu cho hệ thực nghiệm.

Bảng 1 thể hiện bộ dữ liệu thực nghiệm (Asfaram và cs., 2016), bộ dữ liệu này có 4 đầu vào là độ pH (X1), lượng chất hấp phụ (X2)(gram), nồng độ MB (X3)(mg/L) và thời gian siêu âm (X4)(min), có 1 đầu ra là khả năng hấp phụ MB trong nước (Y).

*Bảng 1. Bộ dữ liệu thực nghiệm*

Số mẫu	X1	X2	X3	X4	Y	Số mẫu	X1	X2	X3	X4	Y
1	5	0.01	16	2	85.59	16	8	0.02	32	4	96.35
2	8	0.01	16	2	94.01	17	3.5	0.015	24	3	85.85
3	5	0.02	16	2	95.434	18	9.5	0.015	24	3	92.01
3	8	0.02	16	2	98.12	19	6.5	0.005	24	3	78.2592
5	5	0.01	32	2	80.45	20	6.5	0.025	24	3	96.03
6	8	0.01	32	2	88.2	21	6.5	0.015	8	3	98.36
7	5	0.02	32	2	91.022	22	6.5	0.015	40	3	92.31
8	8	0.02	32	2	93.31	23	6.5	0.015	24	1	94.1
9	5	0.01	16	4	87.24	24	6.5	0.015	24	5	98.596
10	8	0.01	16	4	90.561	25	6.5	0.015	24	3	97.874
11	5	0.02	16	4	98.5	26	6.5	0.015	24	3	96.51
12	8	0.02	16	4	97.56	27	6.5	0.015	24	3	97.447
13	5	0.01	32	4	86.24	28	6.5	0.015	24	3	96.8
14	8	0.01	32	4	88.5	29	6.5	0.015	24	3	97.76
15	5	0.02	32	4	99.214	30	6.5	0.015	24	3	97.18

### 2.1. Mô hình ANN:

Dựa vào các dữ liệu thực nghiệm nêu trên, mối liên hệ giữa các biến độc lập (thông số đầu vào) và các biến phụ thuộc (thông số đầu ra)

Các biến độc lập:  $X = [X1, X2, X3, X4]$  (1)

Trong đó:

X1: độ pH, X2: khối lượng chất hấp phụ, X3: nồng độ MB, X4: thời gian siêu âm.

Biến phụ thuộc đặc trưng cho đầu ra: Y

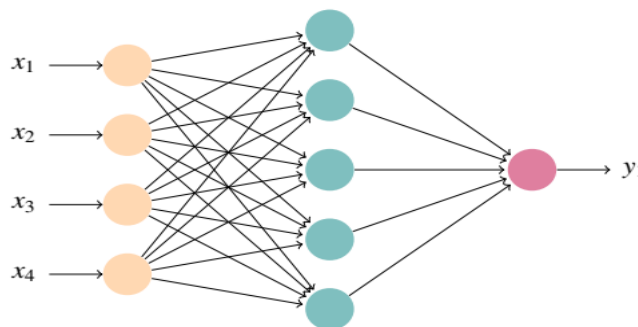
Với, Y: độ hấp phụ R.

Phương trình ma trận hồi quy bậc nhất đơn giản được trình bày như sau:

$$Y = (W_1 \times X + b_1) \times W_2 + b_2 \quad (2)$$

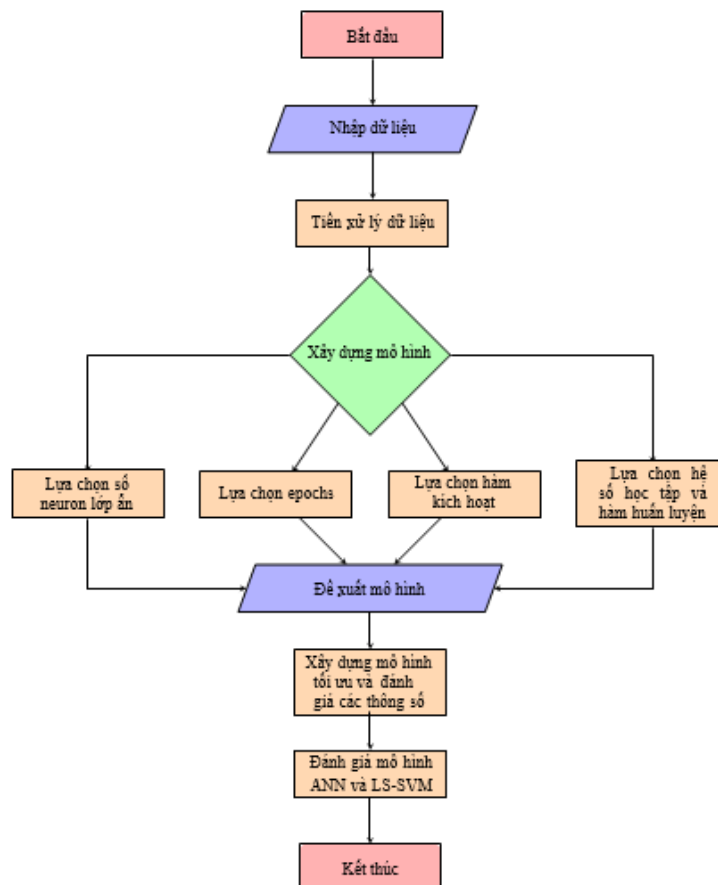
Trong đó, Y là ma trận đầu ra, W<sub>1</sub> là ma trận trọng số lớp đầu vào, X là ma trận đầu vào, b<sub>1</sub> là vector bias đầu vào, W<sub>2</sub> là ma trận trọng số lớp đầu ra, b<sub>2</sub> là vector bias đầu ra.

Mô hình ANN được trình bày như hình 1 với 4 biến đầu vào, 1 biến đầu ra, 5 neuron lớp ẩn



Hình 1. Mô hình ANN

Quá trình huấn luyện được trình bày chi tiết dựa trên lưu đồ ở hình 2.



Hình 2. Lưu đồ tiến trình Xây dựng Mô hình và dự đoán độ hấp phụ MB

**2.2. Mô hình LS-SVM:**

LS-SVM (Least Squares Support Vector Machine) là phương pháp học máy được sử dụng cho các vấn đề phân loại và hồi quy (Suykens & Vandewalle, 1999), phương pháp này tối ưu hóa khoảng cách giữa các lớp, nhưng thay vì giải bài toán tối ưu với ràng buộc, nó chuyển vấn đề về một bài toán hồi quy tuyến tính, với mục tiêu là giảm thiểu sai số bình phương. Hàm mục tiêu của LS-SVM liên quan đến việc tối ưu hóa một hàm mất mát dựa trên sai số dự đoán trong bài toán hồi quy tuyến tính (Moayeri & Hemami, 2023).

Cho một tập dữ liệu  $[(x_i, y_i)]_{i=1}^N$

với  $x_i$  là vectơ gồm các đặc trưng thứ  $i$  và  $y_i$  là giá trị mục tiêu tương ứng, hàm mục tiêu của LS-SVM cho bài toán hồi quy có thể được biểu diễn như sau:

$$L(w, b, \xi) = \frac{1}{2}w^T w + \gamma \sum_{i=1}^N \xi_i^2 \quad (3)$$

Với điều kiện ràng buộc như sau:  $y_i = w^T \phi(x_i) + b + \xi_i, \quad \forall i = 1, \dots, N$

Trong đó:

- $w$  Vector trọng số, biểu diễn đường phân chia (hay hyperplane) trong không gian nhiều chiều,
- $b$  là sai số ( bias),
- $\xi_i$  là biến sai số (lỗi) cho mỗi điểm dữ liệu,
- Tham số điều chỉnh, tham số này cân bằng giữa việc tối đa hóa margin và giảm thiểu lỗi,
- $\phi(x_i)$  là hàm ánh xạ dữ liệu của vectơ đầu vào  $x$  vào không gian đặc trưng nhiều chiều, nơi mà dữ liệu có thể trở nên tách biệt tuyến tính.

Các tham số  $w$  và  $b$  được tìm bằng cách giải một tập hợp các phương trình tuyến tính, đây là điểm khác biệt chính từ SVM chuẩn nơi một vấn đề lập phương trình bậc hai được giải quyết. Mục tiêu là tìm ra mặt phẳng phân chia phù hợp nhất mà tách biệt dữ liệu trong không gian đặc trưng với lỗi tối thiểu.

Hàm hồi quy LSSVM được mô tả như sau:

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b$$

Trong đó:

- $y(x)$  là giá trị dự đoán cho điểm dữ liệu  $x$ .
- $\alpha_k$  là những nhân tử Lagrange.
- $K(x, x_k)$  là hàm nhân kernel, một hàm đặc biệt dùng để so sánh điểm dữ liệu  $x$  với các điểm dữ liệu khác trong tập huấn luyện  $x_k$ .
- $b$  là sai số (bias)

Với hàm nhân cơ sở hàm Radial (RBF):

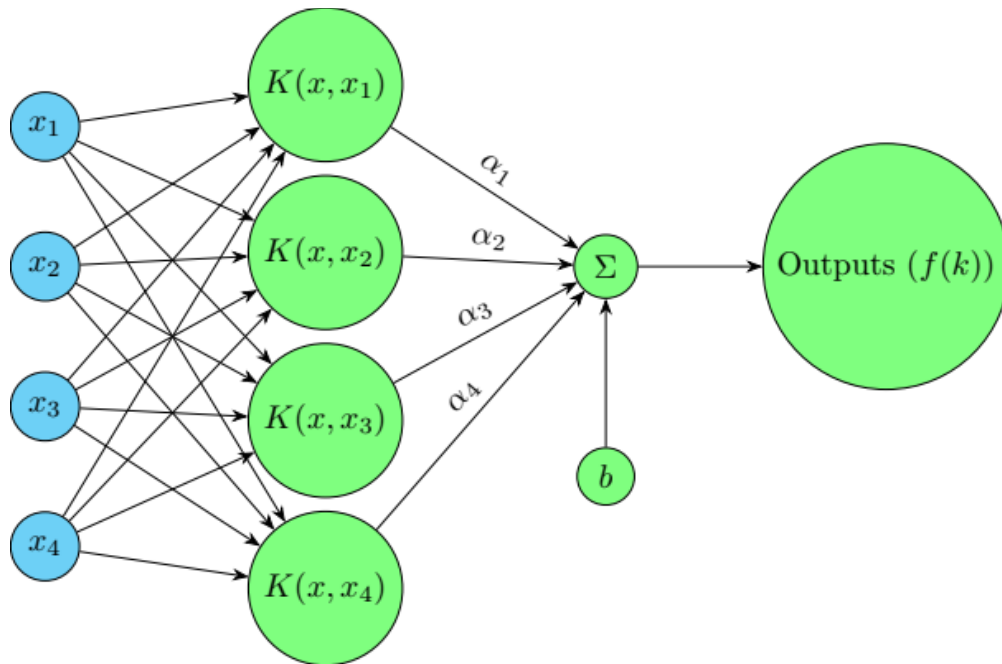
$$K(x_k, x_l) = e^{-\frac{\|x_k - x_l\|^2}{2\sigma^2}}$$

Trong đó:

- $\|x_k - x_l\|^2$  là bình phương khoảng cách Euclidean giữa hai điểm dữ liệu.
- $\sigma^2$  là bình phương độ rộng băng thông của hàm kernel, có thể được tối ưu hóa sử dụng các thuật toán như Genetic Algorithm (GA), Particle Swarm Optimization (PSO), hoặc kết hợp cả hai (HGAPSO).

Tham số  $\gamma$  và  $\sigma^2$  của LSSVM đóng vai trò quan trọng trong việc ảnh hưởng đến độ chính xác dự đoán và khả năng tổng quát hóa của mô hình.

Bài toán tối ưu hóa LS-SVM được giải quyết bằng cách sử dụng phương pháp bình phương nhỏ nhất, làm cho quá trình tính toán trở nên hiệu quả hơn về mặt tính toán, đặc biệt là trong trường hợp của dữ liệu lớn (Valyon & Horváth, 2007). Cấu trúc của mô hình LS SVM được trình bày như hình 3.



Hình 3. Cấu trúc mô hình LS SVM

### 2.3. Các phương pháp đánh giá:

Trong mạng nơ-ron nhân tạo (ANN), có nhiều phương pháp để đánh giá hiệu suất của mô hình. Dưới đây là một số phương pháp đánh giá phổ biến (Hodson, 2022).

+ Sai số căn quân phương (Root Mean Squared Error - RMSE) (Chai & Draxler, 2014; Wang & Lu, 2018).

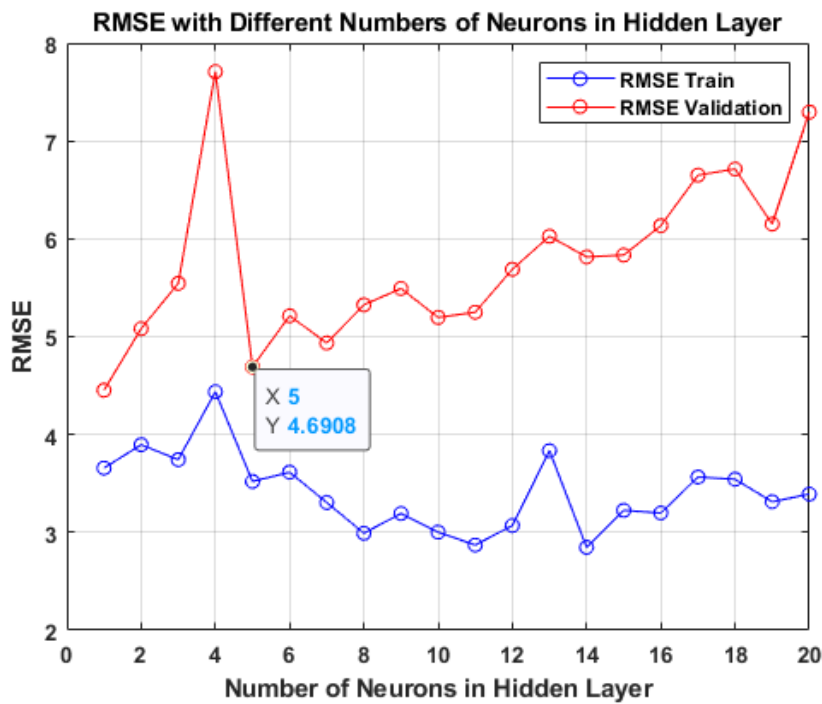
Khi giá trị RMSE càng nhỏ, mô hình hồi quy được coi là có hiệu suất tốt hơn vì nó dự đoán gần giá trị thực tế.

+ Hệ Số  $R^2$  (R-squared) (Barrett, 2000): Hệ số xác định là một thông số thường được sử dụng để đánh giá mức độ phù hợp của một mô hình hồi quy với dữ liệu thực tế. Nó phản ánh tỷ lệ phần trăm biến thiên của biến phụ thuộc được giải thích bởi mô hình hồi quy, thường được giới hạn trong khoảng từ 0 đến 1, giá trị  $R^2$  cao (gần 1) cho thấy rằng mô hình hồi quy đã giải thích được phần lớn biến thiên của dữ liệu, và mô hình phù hợp tốt với dữ liệu quan sát.

## 3. Kết quả và thảo luận

### 3.1. Khảo sát số neuron lớp ẩn:

Số neuron trong lớp ẩn của một mạng nơ-ron nhân tạo (ANN) là một trong những yếu tố quan trọng nhất quyết định cấu trúc và chức năng của mạng. Mỗi neuron trong các lớp ẩn có thể nhận input từ lớp trước đó, áp dụng một hàm kích hoạt, và truyền output của nó đến lớp tiếp theo. Số neuron lớp ẩn trong bài báo này được khảo sát từ 1 đến 20 với 100 lần chạy, dựa trên hình 4 ta thấy số neuron lớp ẩn tối ưu là 5 với giá trị RMSE validation nhỏ nhất. Tại neuron lớp ẩn là 5, giá trị RMSE tập huấn luyện và tập so sánh có độ chênh lệch rất thấp, chứng tỏ mô hình ANN đạt giá trị tối ưu nhất



Hình 4. Số neuron lớp ẩn theo RMSE

3.2. Khảo sát hàm kích hoạt:

Hàm kích hoạt đóng vai trò quan trọng trong các lớp đầu ra của mạng neuron nhân tạo, việc khảo sát hàm kích hoạt tối ưu để mô hình ANN càng trở nên tối ưu giúp hạn chế thời gian khảo sát, ở đây nhóm nghiên cứu sử dụng các hàm Sigmoid(logsig), Purelin, Tanh(tansig), và hàm Softmax, qua khảo sát với 100 lần chạy để tìm ra hàm kích hoạt tối ưu trong cả 2 giai đoạn liên tục là hàm kích hoạt tối ưu cho lớp ẩn (hidden) và lớp đầu ra (output), kết quả theo bảng 2 cho ra cặp hàm kích hoạt lớp ẩn là tansig, và hàm output là purelin có tần suất xuất hiện nhiều nhất(8 lần) với R<sup>2</sup> trên 0.9

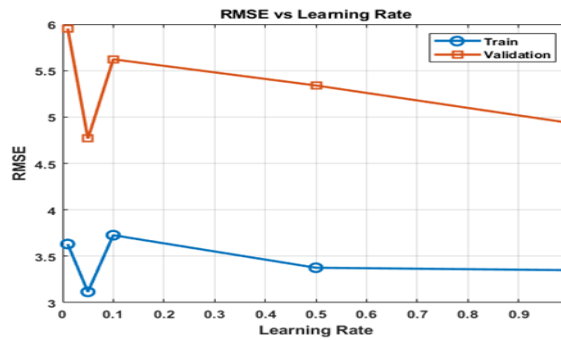
Bảng 2. Khảo sát hàm kích hoạt

Activation								
Hidden	logsig	logsig	logsig	logsig	tansig	tansig	tansig	tansig
Output	logsig	tansig	softmax	purelin	logsig	tansig	softmax	purelin
Số lần xuất hiện với R2 lớn (trên 0.9)	0	5	0	4	0	5	0	8
Activation								
Hidden	softmax	softmax	softmax	softmax	purelin	purelin	purelin	purelin
Output	logsig	tansig	softmax	purelin	logsig	tansig	softmax	purelin
Số lần xuất hiện với R2 lớn (trên 0.9)	0	4	0	5	0	0	0	0

3.3. Khảo sát tốc độ học tập (learning rate):

Tốc độ học tập (learning rate) là một siêu tham số quan trọng trong quá trình huấn luyện mạng nơ-ron nhân tạo (ANN). Hệ số này xác định độ lớn của bước cập nhật được thực hiện trên trọng số của mô hình dựa trên đạo hàm của hàm mất mát. Nếu quá lớn, có thể dẫn đến việc bước cập nhật quá lớn và mô hình có thể không hội tụ. Ngược lại, nếu quá nhỏ, quá trình đào tạo có thể diễn ra rất chậm hoặc mô hình có thể bị mắc kẹt trong các điểm cực tiểu.

learning rate	RMSE_train	RMSE_val
0.01	3.631	5.953
0.05	3.116	4.767
0.1	3.726	5.621
0.5	3.376	5.339
1	3.349	4.936



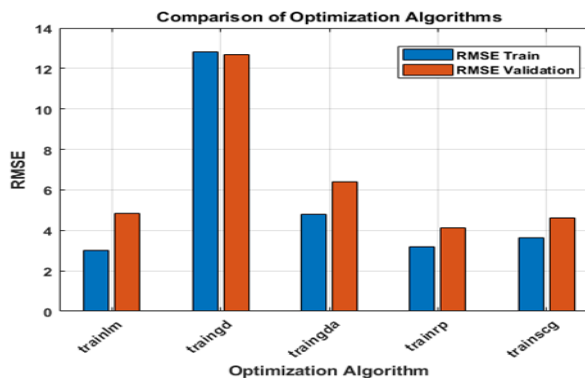
Hình 5. Tốc độ học tập theo RMSE

Dựa vào kết quả hình 5 trên, ta thấy ở tốc độ học tập là 0.05 cho kết quả RMSE của tập train và tập validation thấp nhất. Ở biểu đồ hình 5, ta cũng thấy rõ sự phụ thuộc của RMSE theo tốc độ khá tương đồng giữa tập huấn luyện và tập kiểm định.

**3.4. Khảo sát thuật toán huấn luyện:**

Thuật toán huấn luyện trong ANN là một hàm mục tiêu để điều chỉnh các trọng số của mạng nơ-ron (Fan và cs., 2022). Mục tiêu là làm cho đầu ra của mạng nơ-ron càng gần với đầu ra thực tế nhất có thể. Thuật toán huấn luyện đo lường sự chênh lệch giữa dự đoán của mô hình và giá trị thực tế, với các hàm huấn luyện như hình 6, ta thấy thuật toán huấn luyện ‘trainrp’ có giá trị RMSE trên tập kiểm định nhỏ nhất, chênh lệch giữa giá trị RMSE của tập huấn luyện và tập kiểm định cũng nhỏ nhất, điều này cũng thể hiện rõ qua đồ thị hình 6

Hàm huấn luyện	RMSE_train	RMSE_val
trainlm	3.019	4.841
traingd	12.814	12.688
traingda	4.791	6.380
trainrp	3.187	4.123
trainscg	3.644	4.615

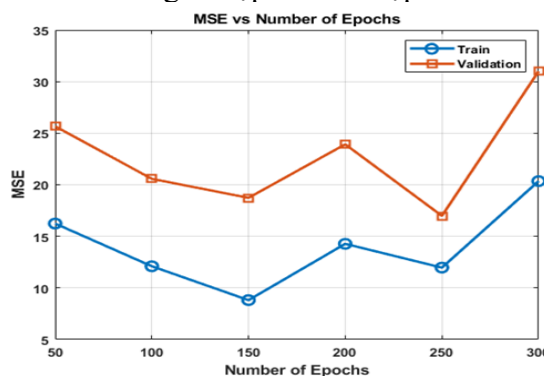


Hình 6. Hàm huấn luyện theo RMSE

**3.5. Khảo sát số epoch:**

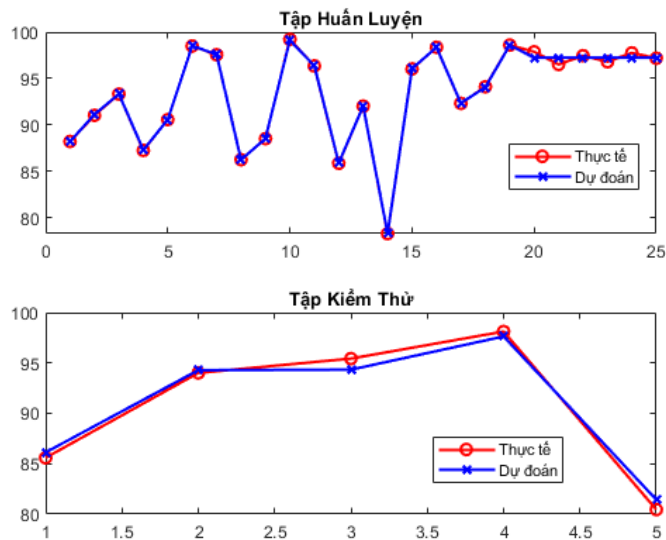
Epoch trong mạng nơ-ron nhân tạo (ANN) đề cập đến số lần toàn bộ tập dữ liệu được sử dụng để huấn luyện mạng. Mỗi epoch bao gồm một quá trình lan truyền tiến qua mạng để đưa ra dự đoán, tiếp theo là một quá trình lan truyền ngược để cập nhật trọng số dựa trên lỗi giữa dự đoán và kết quả thực tế. Trong bài báo này epoch Khảo sát từ 1 đến 300 với 50 lần chạy, dựa vào hình 7 ta thấy ở epoch 150 có sự gần nhau nhất giữa tập train và tập validation.

Epoch huấn luyện	MSE_train	MSE_val
50	16.234	25.662
100	12.096	20.569
150	8.814	18.729
200	14.270	23.919
250	11.965	16.982
300	20.348	30.991



Hình 7. Số epoch theo MSE

Với các thông số tối ưu trên, ta tiếp tục chạy huấn luyện và tìm ra được các chỉ số đánh giá mô hình tốt nhất (cho biết gần chính xác nhất sự phụ thuộc của các biến đầu vào X vào biến đầu ra Y).



**Hình 8.** So sánh giá trị thực tế và dự đoán với tập train và tập validation theo mô hình ANN

Kết quả model ANN với các thông số đã tối ưu như sau: Số neuron lớp ẩn tối ưu là 5, số epoch tối ưu là 150, hệ số học tập tối ưu là 0.05, hàm kích hoạt tối ưu là tansig và purelin.

Hình 8 trên, với biểu đồ "Tập Huấn Luyện" thể hiện mô hình có khả năng theo dõi chặt chẽ sự biến động của dữ liệu thực tế, với giá trị dự đoán phản ánh khá trung thực các đỉnh của chuỗi dữ liệu. Điều này cho thấy mô hình có thể đã học rất tốt từ dữ liệu huấn luyện. Tuy nhiên, khi chuyển sang biểu đồ "Tập Kiểm Thử", mô hình cho thấy dấu hiệu không tổng quát hóa đủ tốt. Cả hai đường giá trị thực tế và dự đoán đều bắt đầu ở mức cao và theo xu hướng giảm dần. Mô hình dự đoán khá phù hợp với giá trị thực tế cho đến giữa chuỗi, sau đó bắt đầu xuất hiện sự chênh lệch. Mô hình ANN phản ánh chính xác dữ liệu trong quá trình huấn luyện nhưng vẫn chưa tương thích tốt với dữ liệu kiểm thử.

**Bảng 3.** Kết quả mô hình ANN

Models	RMSE_train	RMSE_val	R <sup>2</sup> _train	R <sup>2</sup> _val	Độ hấp phụ MB (Y)
ANN	0.2517	0.7447	0.9977	0.9874	90.76%

Kết quả chạy sau khi tối ưu mô hình ANN cho thấy R<sup>2</sup> đã cải thiện rất nhiều, với giá trị 0.9977, mô hình giải thích được 99.77% sự biến thiên của dữ liệu trong tập huấn luyện (train). Với 0.9874 trên tập kiểm thử (val) cũng cao, cho thấy mô hình cũng có sự phù hợp tốt với dữ liệu kiểm thử, sai số căn phương RMSE thấp, tuy nhiên việc RMSE tăng đáng kể từ tập huấn luyện sang tập kiểm thử có thể là dấu hiệu của overfitting, nghĩa là mô hình có thể đã học quá mức những đặc điểm cụ thể của tập huấn luyện mà không có khả năng tổng quát hóa tốt trên dữ liệu mới.

Với mô hình LS-SVM đã tối ưu :

**Bảng 4.** Kết quả mô hình Ls-Svm

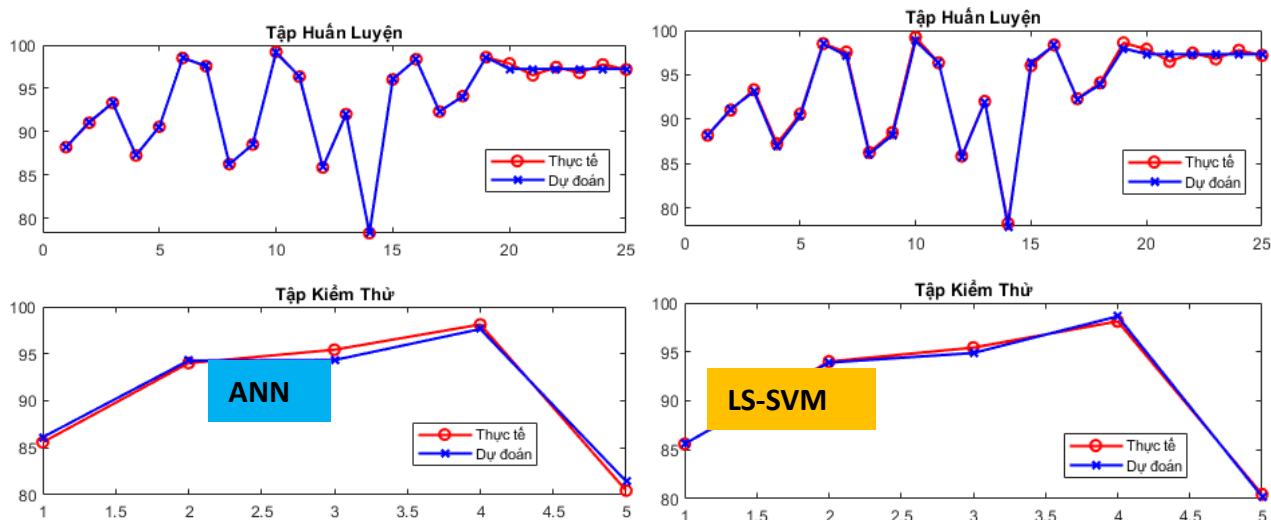
Models	RMSE_train	RMSE_val	R <sup>2</sup> _train	R <sup>2</sup> _val	Độ hấp phụ MB (Y)
ANN	0.2517	0.7447	0.9977	0.9874	90.76%
LS-SVM	0.4677	0.2477	0.9919	0.9986	97.17%

Qua bảng 4 trên, nhận thấy rằng:

- ANN dường như bị overfitting, nghĩa là nó khớp quá chặt với dữ liệu huấn luyện, làm giảm khả năng dự đoán chính xác trên dữ liệu mới.

• LS-SVM, mặc dù có hiệu suất kém hơn một chút trên tập huấn luyện so với ANN, nhưng lại cho kết quả tốt hơn trên tập xác thực, cho thấy mô hình này có khả năng tổng quát hóa tốt hơn.

Hình 9 cũng cho thấy rằng mô hình LS-SVM đã khắc phục hoàn toàn những điểm dữ liệu chưa tốt, giúp cải thiện quá trình huấn luyện và làm giảm đáng kể độ chênh lệch giữa các giá trị dự đoán và giá trị thực nghiệm, với những thông số trên ta thấy mô hình LS-SVM đã cho kết quả huấn luyện tối ưu hơn

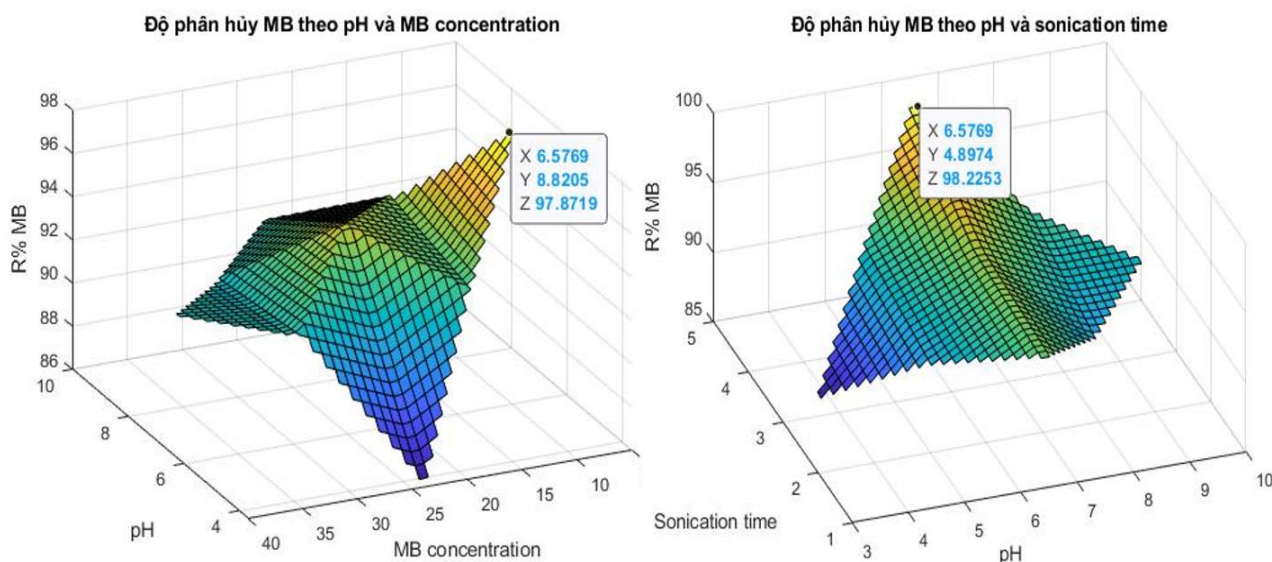


Hình 9. Kết quả dự đoán và thực tế của mô hình ANN và mô hình LS-SVM

### 3.6. Dự đoán dữ liệu đầu vào:

#### Dự đoán pH tối ưu:

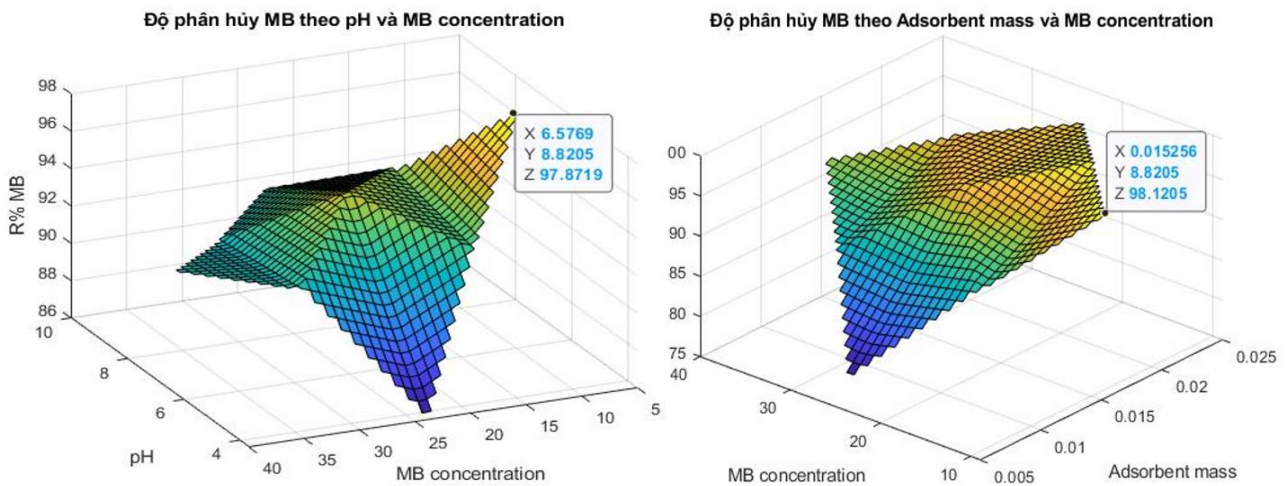
Đồ thị hình 10, khảo sát phần trăm hấp phụ MB (R%) theo lượng hấp phụ và độ pH cũng như thời gian siêu âm, kết quả ta thấy pH = 6.6 cho phần trăm hấp phụ trên 97%



Hình 10. Dự đoán pH tối ưu

#### Dự đoán nồng độ MB tối ưu:

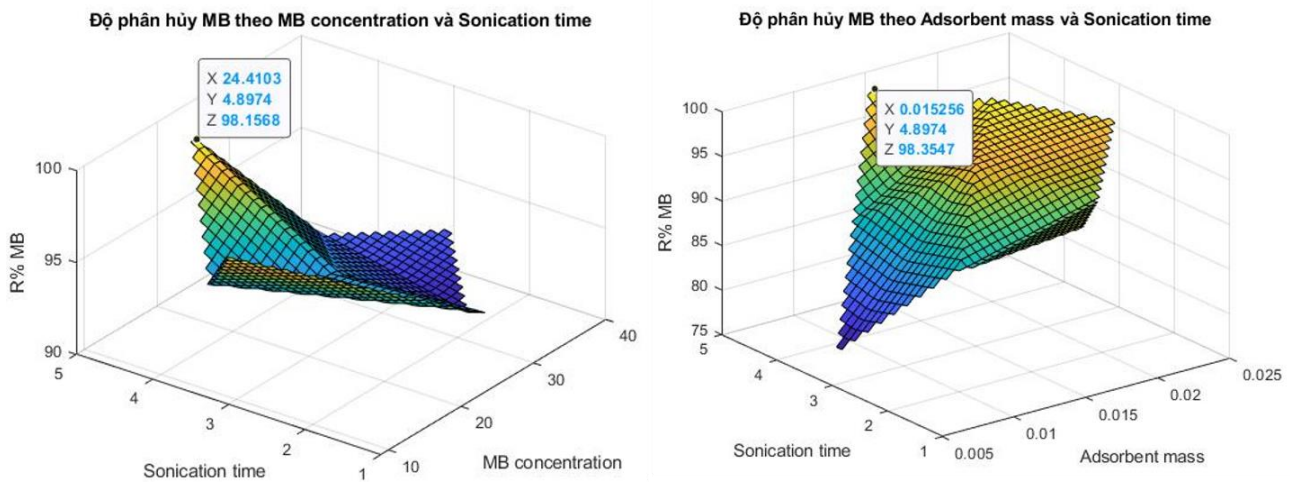
Đồ thị hình 11, khảo sát phần trăm hấp phụ MB theo lượng hấp phụ, nồng độ MB và độ pH, với nồng độ MB là 8.8mg/L với khả năng hấp phụ MB trên 97%.



Hình 11. Dự đoán MB tối ưu

**Dự đoán thời gian siêu âm tối ưu:**

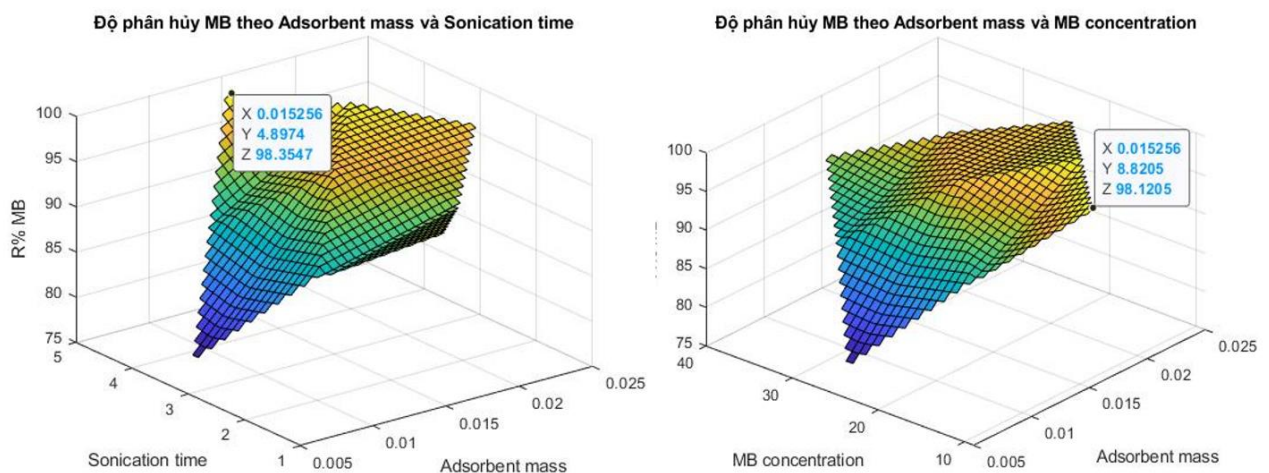
Đồ thị hình 12, khảo sát phần trăm hấp phụ MB theo thời gian siêu âm, nồng độ MB, khối lượng hấp phụ, kết quả cho thấy với thời gian siêu âm là 4.9 phút cho phần trăm hấp phụ đạt trên 97%



Hình 12. Dự đoán thời gian siêu âm tối ưu

**Dự đoán khối lượng chất hấp phụ:**

Đồ thị hình 13, khảo sát phần trăm hấp phụ MB theo khối lượng chất hấp phụ, thời gian siêu âm, nồng độ MB, kết quả cho thấy với khối lượng hấp phụ là 0.015g cho phần trăm hấp phụ đạt trên 97%.



Hình 13. Dự đoán khối lượng chất hấp phụ

Với các kết quả trên, ta có thể dự đoán với các thông số đầu vào như sau: pH là 6.6, khối lượng chất hấp phụ là 0.015g, nồng độ MB là 8.8mg/L, và thời gian siêu âm là 4.9 phút có thể cho phân trăm hấp phụ MB cao nhất trên 97%

#### 4. Kết luận

Nghiên cứu này đã chứng minh hiệu quả của việc sử dụng ANN và LS-SVM trong tối ưu hóa quá trình hấp phụ MB, từ đó đề xuất một phương pháp mới mẽ và hiệu quả cho việc xử lý ô nhiễm nước. Với phương pháp ANN cho kết quả hấp phụ MB trên 90%, với LS-SVM đã cải thiện khả năng hấp phụ MB đáng kể với trên 97%, kết quả  $R^2$  đã cải thiện rất nhiều, đạt giá trị cao trên 0.99. Đồng thời qua mô hình này cũng đã dự đoán được các thông số đầu vào tối ưu như độ pH là 6.6, khối lượng chất hấp phụ là 0.015g, nồng độ MB là 8.8mg/L, thời gian siêu âm là 4.9 phút. Dựa vào mô hình tối ưu đã nghiên cứu trên, chúng ta có thể mở rộng ra, đưa vào thực tế nhằm giải quyết bài toán khử chất độc hại MB trong các khu công nghiệp, khu chế xuất... một cách hiệu quả nhằm bảo vệ sức khỏe cho người lao động.

#### TÀI LIỆU THAM KHẢO

- [1] Abd-Alkhalig Salih Mijwel, Ali Najah Ahmed, Haitham Abdulmohsin Afan, Haiyam Mohammed Alayan, Mohsen Sherif & Ahmed Elshafie (2023). Artificial intelligence models for methylene blue removal using functionalized carbon nanotubes. *Scientific Reports*, 13. <https://www.nature.com/articles/s41598-023-45032-3>
- [2] Ahad Amini Pishro, Shiquan Zhang, Dengshi Huang, Feng Xiong, WeiYu Li & Qihong Yang (2021). Application of artificial neural networks and multiple linear regression on local bond stress equation of UHPC and reinforcing steel bars. *Scientific Reports*, 11. <https://www.nature.com/articles/s41598-021-94480-2>
- [3] Asadollahfardi, G. (2015). Artificial Neural Network. In G. Asadollahfardi (Ed.), *Water Quality Management: Assessment and Interpretation* (pp. 77-91). Springer. [https://doi.org/10.1007/978-3-662-44725-3\\_5](https://doi.org/10.1007/978-3-662-44725-3_5)
- [4] Asfaram, A., Ghaedi, M., Ahmadi Azqhandi, M. H., Goudarzi, A., & Dastkhooon, M. (2016). Statistical experimental design, least squares-support vector machine (LS-SVM) and artificial neural network (ANN) methods for modeling the facilitated adsorption of methylene blue dye. *RSC Advances*, 6(46), 40502-40516. <https://doi.org/10.1039/C6RA01874B>
- [5] Barrett, G. B. (2000). The Coefficient of Determination: Understanding r squared and R squared. *The Mathematics Teacher*, 93(3), 230-234. <https://doi.org/10.5951/MT.93.3.0230>
- [6] Braspenning, P. J., Thuijsman, F., & Weijters, A. J. M. M. (1995). *Artificial Neural Networks: An Introduction to ANN Theory and Practice*. Springer Science & Business Media.
- [7] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- [8] Chengbing Yu, Ziwei Xi, Yilin Lu, Kaixin Tao, Zhong Yi (2020). K/S value prediction of cotton fabric using PSO-LSSVM, Sage journals, 90(23-24). <https://journals.sagepub.com/doi/abs/10.1177/0040517520924750>
- [9] Chicco D, Warrens MJ, Jurman G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7:e623 <https://doi.org/10.7717/peerj-cs.623>
- [10] Fan, G., A.s., E.-S., Eftekhari, S. A., Hekmatifar, M., Toghraie, D., Mohammed, A. S., & Khan, A. (2022). A well-trained artificial neural network (ANN) using the trainlm algorithm for predicting the rheological behavior of water – Ethylene glycol/WO<sub>3</sub> – MWCNTs nanofluid. *International Communications in Heat and Mass Transfer*, 131, 105857. <https://doi.org/10.1016/j.icheatmasstransfer.2021.105857>
- [11] Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481-5487. <https://doi.org/10.5194/gmd-15-5481-2022>

- [12] Moayeri, M. M., & Hemami, M. (2023). Solving Partial Differential Equations by LS-SVM. In J. A. Rad, K. Parand, & S. Chakraverty (Eds.), *Learning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications*, tr. 171-198. Springer Nature. [https://doi.org/10.1007/978-981-19-6553-1\\_8](https://doi.org/10.1007/978-981-19-6553-1_8)
- [13] Palani, S., Liong, S.-Y., & Tkalich, P. (2008). An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), 1586-1597. <https://doi.org/10.1016/j.marpolbul.2008.05.021>
- [14] Seok, K. H. (2010). Semi-supervised classification with LS-SVM formulation. *Journal of the Korean Data and Information Science Society*, 21(3), 461-470.
- [15] Sina Maghsoudy, Pouya Zakerabbasi, Alireza Baghban, Amin Esmaeili & Sajjad Habibzadeh (2024). Connectionist technique estimates of hydrogen storage capacity on metal hydrides using hybrid GAPSO-LSSVM approach. *Scientific Reports*, 14. <https://www.nature.com/articles/s41598-024-52086-4>
- [16] Suykens, J., & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, 9, 293-300. <https://doi.org/10.1023/A:1018628609742>
- [17] Vallyon, J., & Horváth, G. (n.d.). Extended Least Squares LS-SVM, 3(3).
- [18] Wang, H., & Hu, D. (2005). Comparison of SVM and LS-SVM for Regression. *2005 International Conference on Neural Networks and Brain*, 1, 279-283. <https://doi.org/10.1109/ICNNB.2005.1614615>
- [19] Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1), 012049. <https://doi.org/10.1088/1757-899X/324/1/012049>