

# MỤC LỤC

## KHOA HỌC CƠ BẢN, KỸ THUẬT VÀ CÔNG NGHỆ

### Công nghệ thông tin

- ♦ **Lâm Sinh Công:** TỐI ƯU NGUỒN PHÁT HIỆN KÊNH TRỐNG CHO MẠNG WIFI SỬ DỤNG HỌC TĂNG CƯỜNG 3
- ♦ **Nguyễn Thái Dương, Nguyễn Ngọc Tân:** PHÂN TÍCH SỰ ĐÁNH ĐỔI GIỮA HIỆU NĂNG VÀ NĂNG LƯỢNG TRONG MẠNG TRUYỀN THÔNG TÁN XẠ NGƯỢC MÔI TRƯỜNG CHỐNG NHIỀU SỬ DỤNG HỌC TĂNG CƯỜNG ĐA TÁC TỬ 11
- ♦ **Lê Hoàng Vũ, Trần Hà Phương, Trần Thị Thu Phương, Hà Quang Hưng, Phạm Minh Triển, Vũ Minh Trung:** NGHIÊN CỨU PHÁT TRIỂN HỆ THỐNG PHẦN MỀM NHẬN DIỆN SỐ LƯỢNG BỌ HÀ TRONG BÃY PHỤC VỤ QUẢN LÝ SÂU BỆNH HẠI TRÊN CÂY KHOAI LANG 19
- ♦ **Hà Trọng Thắng:** ENHANCING NETWORK SECURITY BASED ON ANOMALY DETECTION USING DEEP LEARNING FOR INTELLIGENT IDS/IPS SYSTEMS 29
- ♦ **Nguyễn Thị Thúy Nga:** TRANSFORMING BUSINESS EDUCATION IN IT: THE ROLE OF MIS IN SHAPING FUTURE LEADERS 43

### Điện - Điện tử - Tự động hóa

- ♦ **Phan Xuân Minh:** SỬ DỤNG THUẬT TOÁN QP CỦA BÀI TOÁN TỐI ƯU TOÀN PHƯƠNG ĐỂ XÂY DỰNG BỘ ĐIỀU KHIỂN MPC CHO HỆ PHI TUYẾN 55
- ♦ **Nguyen Quang Binh, Hoang Van Quyet, Le Van Tuan:** ROBUST STABILITY ANALYSIS AND REAL-TIME SPEED TRACKING CONTROL OF STEPPER MOTORS VIA MATLAB/OPC-PLC INTEGRATION WITH SMITH PREDICTOR 64
- ♦ **Lê Bá Yến, Vũ Quốc Huy:** ĐIỀU KHIỂN THÍCH NGHI PHI MÔ HÌNH CFDL-MFAC BÙ TRỄ CHO MỘT LỚP HỆ PHI TUYẾN SISO RỜI RẠC CÓ TRỄ ĐẦU VÀO BIẾN THIÊN 75

## **Cơ khí - Động lực**

- ♦ **K.Q.Le, J.W.Chew, M.F.Leyva-Mendivil, T.H.Thai, F.Duan:** 84  
SIMULATION OF FIBER DEPOSITION ON 3D PRINTED MOLDED  
FIBER SCREEN USING MULTI-SPHERE DISCRETE ELEMENT  
METHOD
- ♦ **Bùi Vũ Hùng, Bùi Văn Hưng, Đỗ Trung Hiếu, Đỗ Đức Long, Dương  
Công Trung:** 89  
NGHIÊN CỨU TÍNH TOÁN, THIẾT KẾ VÀ CHẾ TẠO  
MÁY CẮT TIA NƯỚC CÔNG SUẤT NHỎ CHO VẬT LIỆU MỀM
- ♦ **Bùi Quang Toàn:** 96  
ĐỀ XUẤT ỨNG DỤNG CÔNG NGHỆ TRÍ TUỆ NHÂN  
TẠO (AI) TRONG HƯỚNG DẪN THỰC HÀNH HỌC PHẦN AUTOCAD  
CHO SINH VIÊN TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á

## **KHOA HỌC KINH TẾ VÀ XÃ HỘI**

### **Kinh tế**

- ♦ **Đỗ Đức Bình:** 103  
NÂNG CAO SẢN XUẤT TINH GỌN TẠI CÔNG TY  
TNHH SẢN XUẤT HÀNG MAY MẶC ESQUEL VIỆT NAM HÒA BÌNH

### **Du lịch**

- ♦ **Lê Thị Hoài Phương:** 119  
DU LỊCH VĂN HÓA VÀ NGHỆ THUẬT BIỂU  
DIỄN: MỐI QUAN HỆ NHÌN TỪ GÓC ĐỘ NGÀNH CÔNG NGHIỆP  
VĂN HÓA
- ♦ **Nguyễn Thị Thu Trang, Lê Mạnh Hùng:** 129  
PHÁT TRIỂN DU LỊCH VĂN  
HÓA TRONG HỆ SINH THÁI CÔNG NGHIỆP VĂN HÓA VIỆT NAM

## **KHOA HỌC GIÁO DỤC**

### **Giáo dục**

- ♦ **Đinh Thị Phương Thảo:** 137  
PERSONAL FACTORS AND TEACHER  
STRESS: IMPLICATIONS FOR EFL LECTURERS IN HIGHER  
EDUCATION

# TỐI ƯU NGƯỠNG PHÁT HIỆN KÊNH TRỐNG CHO MẠNG WIFI SỬ DỤNG HỌC TĂNG CƯỜNG

Lâm Sinh Công

Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

Email: congl@vnu.edu.vn

## TÓM TẮT

Sự gia tăng nhanh chóng của các thiết bị không dây và dịch vụ băng thông cao đã thúc đẩy xu hướng triển khai dày đặc các điểm truy cập WiFi, dẫn đến nhiều nghiêm trọng giữa cùng dịch vụ. Khi đó, kỹ thuật đánh giá kênh trống (Clear Channel Assessment - CCA) được đề xuất trong hệ thống chuẩn IEEE 802.11 để đánh giá nhiễu từ các trạm lân cận và ước lượng hiệu năng quá trình truyền tin. Bài báo này đề xuất phương pháp kết hợp thuật toán Double Deep Q-Network (Double DQN) với kỹ thuật phân bổ công suất Water-Filling nhằm tối ưu hóa đồng thời ngưỡng phát hiện sự chồng lấn giữa các vùng dịch vụ cơ bản (Overlapping Basic Service Set - OBSS) và công suất phát trong mạng WiFi. Trong đó, tác tử học tăng cường sâu đóng vai trò điều chỉnh ngưỡng đánh giá kênh trống (ngưỡng CCA) cho trong các vùng dịch vụ (Basic Service Set - BSS), trong khi Water-Filling phân bổ công suất tối ưu dựa trên trạng thái kênh tức thời. Kết quả mô phỏng cho thấy phương pháp đề xuất đạt mức cải thiện thông lượng mạng trung bình khoảng 30% so với cơ chế tiêu chuẩn của IEEE 802.11ax.

**Từ khóa:** Double DQN, IEEE 802.11ax, Water-Filling

## ABSTRACT

The rapid growth in the number of wireless devices and high-speed services has driven the dense deployment of WiFi access points, which results in a surge of interference. To address this issue, the Clear Channel Assessment (CCA) technique was introduced in the IEEE 802.11 standard to evaluate interference from neighboring stations and estimate transmission performance. This paper proposes a method that combines the Double Deep Q-Network (Double DQN) with the Water-Filling power allocation technique to jointly optimize the detection threshold for the Overlapping Basic Service Set (OBSS) and the transmit power in WiFi networks. In this approach, the DQN agent adjusts the CCA threshold for each Basic Service Set (BSS), while the Water-Filling algorithm allocates optimal transmit power based on instantaneous channel conditions. Simulation results show that the proposed method improves the average network throughput by approximately 30% compared with the standard IEEE 802.11ax mechanism.

**Keyword:** Double DQN, IEEE 802.11ax, Water-Filling

Ngày nhận bài: 10/03/2026; Ngày sửa bài: 27/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. MỞ ĐẦU

Trong bối cảnh sự bùng nổ của các dịch vụ đòi hỏi băng thông lớn như truyền phát video độ phân giải cao, thực tế ảo/tăng cường (VR/AR), nhu cầu về hiệu suất mạng không dây cục bộ (WLAN) ngày càng tăng cao [1]–[3]. Để đáp ứng nhu cầu này, các điểm truy cập (Access Point - AP) được triển khai ngày càng dày đặc, tạo thành các mạng WLAN dày đặc (Dense WLAN). Tuy nhiên, khi mật độ AP tăng, khả năng chống lán của vùng dịch vụ (BSS) ngày càng cao. Điều này gây ra sự gia tăng mạnh của nhiễu đồng kênh và làm suy giảm hiệu suất mạng tổng thể [4], [5].

Chuẩn IEEE 802.11ax (WiFi 6) đã giới thiệu cơ chế tái sử dụng không gian (Spatial Reuse - SR) nhằm tăng số lượng trạm được phép tuyên dẫn đồng thời trong các vùng dịch vụ chồng lán (OBSS) [6]. Cơ chế SR hoạt động thông qua hai kỹ thuật chính: phát hiện vùng chồng lán OBSS dựa trên ngưỡng công suất (OBSS/PD-based SR) và tái sử dụng không gian tham số hóa (PSR). Trong đó, OBSS/PD-based SR cho phép thiết bị truyền đồng thời khi công suất tín hiệu nhận được từ vùng dịch vụ lân cận thấp hơn ngưỡng OBSS/PD cho trước, thay vì trì hoãn truyền dẫn như cơ chế truyền thống [7]. Tuy nhiên, chuẩn IEEE 802.11ax chỉ quy định giá trị biên cho ngưỡng OBSS/PD mà không cung cấp hướng dẫn cụ thể về cách tối ưu hóa tham số này [8]. Việc lựa chọn ngưỡng không phù hợp có thể dẫn đến hai hệ quả trái ngược: ngưỡng quá cao làm tăng nhiễu do nhiều truyền dẫn đồng thời, còn ngưỡng quá thấp hạn chế khả năng tái sử dụng không gian.

Nhiều công trình nghiên cứu đã đề xuất các phương pháp tối ưu tham số SR trong WLAN. Thuật toán COST (Control OBSS/PD Sensitivity Threshold) [9] điều chỉnh ngưỡng OBSS/PD dựa trên mức nhiễu và cường độ tín hiệu nhận được, đạt mức cải thiện thông lượng lên tới

57% so với cấu hình OBSS/PD cố định. Tuy nhiên, COST hoạt động theo nguyên tắc phản ứng tĩnh, không có khả năng học hỏi và thích ứng với sự biến đổi phức tạp của môi trường mạng. Trong [10], tác giả đề xuất phương pháp tập trung dựa trên Multi-Armed Bandit (MAB) kết hợp Gaussian Mixture Sampling để tìm kiếm cấu hình TX\_PWR và OBSS/PD tối ưu cho toàn bộ WLAN. Phương pháp này cho thấy hiệu quả trong việc phân bổ công bằng tài nguyên vô tuyến, nhưng đòi hỏi không gian khám phá lớn khi số lượng AP tăng.

Với sự phát triển của học tăng cường sâu (Deep Reinforcement Learning - DRL), nhiều nghiên cứu gần đây đã áp dụng công cụ này vào bài toán tối ưu SR. Szott và cộng sự [11] đã trình bày tổng quan về việc ứng dụng học máy để cải thiện hiệu suất IEEE 802.11, trong đó nhấn mạnh tiềm năng của DRL trong tối ưu tái sử dụng không gian và quản lý nhiễu. Tài liệu [12] đề xuất sử dụng Deep Q-Network (DQN) với hai chế độ học để lựa chọn đồng thời ngưỡng CCA và công suất phát. Kết quả mô phỏng chỉ ra rằng phương pháp đề xuất có thể đạt mức cải thiện thông lượng trung bình 62.4% so với thuật toán điều khiển truyền thống. Nhóm tác giả Jung và Baik [13] đề xuất sơ đồ OTOP tối ưu công suất phát cho OBSS/PD-based SR dựa trên phân tích hình học ngẫu nhiên, cho phép tối đa hóa xác suất truyền thành công. Gần đây, nghiên cứu tái sử dụng không gian phối hợp đa trạm thu phát cho mạng sau thế hệ WiFi 7 bắt đầu mở ra hướng tiếp cận mới về phối hợp tập trung giữa các trạm [14].

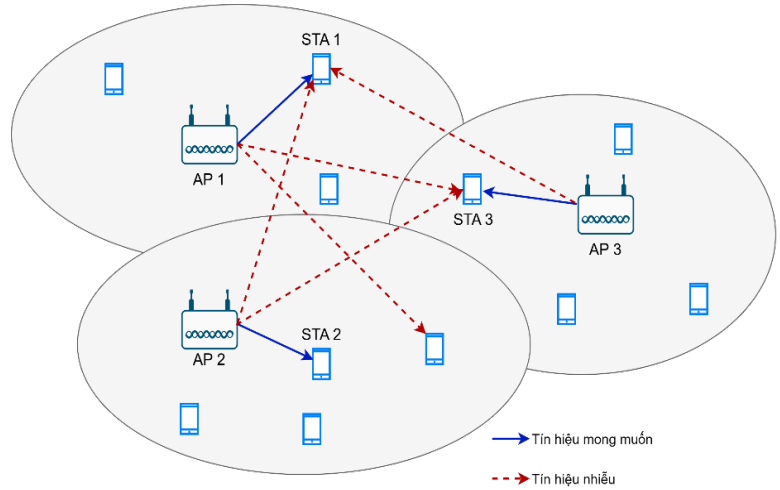
Mặc dù các nghiên cứu nêu trên đã đạt được kết quả đáng kể, hầu hết chỉ tập trung vào việc điều chỉnh riêng lẻ ngưỡng CCA hoặc công suất phát, mà chưa kết hợp đồng thời tối ưu cả hai tham số. Việc sử dụng học máy để tối ưu đồng thời ngưỡng CCA và công suất phát như trong nghiên cứu [12] có thể dẫn đến sự

mở rộng của không gian hành động và làm cho quá trình hội tụ của chương trình học máy diễn ra chậm hơn. Ngoài ra, DQN chỉ làm việc với các hành động rời rạc, trong khi dải công suất khả thi của các trạm phát là giá trị liên tục. Do đó, việc sử dụng DQN để giải bài toán công suất tối ưu có thể chưa phù hợp.

Bài báo này đề xuất phương pháp kết hợp Double DQN với kỹ thuật Water-Filling để đồng thời tối ưu ngưỡng OBSS/PD và phân bổ công suất phát trong mạng WiFi dày đặc. Cụ thể, DQN đảm nhận vai trò tối ưu tham số CCA rời rạc và Water-Filling giải bài toán phân bổ công suất liên tục. Ngoài ra, phương pháp đề xuất thiết kế hàm thưởng đa thành phần giúp cân bằng giữa thông lượng mạng, quản lý nhiễu, hiệu quả năng lượng và đảm bảo chất lượng dịch vụ (QoS). Trong bài báo này, mô hình hệ thống được trình bày trong phần 2. Phần 3 trình bày thuật toán tối ưu. Kết quả phân tích hiệu năng được trình bày trong phần 4. Các kết luận được trình bày trong phần cuối cùng.

## 2. MÔ HÌNH HỆ THỐNG

Bài báo nghiên cứu mô hình mạng WiFi dày đặc bao gồm  $B$  AP được bố trí cố định trong vùng dịch vụ. Trong đó, mỗi AP quản lý một BSS và phục vụ  $K$  người dùng (Station - STA) phân bố ngẫu nhiên trong bán kính phủ sóng của AP đó như được mô tả trong Hình 1. Trong mô hình này, các AP hoạt động trên cùng dải tần số và tạo thành các vùng dịch vụ chồng lấn OBSS. Khi đó, mỗi STA nhận tín hiệu mong muốn từ một AP nhưng chịu nhiễu từ các AP lân cận.



Hình 1: Mô hình hệ thống

Mô hình suy hao đường truyền được xây dựng dựa trên khuyến nghị ITU-R P.1238 cho môi trường trong nhà [15], bao gồm thành phần suy hao theo khoảng cách và hiệu ứng che khuất (shadowing). Cụ thể, suy hao đường truyền (tính bằng dB) giữa thiết bị phát và thu cách nhau khoảng cách được tính theo công thức sau:

$$PL(d) = 40.05 + 20 \log_{10} \left( \frac{f_c}{2.4} \right) + 20 \log_{10} (\min(d, d_{bp})) + 35 \log_{10} \left( \frac{d}{d_{bp}} \right) \mathbb{1}(d > d_{bp}) + L_w + X_\sigma \quad (1)$$

trong đó  $f_c$  là tần số sóng mang,  $d_{bp}$  là khoảng cách điểm gãy (breakpoint distance),  $L_w$  là suy hao do vật cản,  $X$  là biến ngẫu nhiên Gauss với độ lệch chuẩn  $\sigma$ . Hàm chỉ thị  $\mathbb{1}(\cdot)$  nhận giá trị 1 khi điều kiện đúng và 0 trong trường hợp ngược lại.

Mô hình truy nhập kênh CSMA/CA xác định tính được áp dụng. Cụ thể, AP  $b$  được phép truyền (active) khi và chỉ khi tổng nhiễu từ các AP lân cận thấp hơn ngưỡng CCA tương ứng:

$$\text{active}(b) = \mathbb{1} \left( \sum_{b' \neq b} P_{max} \cdot G(b, b') < CCA_w(b) \right) \quad (2)$$

trong đó  $G(b, b')$  là độ lợi kênh giữa cặp AP  $(b, b')$ ,  $P_{max}$  công suất phát tối đa, và  $CCA_w(b)$  là ngưỡng CCA của AP  $b$  tính bằng watt. Ràng buộc này đảm bảo rằng chỉ những AP không gây nhiễu

vượt ngưỡng mới được truyền đồng thời, phản ánh đúng nguyên tắc hoạt động của OBSS/PD-based SR.

Khi AP  $b$  được kích hoạt truyền, tỷ số tín hiệu trên nhiễu cộng tạp âm (SINR) được tính:

$$\gamma_b = \frac{p_b h_{b,b}}{\sum_{b' \neq b, \text{active}(b')} p_{b'} h_{b,b'} + \sigma^2} \quad (3)$$

trong đó  $p_{b'}$  là công suất phát của AP  $b$  dành cho STA  $b'$ ;  $h_{b,b}$  là độ lợi kênh trung bình từ AP  $b$  đến các STA  $b$ ;  $h_{b,b'}$  là độ lợi kênh nhiễu từ AP  $b'$  gây ra tại các STA của AP  $b$ , và  $\sigma^2$  là công suất tạp âm nhiệt.

Thông lượng của mỗi BSS và thông lượng tổng của mạng được xác định theo định lý Shannon:

$$R_b = W \cdot \log_2(1 + \gamma_b), R = \sum_{b=1}^B R_b \quad (4)$$

trong đó  $W$  là băng thông kênh.

Khi đó, bài toán tối ưu tổng thể được phát biểu như sau:

$$\max_{CCA, p} R = \sum_{b=1}^B R_b \quad (5)$$

với các ràng buộc:

$$(a) \quad R_b \geq R_{min}, \forall b \quad (6a)$$

$$(b) \quad 0 \leq p_b \leq P_{max}, \forall b \quad (6b)$$

$$(c) \quad CCA_b \in \mathcal{C}, \forall b \quad (6c)$$

trong đó ràng buộc (6a) đảm bảo QoS tối thiểu cho mỗi BSS, (6b) giới hạn công suất phát, và (6c) yêu cầu ngưỡng CCA thuộc tập rời rạc cho phép  $\mathcal{C}$ . Bài toán (6) là bài toán tối ưu hỗn hợp rời rạc-liên tục, trong đó biến CCA rời rạc và biến công suất liên tục. Hơn nữa, quyết định CCA của các BSS ảnh hưởng qua lại thông qua mô hình CSMA/CA trong biểu thức (2), khiến bài toán trở thành NP-hard. Phần tiếp theo trình bày thuật toán đề xuất để giải quyết bài toán này.

### 3. MÔ TẢ THUẬT TOÁN

Để giải bài toán (6), bài báo đề xuất phân tách thành hai pha: Tại pha 1, học tăng cường DQN tối ưu ngưỡng CCA rời rạc, và Water-Filling tối ưu phân bổ công suất liên tục tại mỗi bước thời gian.

Thuật toán 1: Tối ưu tái sử dụng không gian dựa trên Double DQN kết hợp Water-Filling

Đầu vào: Cấu hình mạng WiFi ( $B$  AP,  $K$  STA/BSS), tập ngưỡng CCA  $\mathcal{C}$ ,  $P_{max}$

Đầu ra: Chuỗi ngưỡng CCA tối ưu, phân bổ công suất  $p^*$ , thông lượng mạng

1: Khởi tạo replay buffer  $\mathcal{D}$  dung lượng  $N_{buf}$ , bộ đệm n-step

2: Khởi tạo mạng hành vi  $Q(s, a; \theta)$  và mạng mục tiêu  $Q(s, a; \theta^-)$

3: for  $episode = 1 : E$  do

4: Khởi tạo môi trường WiFi, nhận trạng thái ban đầu  $s_1$

5: for  $t = 1 : T$  do

6: Chọn hành động phối hợp  $a_t$  bằng  $\epsilon$ -greedy từ  $Q(s_t, a; \theta)$

7: Giải mã  $a_t$  thành hành động CCA cho từng AP

8: Cập nhật ngưỡng CCA, xác định tập AP hoạt động theo (2)

9: Áp dụng Water-Filling tính công suất tối ưu  $p^*$  theo (5)

10: Tính SINR theo (3), thông lượng theo (4)

11: Tính thưởng đa thành phần  $r_t$  theo (7)–(8)

12: Lưu  $(s_t, a_t, r_t, s_{t+1})$  vào bộ đệm n-step

13: if bộ đệm n-step đủ  $n$  mẫu then

14: Tính lợi nhuận n-step  $G_t$  theo (12), lưu vào

15: end if

16: if  $\mathcal{D} \geq N_{warm}$  then

17: Lấy mẫu batch từ  $\mathcal{D}$  theo PER (11)

18: Tính giá trị mục tiêu  $y_j$  theo Double DQN (9)

19: Cập nhật  $\theta$  bằng gradient descent với hàm Huber loss

20: Cập nhật mềm  $\theta^-$  theo (10)  
 21: end if  
 22: end for  
 23: end for

### 3.1. Mô hình quá trình quyết định Markov

Bài toán điều chỉnh ngưỡng CCA được mô hình hóa dưới dạng quá trình quyết định Markov (MDP) với các thành phần sau:

**Không gian trạng thái:** Tại mỗi bước thời gian  $t$ , trạng thái bao gồm vector đặc trưng cục bộ của từng BSS, được ghép nối thành trạng thái toàn cục  $\mathbf{s}_t = [f_1(t), f_2(t), \dots, f_B(t)]$ . Mỗi vector  $f_b(t)$  chứa: độ lợi kênh chuẩn hóa theo thang logarit, mức nhiễu trung bình chuẩn hóa, giá trị ngưỡng CCA hiện tại, mức tải hàng đợi, số lượng người dùng chuẩn hóa, và trạng thái hoạt động của các BSS lân cận. Không gian trạng thái có số chiều  $B \times d_{\text{local}}$ , với  $d_{\text{local}} = 5 + (B - 1)$

**Không gian hành động:** Tác tử tập trung lựa chọn hành động phối hợp cho toàn bộ BSS AP đồng thời. Mỗi AP có 3 hành động: giảm ngưỡng CCA một bậc, giữ nguyên, hoặc tăng một bậc. Không gian hành động phối hợp có kích thước  $|\mathcal{A}| = 3^B$ . Việc sử dụng tác tử tập trung thay vì nhiều tác tử phân tán cho phép phối hợp toàn cục giữa các BSS, tránh xung đột trong quá trình điều chỉnh tham số.

**Hàm thưởng:** Hàm thưởng được thiết kế đa thành phần, kết hợp giữa hiệu suất cục bộ và toàn cục:

$$r_b(t) = \alpha \cdot \bar{R}_b - \beta \cdot \tilde{I}_b - \gamma \cdot \tilde{p}_b + \delta \cdot \mathbb{1}(R_b \geq R_{\min}) \tag{7}$$

$$r(t) = \frac{1}{B} \sum_{b=1}^B [w_{\text{loc}} \cdot r_b(t) + w_{\text{glob}} \cdot r_{\text{global}}(t)] \tag{8}$$

trong đó  $\bar{R}_b$  là thông lượng chuẩn hóa của BSS  $b$ ,  $\tilde{I}_b$  là nhiễu mà BSS  $b$  gây ra cho các BSS lân cận (chuẩn hóa),  $\tilde{p}_b$  là công suất phát chuẩn hóa,  $\mathbb{1}(R_b \geq R_{\min})$  thành phần thưởng QoS,  $r_{\text{global}}(t)$  phản

ánh hiệu suất tổng thể mạng, và  $\alpha, \beta, \gamma, \delta, w_{\text{loc}}, w_{\text{glob}}$  là các hệ số trọng số.

### 3.2. Chiến lược Double Deep Q-Network

Double DQN [15] được sử dụng nhằm khắc phục nhược điểm ước lượng quá cao giá trị Q trong DQN tiêu chuẩn [18]. Trong Double DQN, mạng hành vi (behavior network) được dùng để chọn hành động tốt nhất, còn mạng mục tiêu (target network) đánh giá giá trị của hành

$$y_j = r_j + \gamma_a^n \cdot Q_{\text{target}}(s_{j+1}, \arg \max_a Q(s_{j+1}, a; \theta); \theta^-) \tag{9}$$

trong đó  $\theta$  và  $\theta^-$  lần lượt là tham số mạng hành vi và mạng mục tiêu,  $\gamma_a$  là hệ số chiết khấu, và  $n$  là số bước trong cơ chế n-step returns.

Kiến trúc mạng nơ-ron bao gồm: lớp đầu vào với kích thước bằng số chiều trạng thái; lớp ẩn thứ nhất với hàm kích hoạt ReLU; lớp chuẩn hóa (Layer Normalization) giúp ổn định quá trình huấn luyện; lớp ẩn thứ hai; và lớp đầu ra với số nơ-ron tương ứng kích thước không gian hành động. Mạng mục tiêu được cập nhật mềm (soft update) với hệ số  $\tau$ :

$$\theta^- \leftarrow \tau \cdot \theta + (1 - \tau) \cdot \theta^-$$

Cập nhật mềm cho phép mạng mục tiêu thay đổi từ từ, tránh dao động mạnh trong quá trình huấn luyện so với cách cập nhật cứng (hard update) theo chu kỳ cố định.

### 3.3. Bộ nhớ kinh nghiệm ưu tiên và N-step Returns

Bộ nhớ kinh nghiệm ưu tiên (Prioritized Experience Replay – PER) [19] được áp dụng nhằm tăng hiệu quả sử dụng dữ liệu huấn luyện. Thay vì lấy mẫu ngẫu nhiên đều, PER gán xác suất lấy mẫu tỷ lệ với sai số TD (Temporal Difference):

$$P(i) = \frac{P_i^{\alpha_{\text{per}}}}{\sum_k P_k^{\alpha_{\text{per}}}} \tag{11}$$

trong đó  $p_i = |\delta_i| + \epsilon$  là độ ưu tiên của kinh nghiệm thứ  $i$ ,  $\alpha_{per}$ , điều chỉnh mức độ ưu tiên, và  $\epsilon$  là hằng số nhỏ đảm bảo mọi kinh nghiệm đều có xác suất dương.

Cơ chế n-step returns được kết hợp nhằm giảm phương sai trong ước lượng giá trị và tăng tốc độ lan truyền tín hiệu thưởng:

$$G_t = \sum_{k=0}^{n-1} \gamma_d^k \cdot r_{t+k} \quad (12)$$

### 3.4 Kỹ thuật Water-Filling

Kỹ thuật Water-Filling được áp dụng để tối ưu phân bổ công suất phát cho tập các AP đang hoạt động. Nguyên lý Water-Filling phân bổ nhiều công suất hơn cho các kênh có chất lượng tốt và ít công suất hơn cho các kênh kém. Cụ thể, công suất tối ưu cho AP là:

$$p_b^* = \min \left( \max \left( \mu - \frac{\sigma^2 + I_b}{h_{b,b}}, 0 \right), P_{max} \right) \quad (13)$$

trong đó  $\mu$  là mức nước (water level) được xác định thông qua tìm kiếm nhị phân để thỏa mãn ràng buộc công suất, và  $I_b$  là nhiễu tổng tại AP  $b$  từ các AP khác.

## 4. ĐÁNH GIÁ HIỆU NĂNG

Để đánh giá hiệu năng của thuật toán đề xuất Double DQN kết hợp Water-Filling và so sánh với cơ chế OBSS/PD tiêu chuẩn của IEEE 802.11ax, kịch bản mô phỏng được xây dựng gồm 3 trạm AP được đặt tại 3 đỉnh của tam giác đều cạnh 50 m. Trong mỗi AP, các STA được phân bổ một cách ngẫu nhiên theo phân bố đều. Các tham số mô phỏng chính được mô tả trong Bảng 1.

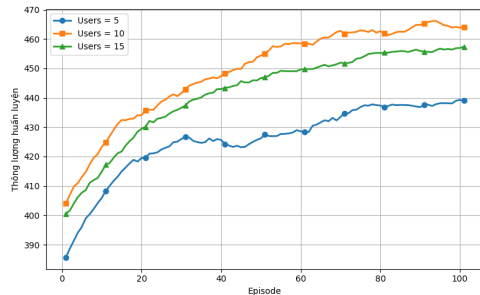
Bảng 1: Tham số mô phỏng

Tham số	Giá trị
Số lượng AP (B)	3
Số lượng STA trên mỗi BSS (K)	5 – 40

Băng thông kênh ( $W$ )	20 MHz
Tần số sóng mang ( $f_c$ )	5 GHz
Suy hao do vật cản (L)	7 dB
Công suất phát tối đa ( $P_{max}$ )	0.1 W (20 dBm)
Tốc độ tối thiểu QoS ( $R_{min}$ )	5 Mbps/BSS
Độ lệch chuẩn shadowing	1.5 dB
Kích thước vùng phục vụ	50 m × 50 m
Tập ngưỡng CCA ( $\mathcal{C}$ )	{-82, -78, -74, -70, -66, -62, -58} dBm
Ngưỡng CCA mặc định	-82 dBm
Tốc độ học ( $\eta$ )	0.001
Số episode huấn luyện ( $E$ )	200
Số bước mỗi episode ( $T$ )	720

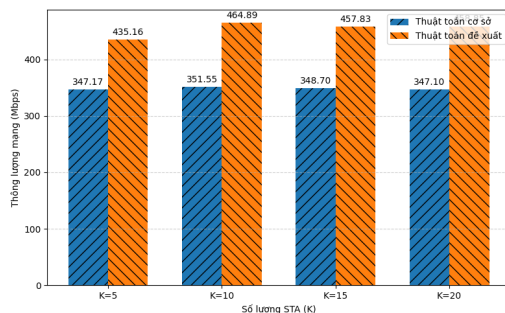
### 4.2. Kết quả mô phỏng và thảo luận

Hình 2 minh họa sự thay đổi của thông lượng hệ thống trong quá trình huấn luyện của thuật toán Double DQN kết hợp Water-Filling trong mạng WLAN dày đặc với số lượng người dùng trên mỗi BSS là  $K = 5, 10$  và  $15$ . Trong giai đoạn đầu của quá trình huấn luyện (episode trong khoảng 1-20), thông lượng tăng nhanh do tác tử học tăng cường bắt đầu khám phá không gian hành động và dần tìm được các cấu hình ngưỡng CCA phù hợp kết hợp với phân bổ công suất Water-Filling. Khi số episode tăng, tốc độ cải thiện thông lượng giảm dần và các đường cong trở nên mượt hơn, cho thấy chính sách học được dần ổn định.



Hình 2: Thông lượng mạng trong quá trình huấn luyện

Sau khoảng 70-80 episode, các đường cong gần như hội tụ và chỉ dao động nhỏ quanh giá trị trung bình, phản ánh tính ổn định của quá trình huấn luyện. Điều này cho thấy tác tử học tăng cường đã tìm được chính sách điều chỉnh ngưỡng CCA phù hợp và duy trì phân bổ công suất hiệu quả thông qua cơ chế Water-Filling. Nhờ đó, hệ thống đạt được sự cân bằng giữa tăng thông lượng mạng, kiểm soát nhiễu liên BSS và đảm bảo các ràng buộc QoS trong môi trường WLAN dày đặc.



Hình 3 so sánh thông lượng mạng trung bình giữa thuật toán cơ sở OBSS/PD [6] và thuật toán đề xuất DQN kết hợp Water-Filling với số lượng người dùng trên mỗi BSS lần lượt là  $K = 5, 10, 15$  và  $20$ . Trong đó, phương pháp OBSS/PD sử dụng ngưỡng CCA cố định để đánh giá để tính trọng của kênh truyền. Kết quả cho thấy thuật toán đề xuất luôn đạt hiệu năng vượt trội so với phương pháp cơ sở trong tất cả các trường hợp của số lượng người dùng được xem xét.

Hình 3: Thông lượng mạng trong quá trình kiểm tra

Cụ thể, khi  $K = 5$ , thông lượng tăng từ 347.17 Mbps lên 435.16 Mbps, tương đương mức cải thiện khoảng 25.4%. Với  $K = 10$ , thông lượng tăng từ 351.55 Mbps lên 464.89 Mbps, tương ứng 32.2%. Trong trường hợp  $K = 15$ , thông lượng tăng từ 348.70 Mbps lên 457.83 Mbps, tương đương 31.3%. Cuối cùng, khi  $K = 20$ , thông lượng đạt 458.85 Mbps, cao

hơn khoảng 32.2% so với 347.10 Mbps của thuật toán cơ sở. Những kết quả này cho thấy phương pháp đề xuất có khả năng tối ưu hiệu quả ngưỡng CCA và phân bổ công suất phát, từ đó khai thác tốt hơn tài sử dụng không gian và giảm nhiễu giữa các BSS trong mạng WLAN dày đặc.

## 5. KẾT LUẬN

Bài báo đã đề xuất phương pháp tối ưu tài sử dụng không gian cho mạng WLAN dày đặc dựa trên Double DQN kết hợp kỹ thuật Water-Filling. Trong đó DQN tối ưu tham số ngưỡng CCA rời rạc, Water-Filling thực hiện phân bổ công suất liên tục theo chất lượng kênh. Cơ chế học tăng cường kết hợp với thiết kế hàm thưởng đa thành phần cho phép cân bằng giữa thông lượng mạng, kiểm soát nhiễu và đảm bảo ràng buộc QoS. Kết quả mô phỏng cho thấy thuật toán đề xuất đạt quá trình huấn luyện ổn định và hội tụ nhanh, với thông lượng tăng khoảng 30% so với cơ chế OBSS/PD tiêu chuẩn của IEEE 802.11ax.

## TÀI LIỆU THAM KHẢO

[1] Y. Li, Y. Zhang, K. Luo, T. Jiang, Z. Li and W. Peng, "Ultra-Dense HetNets Meet Big Data: Green Frameworks, Techniques, and Approaches," in IEEE Communications Magazine, vol. 56, no. 6, pp. 56-63, June 2018, doi: 10.1109/MCOM.2018.1700425.

[2] G. Naik, J. -M. Park, J. Ashdown and W. Lehr, "Next Generation Wi-Fi and 5G NR-U in the 6 GHz Bands: Opportunities and Challenges," in IEEE Access, vol. 8, pp. 153027-153056, 2020, doi: 10.1109/ACCESS.2020.3016036.

[3] Mozaffariahrar E, Theoleyre F, Menth M. A Survey of Wi-Fi 6: Technologies, Advances, and Challenges. Future Internet. 2022; 14(10):293. <https://>

doi.org/10.3390/fi14100293

[4] Karakoç, A, Yilmaz, H. B, & Kuran, M. Ş (2023). More WiFi for everyone: Increasing spectral efficiency in WiFi6 networks using a distributed OBSS/PD mechanism. *Turkish Journal of Electrical Engineering and Computer Sciences* 31 (3): 660-677. <https://doi.org/10.55730/1300-0632.4008>

[5] C. -C. Lai, A. -H. Tsai, C. -W. Ting, K. -H. Lin, J. -C. Ling and C. -E. Tsai, "Interference-Aware Deployment for Maximizing User Satisfaction in Multi-UAV Wireless Networks," in *IEEE Wireless Communications Letters*, vol. 12, no. 7, pp. 1189-1193, July 2023, doi: 10.1109/LWC.2023.3266011

[6] F. Wilhelmi, S. Barrachina-Muñoz, C. Cano, I. Selinis, B. Bellalta,, "Spatial reuse in IEEE 802.11ax WLANs," *Comput. Commun.*, vol. 170, pp. 65–83, Mar. 2021.

[7] F. Wilhelmi, S. Barrachina-Muñoz and B. Bellalta, "On the Performance of the Spatial Reuse Operation in IEEE 802.11ax WLANs," *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*, Granada, Spain, 2019, pp. 1-6, doi: 10.1109/CSCN.2019.8931315.

[8] L. Lanante and S. Roy, "Performance Analysis of the IEEE 802.11ax OBSS\_PD-Based Spatial Reuse," in *IEEE/ACM Transactions on Networking*, vol. 30, no. 2, pp. 616-628, April 2022, doi: 10.1109/TNET.2021.3117816

[9] I. Selinis, K. Katsaros, S. Vahid and R. Tafazolli, "Control OBSS/PD Sensitivity Threshold for IEEE 802.11ax BSS Color," *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Bologna, Italy, 2018, pp. 1-7, doi: 10.1109/PIMRC.2018.8580778.

[10] Anthony Bardou, Thomas Begin, and Anthony Busson. 2021. Improving the Spatial Reuse in IEEE 802.11ax WLANs: A Multi-Armed Bandit Approach. In *Proceedings of the 24th International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '21)*. Association for Computing Machinery, New York, NY, USA, 135–144. <https://doi.org/10.1145/3479239.3485715>

[11] S. Szott, K. Kosek-Szott, P. Gawłowicz, J. T. Gómez, B. Bellalta, A. Zubow, and F. Dressler, "Wi-Fi meets ML: A survey on improving IEEE 802.11 performance with machine learning," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1843–1893, 3rd Quart. 2022.

[12] Y. Huang and K. -W. Chin, "A Deep Q-Network Approach to Optimize Spatial Reuse in WiFi Networks," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6636-6646, June 2022, doi: 10.1109/TVT.2022.3160446

[13] J. Jung, J. Baik, Y. Kim, H. -S. Park and J. -M. Chung, "OTOP: Optimized Transmission Power Controlled OBSS PD-Based Spatial Reuse for High Throughput in IEEE 802.11be WLANs," in *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 17110-17123, 1 Oct.1, 2023, doi: 10.1109/JIOT.2023.3275544.

[14] Pasquale Imputato, Stefano Avallone, Malcolm Smith, David Nunez, Boris Bellalta, *Beyond Wi-Fi 7: Spatial reuse through multi-AP coordination*, *Computer Networks*, Vol 239, 2024, ISSN 1389-1286, <https://doi.org/10.1016/j.comnet.2023.110160>.

[15] ITU-R, "Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 450 GHz," *Recommendation ITU-R P.1238-12*, 2021.

# PHÂN TÍCH SỰ ĐÁNH ĐỔI GIỮA HIỆU NĂNG VÀ NĂNG LƯỢNG TRONG MẠNG TRUYỀN THÔNG TÁN XẠ NGƯỢC MÔI TRƯỜNG CHỐNG NHIỀU SỬ DỤNG HỌC TĂNG CƯỜNG ĐA TÁC TỬ

Nguyễn Thái Dương<sup>1</sup>, Nguyễn Ngọc Tân<sup>1\*</sup>

<sup>1</sup> Khoa Công nghệ thông tin, Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội

\*Email: tan.nguyen@vnu.edu.vn

## TÓM TẮT

Sự bùng nổ của các thiết bị Internet vạn vật (IoT) trong kỷ nguyên 6G đặt ra thách thức kép về tiết kiệm năng lượng và khả năng chống chịu trước các cuộc tấn công gây nhiễu chủ động. Kỹ thuật Truyền thông Tán xạ ngược Môi trường (Ambient Backscatter Communication - AmB) nổi lên như một giải pháp hứa hẹn nhờ khả năng tận dụng sóng vô tuyến có sẵn để truyền tin với công suất cực thấp. Tuy nhiên, việc tối ưu hóa đồng thời thời độ tin cậy và hiệu quả năng lượng trong môi trường nhiễu là bài toán phức tạp. Bài báo này nghiên cứu bài toán cân bằng tài nguyên động, trong đó tín hiệu gây nhiễu được thu thập để cung cấp năng lượng vận hành hệ thống. Chúng tôi đề xuất khung giải pháp IA-MADDPG, kết hợp Học tăng cường sâu đa tác tử với cơ chế Học bắt chước nhằm khắc phục vấn đề hội tụ chậm trong không gian hành động liên tục. Kết quả mô phỏng cho thấy, thay vì tối đa hóa thông lượng một cách mù quáng, phương pháp đề xuất chấp nhận mức tiêu thụ năng lượng cao hơn để duy trì sự ổn định của hiệu suất thông lượng trên năng lượng trong các điều kiện khắc nghiệt.

**Từ khóa:** Truyền thông tán xạ ngược môi trường, Học tăng cường sâu đa tác tử, Chống nhiễu, Hiệu quả năng lượng, mạng truyền trực tiếp.

## ABSTRACT

The deployment of Internet of Things (IoT) devices in 6G networks introduces dual challenges regarding energy efficiency and resilience against active jamming attacks. Ambient Backscatter Communication (AmB) provides a mechanism to address these constraints by modulating existing radio frequency (RF) waves for ultra-low-power transmission. However, the joint optimization of transmission reliability and energy efficiency in non-stationary jamming environments constitutes a complex resource allocation problem. This paper investigates a dynamic resource balancing model wherein adversarial jamming signals are harvested to supply operational energy for the AmB system. We formulate the joint anti-jamming and energy allocation problem as a partially observable Markov decision process (POMDP). To solve this, we propose an Imitation-Augmented Multi-Agent Deep Deterministic Policy Gradient (IA-MADDPG) framework. This architecture integrates multi-agent deep reinforcement learning (MARL) with an imitation learning mechanism to mitigate the slow convergence and high exploration overhead inherent to continuous action spaces. Simulation results indicate that the IA-MADDPG framework strictly avoids unconstrained throughput maximization; instead, it strategically allocates higher energy consumption to guarantee the stability of the throughput-to-energy efficiency metric under severe jamming conditions.

**Keyword:** Ambient Backscatter Communication, Multi-agent deep reinforcement learning, Anti-Jamming, Energy efficiency, Direct transmission network.

Ngày nhận bài: 05/01/2026; Ngày sửa bài: 10/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. MỞ ĐẦU

Trong lộ trình phát triển của mạng di động thế hệ thứ 6 (6G), việc kết nối hàng tỷ thiết bị IoT công suất thấp đòi hỏi các giải pháp truyền thông mang tính đột phá về hiệu quả năng lượng. Mặc dù các công nghệ hiện tại đã đạt được những bước tiến lớn về tốc độ, rào cản về tuổi thọ pin và sự khan hiếm phổ tần vẫn là những nút thắt cổ chai [1]. Đặc biệt, trong các kịch bản quân sự hoặc an ninh cao, các thiết bị IoT thường xuyên phải đối mặt với các cuộc tấn công gây nhiễu nhằm làm gián đoạn kết nối.

Các phương pháp chống nhiễu truyền thông như Trải phổ nhảy tần (FHSS) hay Trải phổ chuỗi trực tiếp (DSSS) thường yêu cầu các thiết bị phải tiêu tốn một lượng năng lượng lớn để phát tín hiệu mạnh hơn nhiều hoặc liên tục thay đổi tần số công tác [2]. Điều này đi ngược lại với triết lý thiết kế sạch và bền vững của mạng IoT mật độ cao. Gần đây, kỹ thuật Truyền thông tán xạ ngược môi trường đã được đề xuất như một hướng đi mới, cho phép các thiết bị giao tiếp bằng cách điều biến và phản xạ lại các sóng vô tuyến có sẵn trong môi trường (như sóng TV, Wi-Fi, 5G) thay vì tự tạo ra sóng mang [3].

Một hướng tiếp cận đầy tiềm năng nhưng chưa được khai thác triệt để là việc coi tín hiệu gây nhiễu không phải là nhiễu, mà là một nguồn năng lượng sóng mang công suất lớn [4]. Trong kịch bản này, thiết bị Truyền thông tán xạ ngược môi trường có thể tận dụng chính năng lượng từ kẻ tấn công để phản xạ tín hiệu của mình, biến nguy cơ thành cơ hội. Tuy nhiên, việc hiện thực hóa ý tưởng này gặp phải thách thức lớn trong việc quản lý tài nguyên: Điều chỉnh hệ số phản xạ () để cân bằng giữa cường độ tín hiệu phát đi và năng lượng tiêu hao cho mạch điện và Lựa chọn chế độ truyền trực tiếp hoặc truyền qua trung gian để tối ưu hóa năng

lượng toàn trình [5].

Các phương pháp tối ưu hóa truyền thông thường dựa trên các mô hình tĩnh hoặc tham lam, vốn không thể nắm bắt được tính động của môi trường nhiễu và sự tương tác phức tạp giữa các người dùng [6]. Học tăng cường sâu (Deep Reinforcement Learning - DRL), đặc biệt là mô hình đa tác tử (Multi-Agent DRL), cung cấp một công cụ mạnh mẽ để giải quyết các bài toán ra quyết định trong môi trường phi tập trung. Tuy nhiên, các thuật toán nền tảng như MADDPG thường gặp vấn đề “khởi đầu lạnh”, dẫn đến tốn kém thời gian và năng lượng cho quá trình thăm dò [7] [8].

Bài báo này đề xuất và phân tích một khung giải pháp dựa trên thuật toán IA-MADDPG (Imitation-Augmented MADDPG). Đóng góp chính của bài báo tập trung vào việc phân tích sâu sắc khía cạnh hiệu quả năng lượng, làm rõ cơ chế mà tại đó các tác tử học máy chấp nhận hy sinh một phần năng lượng tiêu thụ để đổi lấy sự đảm bảo về chất lượng dịch vụ (QoS) và thông lượng trong điều kiện nhiễu khắc nghiệt.

## 2. MÔ HÌNH HỆ THỐNG VÀ BÀI TOÁN TỐI ƯU

### 2.1. Mô hình Kịch bản Mạng

Chúng tôi xem xét một mạng truyền thông không dây trong vùng diện tích  $120 \times 120$  mét, bao gồm  $N$  cặp người dùng nguồn - đích (Source User - SU, Destination User - DU) và một trạm chuyển tiếp (Relay Base Station - rBS) đặt tại trung tâm. Hệ thống chịu tác động bởi một tác nhân gây nhiễu công suất lớn di chuyển ngẫu nhiên và phát nhiễu liên tục.

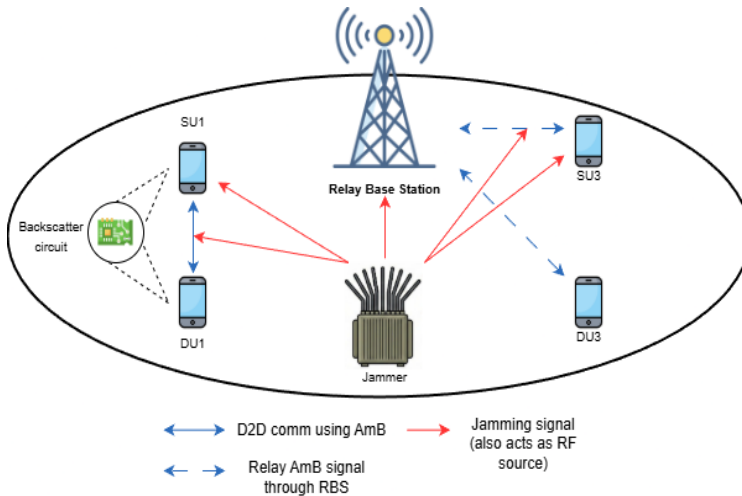
Khác với các hệ thống vô tuyến chủ động, mỗi  $SU_i$  trong mô hình này được trang bị thẻ Backscatter thụ động. Thay vì sử dụng bộ dao động nội để tạo sóng mang,  $SU_i$  thu nhận tín hiệu từ tác nhân gây nhiễu

và phản xạ lại nó tới  $DU_i$  hoặc rBS.

Quá trình truyền tin có thể diễn ra theo hai chế độ ( $m_i$ ):

- **Chế độ truyền trực tiếp:** Truyền trực tiếp từ  $SU_i$  đến  $DU_i$  thông qua phản xạ tín hiệu tác nhân gây nhiễu. Chế độ này tiết kiệm năng lượng truyền dẫn nhưng dễ bị ảnh hưởng nếu khoảng cách xa hoặc nhiễu tại máy thu quá lớn.

- **Chế độ truyền qua trung gian:** Tín hiệu được phản xạ từ  $SU_i$  đến rBS, sau đó rBS giải mã và chuyển tiếp đến  $DU_i$ . Chế độ này tăng độ tin cậy nhờ lợi ích phân tập không gian nhưng có thể gây độ trễ và tiêu tốn tài nguyên hệ thống.



## 2.2. Mô hình Tín hiệu và Tiêu thụ Năng lượng

Đây là phần trọng tâm để phân tích hiệu quả năng lượng. Quá trình backscatter được thực hiện bằng cách thay đổi trở kháng anten để hấp thụ hoặc phản xạ tín hiệu tới. Gọi  $\alpha_i \in [0, 1]$  là hệ số phản xạ của  $SU_i$ .

Tín hiệu thu được tại đích đến  $D_i$  trong chế độ truyền trực tiếp được biểu diễn:

$$y_{D_i}(t) = \sqrt{P_J} g_{J,S_i} g_{S_i,D_i} \alpha_i x_{S_i}(t) + n_{D_i}(t) \quad (1)$$

Trong đó,  $P_J$  là công suất phát của tác nhân gây nhiễu,  $g_{A,B}$  là độ lợi kênh giữa nút A và B, và  $n_{D_i}$  là tạp âm nhiệt.

Mặc dù Truyền thông tán xạ ngược

môi trường không tiêu tốn năng lượng để tạo sóng mang, nhưng nó vẫn tiêu thụ năng lượng cho việc chuyển mạch trở kháng và hoạt động của vi mạch điều khiển. Tổng năng lượng tiêu thụ  $E_i$  cho một lần truyền tin của người dùng  $i$  được mô hình hóa như sau:

$$E_i = E_{\text{circuit}} + \eta |\alpha_i|^2, \quad (2)$$

Trong đó:

$E_{\text{circuit}} = c_{\text{switch}} + c_{\text{sense}}$  là năng lượng cố định cho việc chuyển mạch và cảm biến môi trường.

$\eta |\alpha_i|^2$  là thành phần năng lượng động, phụ thuộc bình phương vào hệ số phản xạ.

Từ phương trình trên, việc tăng sẽ làm tăng cường độ tín hiệu phản xạ (tăng SINR), nhưng đồng thời làm tăng chi phí năng lượng theo hàm bậc hai. Đây chính là trọng tâm của sự đánh đổi (trade-off) mà bài báo này phân tích.

## 2.3. Thiết lập Bài toán Tối ưu Đa mục tiêu

Mục tiêu của hệ thống là tối đa hóa thông lượng đồng thời kiểm soát mức tiêu thụ năng lượng. Tỷ số tín hiệu trên nhiễu cộng tạp âm (SINR) tại máy thu phụ thuộc vào chế độ truyền dẫn:

- SINR chế độ truyền trực tiếp ( $SINR_i^{D2D}$ ) tỉ lệ thuận với  $\alpha_i^2 P_J$ .

- SINR chế độ truyền qua trung gian ( $SINR_i^{Relay}$ ) là giá trị nhỏ nhất của SINR hai chặng (SU-rBS và rBS-DU).

Bài toán được mô hình hóa dưới dạng Quá trình Quyết định Markov Quan sát Một phần (POMDP), trong đó mỗi tác tử  $SU_i$  quan sát trạng thái kênh cục bộ và đưa ra hành động  $a_i = [\alpha_i, m_i]$ .

Hàm phần thưởng (Reward Function)  $r_i$  cho mỗi tác tử được thiết kế để cân bằng giữa thông lượng và năng lượng tiêu thụ:

$$r_i = w_1 \log_2(1 + \text{SINR}_i) + w_2 \tanh(\text{SINR}_i - \Gamma) - w_3 \alpha_i^2, \quad (3)$$

Trong đó, thành phần thứ nhất khuyến khích tăng thông lượng, thành phần thứ hai đảm bảo SINR vượt ngưỡng an toàn  $\Gamma$ , và thành phần thứ ba ( $-w_3 \alpha_i^2$ ) phạt nặng cho việc tiêu tốn năng lượng quá mức. Việc điều chỉnh các trọng số  $w$  sẽ quyết định chiến lược của mạng: thiên về hiệu năng (Performance-centric) hay thiên về tiết kiệm năng lượng (Energy-centric).

### 3. THUẬT TOÁN ĐỀ XUẤT: IA-MADDPG

Để giải quyết bài toán tối ưu đa mục tiêu trong môi trường nhiều động như đã mô tả, chúng tôi đề xuất khung giải pháp IA-MADDPG. Khung giải pháp này được xây dựng dựa trên nền tảng của thuật toán MADDPG [7], nhưng tích hợp các cơ chế tiên tiến để khắc phục nhược điểm hội tụ chậm và kém ổn định («khởi đầu lạnh») của các phương pháp Học tăng cường (RL) truyền thống.

#### 3.1. Học Bắt chước (Imitation Learning) để Tăng tốc Hội tụ

Trong giai đoạn đầu của quá trình huấn luyện, việc để các tác tử khám phá ngẫu nhiên không gian hành động liên tục là cực kỳ lãng phí năng lượng và thời gian. Để giải quyết vấn đề này, chúng tôi giới thiệu một chính sách chuyên gia phân tích  $\mu_E(o_i)$ . Chuyên gia này sử dụng chiến lược tham lam (Greedy) dựa trên thông tin kênh truyền tức thời để tối ưu hóa SINR [8].

Cơ chế thu thập dữ liệu từ chiến lược tham lam: Cụ thể, tại mỗi bước thời gian của giai đoạn khởi động, chính sách chuyên gia quan sát trạng thái kênh cục bộ  $o_i$  (bao gồm  $h_{J,SU}$ ,  $h_{SU,DU}$  và mức

nhều) để tính toán cấu hình hành động  $a_i = \{\alpha_i, m_i\}$  nhằm cực đại hóa giá trị hàm phần thưởng  $r_i$  tại phương trình (3). Quá trình ra quyết định này mang tính thiên cận, chỉ tập trung vào tối ưu hóa lợi ích tức thời mà không cập nhật hàm giá trị dài hạn. Các bộ dữ liệu chuyển đổi trạng thái  $(s, a, r, s')$  sinh ra từ chiến lược này được thu thập và đẩy liên tục vào bộ đệm kinh nghiệm  $D$  cho đến khi đạt ngưỡng dung lượng.

Quá trình học được tăng tốc thông qua hai cơ chế:

- **Khởi động chuyên gia (Expert Warm-up):** Trước khi bắt đầu quá trình huấn luyện RL, bộ đệm kinh nghiệm  $D$  được lấp đầy một phần bởi các mẫu chuyển đổi trạng thái  $(s, a, r, s')$  sinh ra từ chiến lược tham lam (Greedy). Điều này đảm bảo rằng ngay từ các bước cập nhật gradient đầu tiên, mạng nơ-ron đã được học từ các dữ liệu chất lượng cao thay vì dữ liệu ngẫu nhiên.

- **Sao chép hành vi (Behavior Cloning - BC):** Hàm mất mát của mạng Actor được sửa đổi để bao gồm thành phần sai số bất chước. Điều này khuyến khích tác tử mô phỏng hành vi của chuyên gia trong giai đoạn đầu, đồng thời vẫn giữ khả năng khám phá để tìm ra chiến lược tốt hơn về sau. Hàm mất mát tổng hợp được định nghĩa:

$$L(\theta_i) = L_{PG} + \lambda_{IL} E \left[ \left| \mu_i(o_i) - \mu_E(o_i) \right|^2 \right] \quad (4)$$

Trong đó,  $\mathcal{L}_{PG}$  là mất mát theo Policy Gradient tiêu chuẩn,  $\mu_E$  là chính sách chuyên gia, và  $\lambda_{IL}$  là hệ số trọng số giảm dần theo thời gian.

Cơ chế suy giảm của hệ số  $\lambda_{IL}$ : Để đảm bảo tính tái lập của quá trình huấn luyện, cơ chế giảm dần của hệ số  $\lambda_{IL}$  được mô hình hóa theo hàm suy giảm mũ. Cụ thể, tại tập huấn luyện thứ  $e$ , trọng số này được cập nhật theo công thức:

$$\lambda_{IL}(e) = \lambda_{IL}(0) \cdot (\gamma_{IL})^e$$

Trong đó,  $\lambda_{IL}(0)$  là giá trị trọng số khởi tạo ban đầu và  $\gamma_{IL} \in (0,1)$  là hằng số tốc độ suy giảm. Việc sử dụng hàm mũ ép buộc mô hình bám sát nhân của chuyên gia ở các tập đầu tiên để ổn định không gian trọng số mạng nơ-ron. Sau đó,  $\lambda_{IL}$  suy giảm tiệm cận về 0 ở các tập sau để triệt tiêu độ chệch của chiến lược tham lam, mở rộng không gian khám phá để thuật toán MARL tự hội tụ về điểm cân bằng tối ưu hơn.

### 3.2. Nâng cao Tính Ổn định với PER và TD3

Để tăng cường độ ổn định trong môi trường đa tác tử biến đổi nhanh, chúng tôi tích hợp hai kỹ thuật:

- **Ưu tiên Hồi tưởng Kinh nghiệm (Prioritized Experience Replay - PER):** Thay vì lấy mẫu ngẫu nhiên đồng đều, thuật toán ưu tiên lấy mẫu các chuyển đổi có sai số dự báo (TD-error) lớn. Điều này giúp mạng nơ-ron tập trung học các tình huống “khó” hoặc bất ngờ (ví dụ: khi tác nhân gây nhiễu thay đổi vị trí đột ngột) [9].

- **Làm mượt Chính sách Mục tiêu (Target Policy Smoothing):** Lấy cảm hứng từ thuật toán TD3, chúng tôi thêm nhiễu được cắt (clipped noise) vào hành động mục tiêu khi cập nhật mạng Critic. Kỹ thuật này giúp giảm thiểu hiện tượng ước lượng quá cao giá trị Q (Q-value overestimation), một vấn đề thường gặp trong DDPG.

### 3.3. Quy trình Thuật toán Chi tiết

Quy trình huấn luyện đầy đủ của IA-MADDPG được trình bày chi tiết trong Thuật toán 1.

#### Thuật toán 1: Imitation-Augmented MADDPG (IA-MADDPG)

**Đầu vào:** Các siêu tham số mạng, Môi trường truyền thông, Chuyên gia  $\mu^E$ .

**Đầu ra:** Chính sách tối ưu cho các tác tử  $SU_i$ .

#### 1. Khởi tạo:

- Khởi tạo mạng Actor  $\mu_i(o_i|\theta_i)$  và Critic  $Q_i(s,a|\phi_i)$  với trọng số ngẫu nhiên.

- Khởi tạo các mạng mục tiêu (Target networks)  $\mu'_i$  và  $Q'_i$  với trọng số tương ứng.

- Khởi tạo Bộ đệm ưu tiên (Prioritized Replay Buffer)  $\mathcal{D}$ .

#### 2. Giai đoạn Khởi động (Warm-up Phase):

- For  $k = 1$  to  $K_{warmup}$  do:
  - Quan sát trạng thái  $s$ ; Chọn hành động từ chuyên gia  $a = \mu^E(s)$ .
  - Thực thi hành động, nhận phần thưởng  $r$  và trạng thái mới  $s'$ .
  - Lưu bộ chuyển đổi  $(s, a, r, s')$  vào  $\mathcal{D}$ .

- End For

#### 3. Giai đoạn Huấn luyện (Training Phase):

- For episode = 1 to  $M$  do:
  - Khởi tạo lại môi trường, nhận trạng thái ban đầu  $s$ .
  - Khởi tạo nhiễu ngẫu nhiên  $\mathcal{N}$  cho việc khám phá.
  - For  $t = 1$  to  $T$  (số bước mỗi tập) do:
    - Mỗi tác tử chọn hành động  $a_i = \mu_i(o_i) + \mathcal{N}_t$ .
    - Thực thi hành động kết hợp  $a = (a_1, \dots, a_N)$ , nhận phần thưởng  $r$  và trạng thái mới  $s'$ .
    - Lưu  $(s, a, r, s')$  vào  $\mathcal{D}$  với độ ưu tiên cao nhất.
    - Lấy mẫu ngẫu nhiên một minibatch  $B$  từ  $\mathcal{D}$  dựa trên trọng số ưu tiên PER.
    - Cập nhật tham số (cho mỗi tác tử  $i$ ):
      - ◊ Tính mục tiêu  $y_i$  sử dụng Target Policy Smoothing (TD3).
      - ◊ Cập nhật Critic bằng cách tối thiểu hóa hàm mất mát Bellman có trọng số PER.
      - ◊ Cập nhật Actor bằng cách tối thiểu hóa hàm mất mát tổng hợp.
      - ◊ Cập nhật độ ưu tiên trong  $\mathcal{D}$  dựa trên

TD-error mới.

$$\diamond \theta'_i \leftarrow \tau\theta_i + (1 - \tau)\theta'_i$$

$$\diamond \phi'_i \leftarrow \tau\phi_i + (1 - \tau)\phi'_i$$

- End For (Time step)
- End For (Episode)

#### 4. ĐÁNH GIÁ HIỆU SUẤT VÀ PHÂN TÍCH

##### 4.1. Thiết lập Mô phỏng và Kịch bản đánh giá

Hệ thống được triển khai trên nền tảng PyTorch với các tham số mô phỏng bám sát thực tế môi trường nhiễu động. Cụ thể, công suất tác nhân gây nhiễu được thiết lập  $P_j = 1.0W$ , tạo ra áp lực nhiễu lớn lên mạng lưới. Các tác tử (Agent) sử dụng mạng nơ-ron sâu với cấu trúc Actor [256, 256, 128] và Critic [512, 256, 128, 64] để học chiến lược đối phó.

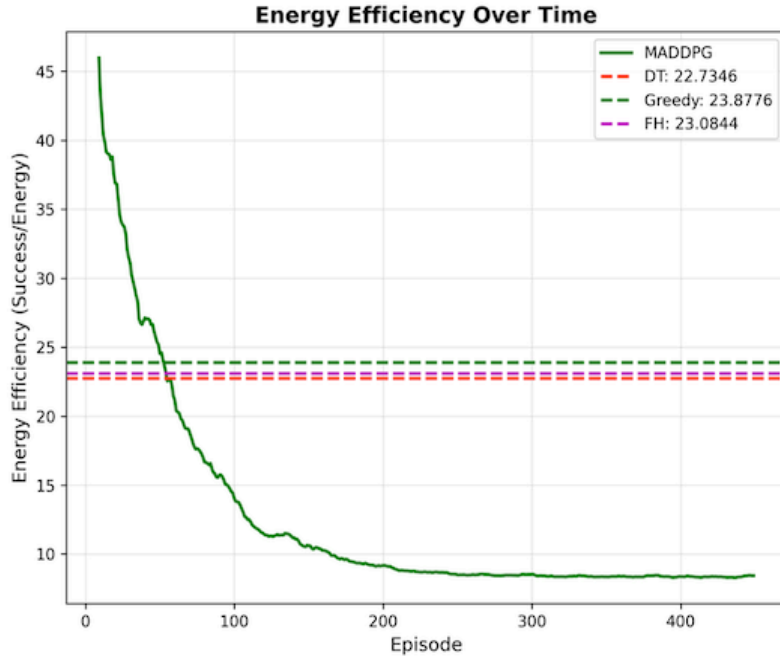
Để làm nổi bật khả năng thích nghi của thuật toán đề xuất (IA-MADDPG), chúng tôi so sánh với ba chiến lược cơ sở đại diện cho tư duy thiết kế tĩnh:

- Truyền dẫn trực tiếp (Direct Transmission - DT) & Tham lam (Greedy): Sử dụng hệ số phản xạ cố định ở mức thấp ( $\alpha = 0.2$ ). Đây là chiến lược "an toàn" nhằm tối ưu hóa chỉ số tiết kiệm năng lượng lý thuyết.
- Nhảy tần ngẫu nhiên (Frequency Hopping - FH): Chiến lược nhảy tần ngẫu nhiên, đại diện cho phương pháp tránh nhiễu thụ động.

##### 4.2. Phân tích Hiệu quả Năng lượng

##### (Energy Efficiency Analysis)

Thay vì chỉ đánh giá mức tiêu thụ năng lượng một cách cơ học, phần này đi sâu phân tích hành vi thông minh của hệ thống thông qua chỉ số Hiệu quả Năng lượng (Energy Efficiency).



Hình 1: Hiệu quả Năng lượng theo thời gian (Energy Efficiency Over Time)

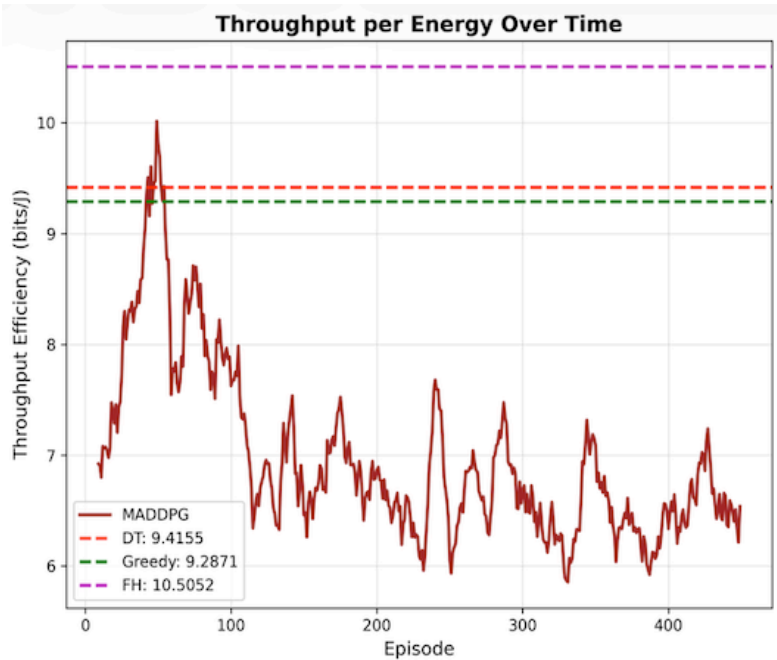
Ở Hình 1, đường biểu diễn của IA-MADDPG (màu xanh lá) hội tụ về mức thấp hơn (~9) so với các đường cơ sở (~23). Trong các nghiên cứu truyền thống, điều này thường bị hiểu nhầm là kém hiệu quả. Tuy nhiên, trong bối cảnh chống nhiễu, đây chính là bằng chứng cho khả năng thích nghi vượt trội của hệ thống đề xuất. Các phương pháp DT, Greedy và FH duy trì hiệu quả năng lượng cao giả bằng cách giữ cố định hệ số phản xạ ở mức thấp ( $\alpha = 0.2$ ). Chiến lược này giống như việc một chiếc xe chạy chậm để tiết kiệm xăng bất chấp việc không thể leo qua dốc cao. Trong môi trường nhiễu mạnh ( $P_j = 1.0W$ ), mức năng lượng thấp này không đủ để tạo ra tỷ số tín hiệu trên nhiễu (SINR) cần thiết để giải mã gói

tin. Do đó, sự hiệu quả của các phương pháp này thực chất là sự bất lực trong việc duy trì kết nối. Phương pháp này tiết kiệm năng lượng, nhưng đánh mất mục tiêu được ưu tiên hàng đầu là truyền tin tin cậy. Ngược lại, sự sụt giảm hiệu quả năng lượng của IA-MADDPG phản ánh hành động phản kháng tích cực. Tác tử học được rằng để duy trì kết nối dưới tác động của nhiễu, nó buộc phải đẩy hệ số phản xạ lên mức cao (gần 1.0). Mặc dù điều này làm tăng chi phí năng lượng tức thời (theo hàm bậc hai  $\eta\alpha^2$ ), nhưng nó đảm bảo tín hiệu vượt qua ngưỡng giải mã. Như vậy, mức hiệu quả năng lượng thấp hơn của IA-MADDPG không phải là lãng phí, mà là chi phí bảo hiểm cần thiết để đảm bảo độ tin cậy của dịch vụ (QoS), điều mà các chiến lược tiết kiệm năng lượng cục bộ đơn giản không thể đáp ứng.

### 4.3. Phân tích Hiệu suất Thông lượng trên Năng lượng (Throughput Efficiency)

Để minh chứng rõ hơn trí thông minh của hệ thống, chúng ta xem xét chỉ số Hiệu suất Thông lượng (Throughput Efficiency - bits/Joule) trong Hình 2. Đây là thước đo cho thấy mỗi đơn vị năng lượng bỏ ra mang lại bao nhiêu giá trị thông tin thực tế.

Sự khác biệt giữa đường IA-MADDPG (dao động mạnh) và các đường cơ sở (đi ngang) cho thấy hai triết lý vận hành hoàn toàn trái ngược.



Hình 2: Hiệu suất Thông lượng trên Năng lượng theo thời gian

- Sự cứng nhắc của các phương pháp tĩnh: Các đường nét đứt nằm ngang thể hiện sự cứng nhắc. Dù môi trường nhiều thay đổi như thế nào, các phương pháp này vẫn áp dụng một cấu hình duy nhất. Điều này dẫn đến rủi ro “Outage” (mất kết nối) cực cao khi nhiễu biến động vượt quá khả năng chịu đựng của cấu hình  $\alpha = 0.2$ .

- Đường dao động của IA-MADDPG (màu nâu đỏ) là minh chứng cho quá trình dò tìm và thích nghi liên tục. Các đỉnh thể hiện những khoảnh khắc tác tử phát hiện nhiễu giảm và thông minh hạ thấp công suất để tiết kiệm năng lượng. Các đáy thể hiện những thời điểm tác tử phát hiện nhiễu tăng cao và quyết định hy sinh hiệu suất bits/Joule để dồn toàn lực duy trì kết nối. Hành vi này tương tự như cơ chế điều tiết sinh học: hệ thống tự động cân bằng giữa trạng thái tiết kiệm và tiêu hao tùy thuộc vào mức độ đe dọa của máy gây nhiễu.

Kết quả mô phỏng khẳng định sự uy việt của IA-MADDPG không nằm ở con

số tiết kiệm năng lượng thô, mà nằm ở khả năng quản trị rủi ro. Trong khi các phương pháp cơ sở chấp nhận rủi ro mất tin để bảo toàn năng lượng, IA-MADDPG sử dụng năng lượng như một chiến lược để đảm bảo tính liên tục của dịch vụ. Đây là đặc tính tiên quyết cho các mạng IoT quân sự hoặc các ứng dụng y tế, an ninh quan trọng, nơi sự gián đoạn thông tin là không thể chấp nhận được.

## 5. KẾT LUẬN

Bài báo đã trình bày một phân tích toàn diện về sự đánh đổi giữa hiệu năng và năng lượng trong mạng Truyền thông tán xạ ngược môi trường dưới tác động của nhiễu chủ động. Thông qua việc áp dụng thuật toán IA-MADDPG, chúng tôi đã chứng minh rằng việc sử dụng các chiến lược cố định tuy tiết kiệm năng lượng nhưng không đảm bảo độ tin cậy trong môi trường thù địch. Ngược lại, việc cho phép các tác tử học máy tự chủ điều chỉnh hệ số phản xạ và chế độ truyền dẫn - mặc dù tiêu tốn năng lượng hơn - là một sự đầu tư chiến lược cần thiết. Giải pháp đề xuất cho phép duy trì kết nối ổn định bằng cách biến đổi linh hoạt cấu hình hệ thống, phù hợp với các ứng dụng yêu cầu chất lượng dịch vụ (QoS) cao trong kỷ nguyên 6G. Các hướng phát triển tiếp theo sẽ tập trung vào việc tích hợp các ràng buộc năng lượng cứng vào không gian hành động để tìm ra điểm cân bằng Pareto tối ưu hơn nữa.

## TÀI LIỆU THAM KHẢO

- [1] H. Tataria, M. Shafi, A. Molisch, M. Dohler, H. Sjoland, and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," *Proceedings of the IEEE*, pp. 1-34, 2021.
- [2] H. Pirayesh and H. Zeng, "Jamming

attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 767-809, 2022.

- [3] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 39-50, 2013.

- [4] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient Backscatter Communications: A Contemporary Survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2889-2922, 2018.

- [5] N. Van Huynh, D. N. Nguyen, D. Thai Hoang, E. Dutkiewicz, and M. Mueck, "Ambient Backscatter: A Novel Method to Defend Jamming Attacks for Wireless Networks," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 175-178, 2020.

- [6] L. Jia, N. Qi, Z. Su, F. Chu, S. Fang, K. Wong, and C. Chae, "Game theory and reinforcement learning for anti-jamming defense in wireless communications," *IEEE Communications Surveys & Tutorials*, 2024.

- [7] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *NIPS*, 2017.

- [8] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *ACM Computing Surveys*, vol. 50, no. 2, pp. 1-35, 2017.

- [9] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv:1511.05952*, 2015.

# NGHIÊN CỨU PHÁT TRIỂN HỆ THỐNG PHẦN MỀM NHẬN DIỆN SỐ LƯỢNG BỌ HÀ TRONG BẦY PHỤC VỤ QUẢN LÝ SÂU BỆNH HẠI TRÊN CÂY KHOAI LANG

Lê Hoàng Vũ<sup>1</sup>, Trần Hà Phương<sup>1</sup>, Trần Thị Thu Phương<sup>2</sup>,  
Hà Quang Hưng<sup>3</sup>, Phạm Minh Triển<sup>1</sup>, Vũ Minh Trung\*

<sup>1</sup> Khoa Công nghệ Nông nghiệp, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

<sup>2</sup> Khoa Nông học, Học viện Nông nghiệp Việt Nam

<sup>3</sup> Khoa Công nghệ Thông tin, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

\*Email: trungvm@vnu.edu.vn

## TÓM TẮT

Bọ hà khoai lang (*Cylas formicarius*) là loài côn trùng gây hại phổ biến, gây tổn thất lớn về kinh tế do làm giảm năng suất và chất lượng củ. Nghiên cứu phát triển quy trình nhận diện đối tượng nhỏ mật độ cao nhằm cung cấp công cụ dự báo sớm dịch hại, tối ưu hóa sử dụng thuốc bảo vệ thực vật. Tập dữ liệu gồm 334 ảnh thực địa với mật độ từ 10 đến 100 cá thể mỗi ảnh được gán nhãn theo định dạng YOLO. Thực nghiệm so sánh hiệu năng giữa các phiên bản YOLOv8, v9, v10 và v11 được thực hiện trong cùng điều kiện dữ liệu và cấu hình huấn luyện. Kết quả cho thấy YOLOv11n đạt hiệu quả cao nhất với Precision 99,12%, Recall 97,78%, mAP@0.5 đạt 98,77% và mAP@[0.5:0.95] đạt 77,15%. Với cấu trúc tối ưu gồm 2,65 triệu tham số và 6,6 GFLOPs, mô hình này đảm bảo tính gọn nhẹ cho việc triển khai. Một hệ thống phần mềm đã được phát triển để tích hợp mô hình, thực hiện trực quan hóa kết quả và kiểm chứng khả năng ứng dụng thực tế trong giám sát dịch hại nông nghiệp.

**Từ khóa:** Bọ hà khoai lang, học sâu, phát hiện đối tượng, xử lý ảnh, YOLO.

## ABSTRACT

The sweet potato weevil (*Cylas formicarius* (Fabricius, 1798)) is a destructive pest prevalent in sweet potato cultivation, causing significant economic losses by reducing both yield and tuber quality while remaining difficult to control due to its cryptic life cycle within stems and tubers [1]. This study contributes an effective detection workflow for small objects at high densities, providing a supportive tool for farmers and agricultural managers to facilitate early pest forecasting. Consequently, this enables the optimization of pesticide application and enhances crop productivity. The dataset, comprising 334 images collected and annotated in the You Only Look Once (YOLO) format, features 10-100 individual weevils per image represented by single-class bounding boxes. Model selection was conducted by benchmarking various YOLO versions under identical data and training conditions [2]. Performance was evaluated based on Precision, Recall, mAP@0.5, GFLOPs, and parameter count. YOLOv11n achieved mAP@0.5 of 98.77% and mAP@[0.5:0.95] of 77.15% (with 2.65 million parameters and 6.6 GFLOPs) and was subsequently selected for software integration due to its lightweight architecture and high accuracy. Furthermore, a software system was developed to incorporate the trained model for result visualization, serving as a platform for testing and validating the feasibility of its deployment for agricultural pest monitoring.

**Keywords:** Sweet potato weevil, deep learning, object detection, image processing, YOLO.

**Ngày nhận bài:** 26/01/2026; **Ngày sửa bài:** 20/03/2026; **Ngày duyệt đăng:** 02/04/2026

## 1. ĐẶT VẤN ĐỀ

Bọ hà khoai lang còn khó kiểm soát do vòng đời phát triển chủ yếu diễn ra trong thân và củ, khiến việc phát hiện sớm và can thiệp kịp thời gặp nhiều hạn chế [1]. Do đó, việc ước lượng sớm mật độ dịch hại có ý nghĩa quan trọng trong quản lý dịch hại tổng hợp, góp phần giảm thiểu tổn thất và nâng cao hiệu quả sản xuất nông nghiệp. Việc lựa chọn mô hình được thực hiện thông qua so sánh các phiên bản YOLO khác nhau trong cùng điều kiện dữ liệu và huấn luyện [2].

Trên thực tế, việc theo dõi mật độ bọ hà khoai lang hiện nay chủ yếu dựa vào các phương pháp bẫy thủ công và kiểm tra trực quan, dẫn đến nhiều hạn chế trong khả năng phản hồi sớm và quy mô giám sát [3]. Phương pháp thủ công này phụ thuộc vào kinh nghiệm của người nông dân và chuyên gia, tiêu tốn thời gian và tài nguyên, đồng thời khó áp dụng hiệu quả trong quản lý dịch hại trên diện tích lớn. Các nghiên cứu tổng quan về công nghệ nhận dạng đối tượng trong nông nghiệp chỉ ra rằng việc sử dụng các giải pháp tự động dựa trên thị giác máy tính và học sâu là xu hướng tất yếu để giảm chi phí và tăng độ chính xác trong giám sát dịch hại so với phương pháp thủ công [4]. Trong đó, các mô hình một giai đoạn như YOLO (You Only Look Once) đã được chứng minh là phù hợp cho các nhiệm vụ phát hiện đối tượng trong môi trường thực tế nhờ khả năng xử lý nhanh và hiệu quả với đối tượng nhỏ và nền ảnh phức tạp [5]. Các nghiên cứu gần đây trong lĩnh vực phát hiện sâu bệnh và côn trùng gây hại sử dụng YOLO hay biến thể của nó đã mô tả rõ hiệu quả của kiến trúc này trong môi trường nông nghiệp, nơi đối tượng bị che khuất, kích thước nhỏ và nền ảnh đa dạng là thách thức

thường trực. Ví dụ, các mô hình YOLO đã được áp dụng thành công để phát hiện và đếm sâu bệnh trên nhiều loại cây trồng, với độ chính xác cao và chi phí tính toán hiệu quả, đồng thời hỗ trợ triển khai trên thiết bị có tài nguyên tính toán hạn chế [6]. Phát hiện côn trùng là bài toán đối tượng nhỏ đầy thách thức do đặc trưng hình thái dễ bị triệt tiêu khi đi qua các tầng trích xuất sâu. Vì vậy, các biến thể YOLO hiện đại đã tích hợp cơ chế học đặc trưng đa quy mô (multi-scale feature learning) nhằm bảo toàn và khai thác hiệu quả thông tin của đối tượng ở kích thước tối thiểu [7].

Vì vậy, nghiên cứu này hướng tới: (1) Xây dựng tập dữ liệu ảnh bọ hà khoai lang được gán nhãn thống nhất theo định dạng YOLO, hỗ trợ trực tiếp cho việc huấn luyện mô hình phát hiện đối tượng; (2) Huấn luyện và đánh giá các phiên bản YOLO khác nhau trên cùng tập dữ liệu ảnh bọ hà. So sánh này nhằm đánh giá ảnh hưởng của số lượng tham số (params), số phép toán dấu phẩy động (FLOPs) và các chỉ số về độ chính xác của mô hình như độ chính xác (Precision), độ hồi tưởng (Recall), mAP@0.5 (Mean Average Precision tại ngưỡng Intersection over Union – IoU = 0.5, trong đó IoU đo mức độ chồng lấp giữa hộp dự đoán và hộp nhãn thực) và mAP@0.5:0.95 (giá trị mAP trung bình trên dải ngưỡng IoU từ 0.5 đến 0.95 với bước 0.05), nhằm để xác định cấu hình phù hợp nhất cho bài toán phát hiện và đếm côn trùng nhỏ trong thực địa; (3) Phát triển một hệ thống phần mềm để triển khai mô hình YOLO đã huấn luyện, phục vụ cho việc thử nghiệm và đánh giá khả năng triển khai mô hình trong giám sát dịch hại nông nghiệp. Ứng dụng này có mục tiêu tích hợp mô hình học sâu vào quy trình xử lý ảnh và trực quan hóa kết quả

phát hiện, từ đó minh họa khả năng áp dụng thực tế của mô hình cho bài toán nhận diện bọ hà khoai lang.

## 2. VẬT LIỆU VÀ PHƯƠNG PHÁP

### 2.1 Xây dựng bộ dữ liệu bọ hà khoai lang

Nhằm giải quyết bài toán tự động hóa việc phát hiện và đếm bọ hà khoai lang ngoài đồng ruộng, một bộ dữ liệu hình ảnh đã được xây dựng theo quy trình kiểm soát chặt chẽ [3]. Đầu tiên, nhóm nghiên cứu thực hiện việc sử dụng bẫy mồi pheromone và bẫy chuyên dụng nhằm thu hút các mẫu bọ hà trên cánh đồng khoai lang tại tỉnh Bắc Ninh (xem Hình 1). Sau đó, các mẫu bọ hà sẽ được chụp bằng điện thoại iPhone XS Max với camera thường, không chỉnh sửa thông số. Các ảnh được chụp trong môi trường có độ sáng đồng đều, giúp loại bỏ nhiễu từ ánh sáng thay đổi trong môi trường thực tế. Ảnh thu được có độ phân giải 1920x2560 pixels, đảm bảo độ chi tiết cần thiết cho quá trình huấn luyện và suy luận của mô hình huấn luyện.



Hình 1. Thu thập bọ hà ngoài thực tế

Do dữ liệu ảnh ban đầu được thu thập với kích thước 1920x2560 pixels, trước khi thực hiện các bước tiền xử lý tiếp theo, ảnh được cắt về kích thước 1920x1920 pixels nhằm chuẩn hóa tỷ

lệ khung hình. Bước xử lý này được áp dụng để loại bỏ sự không đồng nhất về hình dạng ảnh đầu vào, đồng thời giảm thiểu sai lệch hình học của đối tượng có thể phát sinh trong quá trình điều chỉnh kích thước ảnh về kích thước đầu vào cố định của mô hình. Việc cắt ảnh về dạng vuông trước khi điều chỉnh kích thước theo cơ chế bảo toàn tỉ lệ khung hình giúp giữ lại đặc trưng hình thái của đối tượng, đồng thời đảm bảo tính nhất quán của dữ liệu trong toàn bộ quá trình huấn luyện và đánh giá mô hình.

Quá trình gán nhãn dữ liệu được thực hiện thủ công bằng nền tảng CVAT (Computer Vision Annotation Tool) trong Hình 2 [8]. Mỗi cá thể bọ hà trong ảnh được gán nhãn thủ công bằng hộp bao quanh và các nhãn này được tổ chức theo định dạng YOLO phục vụ cho quá trình huấn luyện mô hình [9]. Bài toán trong nghiên cứu được thiết lập dưới dạng phát hiện đối tượng một lớp, trong đó toàn bộ các cá thể bọ hà khoai lang được gán chung một nhãn. Trong quá trình gán

nhãn, số lượng cá thể bọ hà trong mỗi ảnh có xu hướng dao động trong khoảng từ 10 đến 100, cho thấy đặc trưng phân bố mật độ cao và mức độ chồng lấn đáng kể giữa các đối tượng trong dữ liệu thực địa. Nhằm đảm bảo chất lượng tập dữ liệu, quá trình gán nhãn được kiểm

soát chặt chẽ để các hộp bao quanh đại diện chính xác từng cá thể bọ hà. Các bước này nhằm giảm thiểu sai lệch và tăng độ chính xác trong quá trình huấn luyện mô hình [10, 11].



Hình 2. Gán nhãn dữ liệu trên nền tảng CVAT

## 2.2. Ứng dụng mô hình học sâu

Nghiên cứu này tập trung so sánh thực nghiệm hiệu năng giữa các phiên bản khác nhau của các biến thể mô hình YOLO trong cùng một bài toán phát hiện bọ hà khoai lang. Cách tiếp cận này nhằm đánh giá mức độ phù hợp của từng phiên bản YOLO đối với bài toán phát hiện và đếm côn trùng kích thước nhỏ, xuất hiện dày đặc trong điều kiện ảnh thực địa. YOLO là dòng mô hình phát hiện đối tượng một giai đoạn nổi bật với khả năng suy luận thời gian thực và kiến trúc gọn nhẹ [2]. Các phiên bản YOLO liên tiếp được giới thiệu với những thay đổi về cấu trúc mạng, cơ chế huấn luyện và chiến lược trích xuất đặc trưng, dẫn đến sự khác biệt đáng kể về độ chính xác, tốc độ suy luận và chi phí tính toán [11].

Để đảm bảo tính công bằng và khách quan trong việc đánh giá hiệu năng, tất cả các mô hình (YOLOv8n, v8s, v9n, v9s, v10n và v11n) đều được huấn luyện và kiểm thử trên cùng một tập dữ liệu chuẩn duy nhất. Quá trình thực nghiệm sử dụng phương pháp Transfer Learning với các trọng số tiền huấn luyện (Pre-trained weights) từ tập dữ liệu COCO, đồng thời áp dụng cấu hình siêu tham số thống nhất cho

mọi phiên bản trong suốt quá trình huấn luyện [11]. Cụ thể, quá trình huấn luyện được thiết lập với 80 epochs, ảnh đầu vào được điều chỉnh kích thước về  $1024 \times 1024$  pixels theo cơ chế bảo toàn tỷ lệ khung hình (letterbox), kích thước lô huấn luyện (batch size) được cố định ở mức 8 nhằm đảm bảo tính ổn định của quá trình tối ưu và sự nhất quán giữa các cấu hình mô hình. Việc cố định các thông số này giúp hạn chế tối đa ảnh hưởng của các yếu tố ngoại lai, đảm bảo sự khác biệt về kết quả đầu ra chỉ phản ánh đặc trưng kiến trúc của từng phiên bản mô hình.

Quá trình thực nghiệm được triển khai trên môi trường Kaggle Notebook với cấu hình phần cứng sử dụng 02 GPU NVIDIA Tesla T4 (16GB GDDR6 mỗi card), hỗ trợ tăng tốc tính toán song song. Các siêu tham số huấn luyện được thiết lập dựa trên cấu hình mặc định của YOLOv11 để đảm bảo tính khách quan khi so sánh giữa các phiên bản, cụ thể như sau: (1) Bộ tối ưu hoá (Optimizer) Sử dụng thuật toán tối ưu hóa tự động (auto), mặc định là SGD với xung lượng (momentum) đạt 0.937 và hệ số suy giảm trọng số (weight decay) là 0.0005. Quá trình khởi động được duy

trì trong 3.0 epoch đầu tiên để ổn định gradient; (2) Tốc độ học (Learning rate) được khởi tạo ở mức 0.01. Hệ thống áp dụng cơ chế giảm cấp tốc độ học với hệ số kết thúc bằng 0.01, giúp mô hình hội tụ sâu hơn ở các giai đoạn cuối của quá trình huấn luyện; (3) Tăng cường dữ liệu (Augmentation) biến đổi không gian màu HSV (H: 0.015, S: 0.7, V: 0.4), dịch chuyển ảnh (0.1), thay đổi tỉ lệ (0.5) và lật ngang ảnh (0.5). Đặc biệt, kỹ thuật Mosaic augmentation được thiết lập ở mức tối đa (1.0).

Tập dữ liệu được chia thành ba tập con bao gồm cả ảnh còn trùng riêng lẻ và chồng lấn: tập huấn luyện/ đánh giá/ kiểm thử tương ứng với 8:1:1. Các mô hình sau khi huấn luyện sẽ được đánh giá trên tập kiểm thử để đo lường khả năng tổng quát hóa trên dữ liệu chưa từng xuất hiện trong quá trình huấn luyện trước đó [12]. Hiệu năng phát hiện của mô hình được đánh giá thông qua hai chỉ số cơ bản là Precision và Recall. Trong đó, Precision phản ánh tỷ lệ các dự đoán dương tính đúng (true positives, TP) trên tổng số dự đoán dương, bao gồm cả các dự đoán sai (false positives, FP), còn Recall thể hiện tỷ lệ các đối tượng thực tế được mô hình phát hiện đúng trên tổng số đối tượng cần phát hiện, bao gồm cả các trường hợp bị bỏ sót (false negatives, FN). Hai chỉ số này được xác định theo các công thức chuẩn trong bài toán phát hiện đối tượng, lần lượt được trình bày trong công thức (1) và công thức (2). Bên cạnh Precision và Recall, các chỉ số mAP@0.5 và mAP@[0.5:0.95] được sử dụng để đánh giá toàn diện năng lực phát hiện và chất lượng định vị đối tượng. Công thức (3) xác định mAP@0.5, tính toán tại ngưỡng chồng lấn IoU = 0,5, nhằm đánh giá khả năng nhận diện đối tượng

ở mức cơ bản và phân biệt mục tiêu với nền môi trường. Trong khi đó, công thức (4) mô tả mAP@[0.5:0.95], giá trị trung bình của mAP trên dải ngưỡng IoU từ 0,5 đến 0,95. Đây là thước đo khắt khe hơn, phản ánh độ chính xác tinh vi trong việc thiết lập khung bao sát hình dáng cá thể, yếu tố then chốt đối với bài toán đếm đối tượng kích thước nhỏ và mật độ dày đặc.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$mAP@0.5 = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3)$$

$$mAP@[0.5:0.95] = \frac{1}{10} \sum_{i=1}^{10} mAP_{IoU} \quad (4)$$

Bên cạnh hai chỉ số trên, nghiên cứu sử dụng chỉ số mAP để đánh giá tổng hợp đồng thời khả năng phát hiện và độ chính xác định vị hộp bao quanh của mô hình. Cụ thể, mAP@0.5 được tính tại ngưỡng chồng lấn IoU cố định bằng 0,5, phản ánh hiệu năng phát hiện trong điều kiện yêu cầu định vị ở mức cơ bản, trong khi mAP@[0.5:0.95] được xác định bằng cách lấy trung bình giá trị AP trên dải ngưỡng IoU từ 0,5 đến 0,95 với bước nhảy 0,05, cho phép đánh giá khắt khe hơn về chất lượng định vị trong các kịch bản phát hiện thực tế. Việc kết hợp Precision, Recall và các chỉ số mAP giúp phản ánh toàn diện hiệu năng của mô hình trong bài toán phát hiện và đếm bọ hà khoai lang.

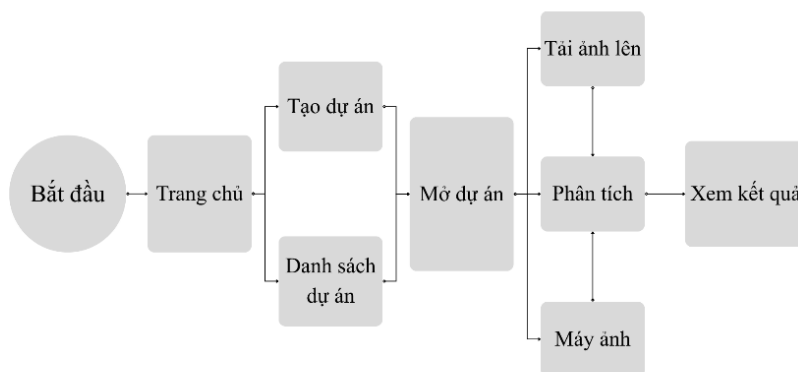
Nghiên cứu đồng thời xem xét chi phí tính toán của từng phiên bản mô hình thông qua hai đại lượng: số lượng tham số và số phép toán dấu phẩy động. Hai chỉ số này được sử dụng như thước đo hỗ trợ để đánh giá mức độ “gọn nhẹ” và yêu cầu tính toán của mô hình. Việc kết hợp các chỉ số độ chính xác và chi

phí tính toán cho phép so sánh hiệu năng mô hình theo hướng toàn diện, phục vụ mục tiêu lựa chọn cấu hình phù hợp cho bài toán phát hiện và đếm bọt hà [13].

### 2.3. Hệ thống phần mềm phát hiện và đếm bọt hà

Trong nghiên cứu này, một hệ thống phần mềm hỗ trợ phát hiện và đếm bọt hà khoai lang được xây dựng nhằm phục vụ quá trình thử nghiệm và minh họa khả năng suy luận của các mô hình YOLO đã huấn luyện. Hệ thống đóng vai trò trung gian giữa mô hình học sâu và dữ liệu ảnh đầu vào, cho phép thực hiện toàn bộ quy trình từ thu ảnh, tiền xử lý, suy luận đến hiển thị kết quả phát hiện. Về mặt kiến trúc chức năng, hệ thống được thiết kế theo hướng mô-đun, bao gồm các thành phần chính: quản lý dự án, thu ảnh đầu vào, xử lý cùng suy luận bằng mô hình tiền huấn luyện và hiển thị kết quả phát hiện. Ảnh đầu vào được người dùng cung cấp thông qua chức năng tải ảnh từ thiết bị lưu trữ hoặc thu nhận trực tiếp từ máy ảnh. Mỗi ảnh được gắn với một dự án cụ thể nhằm phục vụ việc tổ chức và theo dõi quá trình phân tích. Sau khi tiếp nhận, ảnh được đưa qua bước tiền xử lý, bao gồm chuẩn hóa kích thước và định dạng theo yêu cầu của mô hình YOLO đã huấn luyện. Dữ liệu sau tiền xử lý được chuyển tới mô-đun suy luận, tại đó mô hình YOLO thực hiện phát hiện các cá thể bọt hà và sinh ra các khung bao tương ứng. Kết quả suy luận bao gồm vị trí các khung bao và số lượng đối tượng được phát hiện trong mỗi ảnh [2, 11].

Kết quả đầu ra được trực quan hóa trên giao diện người dùng, trong đó các khung bao được hiển thị chồng lên ảnh gốc nhằm biểu diễn vị trí của từng cá thể bọt hà. Luồng xử lý của hệ thống được xây dựng thống nhất với quy trình huấn luyện và đánh giá mô hình, đảm bảo dữ liệu đầu vào và đầu ra của mô-đun suy luận tuân theo cùng một chuẩn định dạng. Quy trình hoạt động tổng thể của hệ thống được minh họa trong Hình 3, bao gồm các bước: khởi tạo hoặc mở dự án, tiếp nhận ảnh đầu vào, thực hiện suy luận bằng mô hình YOLO và hiển thị kết quả phát hiện. Thiết kế này cho phép hệ thống hoạt động độc lập với giai đoạn huấn luyện, tập trung vào việc triển khai và kiểm thử mô hình trong các kịch bản sử dụng ảnh khác nhau. Thông qua giao diện này, người dùng có thể nhanh chóng kiểm tra khả năng phát hiện của mô hình trong các kịch bản ảnh khác nhau, bao gồm ảnh có mật độ đối tượng cao và nền ảnh phức tạp.



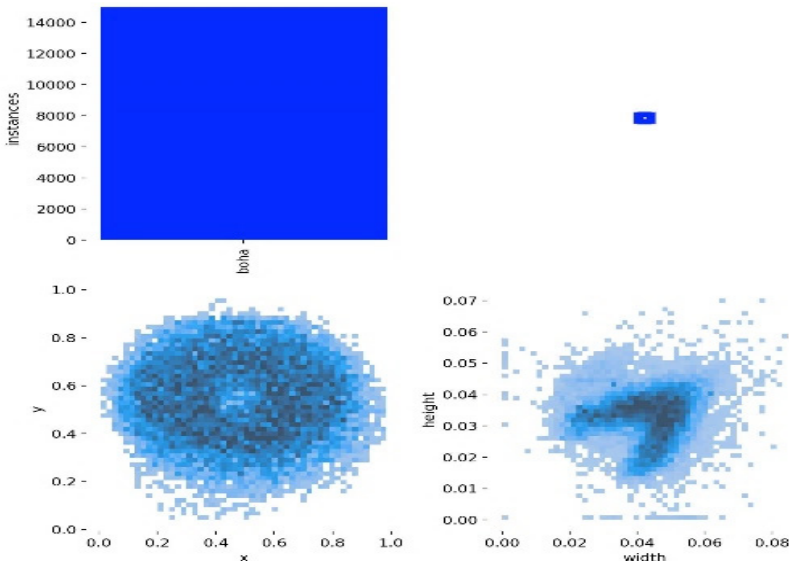
Hình 3. Quy trình hoạt động của ứng dụng

## 3. KẾT QUẢ VÀ BÀN LUẬN

### 3.1. Tập dữ liệu bọt hà khoai lang

Tập dữ liệu bọt hà khoai lang được sử dụng trong nghiên cứu được chia thành ba tập con bao gồm 267 ảnh huấn luyện, 33 ảnh đánh giá và 34 ảnh kiểm thử, với

tổng cộng 334 ảnh. Toàn bộ các ảnh trong ba tập đều có tệp nhãn tương ứng, không tồn tại trường hợp ảnh thiếu nhãn hoặc tệp nhãn rỗng, cho thấy quá trình chuẩn bị và kiểm tra dữ liệu được thực hiện đầy đủ và nhất quán. Tập dữ liệu bao gồm tổng cộng 18.509 hộp bao quanh, trong đó tập huấn luyện chiếm phần lớn với 14.999 nhãn. Phân tích thống kê phân bố nhãn và kích thước hộp bao quanh trong tập huấn luyện, được minh họa trong Hình 4, cho thấy các hộp bao quanh có kích thước tương đối nhỏ so với kích thước ảnh và phân bố không đồng đều trong không gian ảnh. Đặc điểm này phản ánh rõ bản chất của đối tượng cần phát hiện là các cá thể bọ hà kích thước nhỏ, xuất hiện với mật độ cao và thường chồng lấn trong điều kiện ảnh thực địa.



Hình 4. Phân bố nhãn và kích thước hộp bao quanh trong tập huấn luyện

### 3.2. Huấn luyện và so sánh mô hình

Kết quả so sánh hiệu năng giữa các phiên bản YOLO được trình bày trong Bảng 1, có thể nhận thấy các phiên bản YOLO được khảo sát đều đạt hiệu năng phát hiện cao trên tập dữ liệu bọ hà khoai lang, cho thấy tính phù hợp của dòng mô hình phát hiện một giai đoạn đối với bài toán phát hiện và đếm côn trùng mật độ cao. Tuy nhiên, giữa các phiên bản vẫn tồn tại sự khác biệt rõ rệt về mức độ cân bằng giữa độ chính xác phát hiện và chi phí tính toán.

**Bảng 1. So sánh các phiên bản YOLO**

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@[0.5:0.95] (%)	Params (M)	FLOPs (G)
YOLOv8n	97,85	96,36	97,81	71,34	3,01	8,1
YOLOv8s	98,49	97,56	98,17	74,18	11,13	28,4
YOLOv9n	97,38	96,93	98,01	71,82	1,97	7,6
YOLOv9s	98,66	97,45	98,31	74,4	7,17	26,7
YOLOv10n	94,93	93,27	97,34	70,68	2,69	8,2
YOLOv11n	99,12	97,78	98,77	77,15	2,65	6,6

Xét tổng thể các kết quả trình bày trong Bảng 1, YOLOv11n cho thấy hiệu năng toàn diện tốt nhất trong số các mô hình được đánh giá, khi đạt được sự cân bằng hợp lý giữa độ chính xác phát hiện và chi phí tính toán. Mô hình

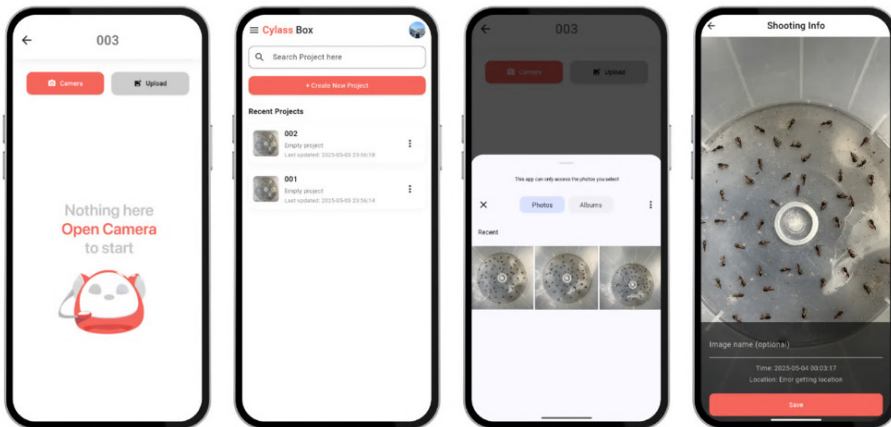
này nổi bật so với các phiên bản YOLO còn lại, đặc biệt trong bối cảnh bài toán

phát hiện và đếm bọ hà khoai lang với mật độ đối tượng cao và kích thước nhỏ. Về hiệu năng phát hiện, YOLOv11n đạt Precision 99,12% và Recall 97,78%, là các giá trị cao nhất trong nhóm các mô hình so sánh. Đồng thời, mô hình đạt  $mAP@0.5 = 98,77\%$ , vượt qua các phiên bản YOLOv8n (97,81%), YOLOv9n (98,01%) và YOLOv10n (97,34%). Đáng chú ý, tại chỉ số  $mAP@[0.5:0.95]$ , YOLOv11n đạt 77,15%, cao hơn rõ rệt so với các mô hình còn lại (từ 70,68% đến 74,40%). Kết quả này cho thấy YOLOv11n có khả năng định vị chính xác hơn trong các điều kiện IoU cao, phù hợp với đặc thù của bài toán phát hiện đối tượng nhỏ và chồng lấn.

Về chi phí tính toán, các phiên bản YOLO cho thấy sự khác biệt rõ rệt giữa mức độ phức tạp mô hình và hiệu năng đạt được. YOLOv11n chỉ sử dụng 2,65 triệu tham số và 6,6 GFLOPs, là một trong những cấu hình nhẹ nhất trong nhóm mô hình được khảo sát, thấp hơn YOLOv8n (3,01 triệu tham số, 8,1 GFLOPs) và YOLOv10n (2,69 triệu tham số, 8,2 GFLOPs), đồng thời giảm đáng kể chi phí so với các phiên bản kích thước lớn

như YOLOv8s (11,13 triệu tham số, 28,4 GFLOPs) và YOLOv9s (7,17 triệu tham số, 26,7 GFLOPs). Mặc dù YOLOv9n có số tham số thấp nhất (1,97 triệu), FLOPs của mô hình này vẫn ở mức 7,6G và không mang lại lợi thế rõ rệt về hiệu năng so với YOLOv11n. Như vậy, YOLOv11n đạt được mức tiết kiệm tài nguyên tính toán đáng kể cả về bộ nhớ mô hình lẫn số phép toán suy luận, trong khi vẫn duy trì hiệu năng phát hiện vượt trội, cho thấy cấu trúc mô hình được tối ưu tốt hơn cho bài toán phát hiện và đếm đối tượng mật độ cao trong điều kiện tài nguyên hạn chế tuy nhiên vẫn tồn tại các sai số đặc trưng: (1) Âm tính giả (False Negative) chỉ số Recall đạt 97,78%, tương ứng tỉ lệ bỏ sót khoảng 2,22%. Lỗi này phát sinh chủ yếu do bọ hà nằm chồng lấn lên nhau hoặc bị khuất ở mép ảnh, khiến mô hình khó trích xuất đủ đặc trưng hình dạng; (2) Dương tính giả (False Positive) với Precision đạt 99,12%, tỉ lệ nhận diện nhầm là 0,88%. Nguyên nhân do các mảnh vụn thực vật hoặc vết bẩn trong ảnh có màu sắc và kích thước tương đồng với đối tượng mục tiêu.

### 3.3. Hệ thống phần mềm phát hiện và đếm bọ hà



Hình 5. Giao diện ứng dụng

Bên cạnh kết quả đánh giá hiệu năng của các mô hình học máy, nghiên cứu đã triển khai một hệ thống phần mềm nhằm tích hợp mô hình YOLO đã huấn luyện vào môi trường ứng dụng thực tế cho bài toán phát hiện và đếm bọ hà khoai lang. Hệ thống đóng vai trò như một công cụ hỗ trợ thử nghiệm và trực quan hóa kết quả suy luận, qua đó kiểm chứng tính nhất quán giữa giai đoạn huấn luyện mô hình và giai đoạn ứng dụng.

Giao diện của hệ thống phần mềm, như minh họa trong Hình 5 cho phép quản lý dữ liệu ảnh theo từng dự án một cách độc lập. Các chức năng chính của hệ thống bao gồm tạo và truy xuất dự án, nhập ảnh đầu vào từ nguồn lưu trữ hoặc thiết bị thu nhận ảnh, cũng như thực hiện suy luận tự động bằng mô hình YOLO đã được huấn luyện. Đối với mỗi ảnh được xử lý, kết quả phát hiện được trực quan hóa thông qua các hộp bao tương ứng với từng cá thể bọ hà, đồng thời cung cấp thông tin về số lượng đối tượng được nhận diện trong ảnh. Việc triển khai hệ thống phần mềm cho phép quan sát trực quan hành vi của mô hình trong các kịch bản ảnh khác nhau, đặc biệt trong điều kiện mật độ đối tượng cao và hiện tượng chồng lấn thường gặp trong dữ liệu thực địa. Thông qua hệ thống này, khả năng chuyển giao mô hình từ môi trường nghiên cứu sang bối cảnh ứng dụng thực tế bước đầu được kiểm chứng ở mức chức năng, tạo tiền đề cho các nghiên cứu tiếp theo hướng tới hoàn thiện hệ thống giám sát dịch hại nông nghiệp dựa trên thị giác máy tính.

#### 4. KẾT LUẬN

Nghiên cứu đã thực hiện so sánh thực nghiệm các phiên bản khác nhau của dòng mô hình YOLO trong bài toán phát hiện và đếm bọ hà khoai lang từ ảnh

thực địa, trên cùng một tập dữ liệu và quy trình huấn luyện thống nhất. Kết quả cho thấy các mô hình có sự khác biệt rõ rệt về độ chính xác phát hiện và chi phí tính toán. Trong số các phiên bản được đánh giá, YOLOv11n thể hiện sự cân bằng tốt nhất giữa hiệu năng phát hiện, độ ổn định định vị và mức chi phí tính toán thấp, qua đó được lựa chọn là cấu hình phù hợp cho bài toán nghiên cứu. Bên cạnh việc xác định mô hình phù hợp, nghiên cứu làm rõ ý nghĩa của việc lựa chọn kiến trúc phát hiện đối tượng trong bối cảnh dữ liệu có mật độ đối tượng cao và điều kiện ảnh thực địa phức tạp. Quy trình so sánh được xây dựng theo hướng thống nhất và có khả năng tái lập, cho phép đánh giá khách quan sự đánh đổi giữa độ chính xác và chi phí tính toán của các phiên bản YOLO khác nhau. Ngoài ra, việc tích hợp mô hình đã huấn luyện vào hệ thống phần mềm phát hiện và đếm bọ hà cho thấy khả năng chuyển giao từ giai đoạn nghiên cứu sang triển khai thử nghiệm trong thực tế.

Tuy nhiên, nghiên cứu vẫn tồn tại một số giới hạn nhất định. Trước hết, tập dữ liệu được xây dựng tập trung vào một loài dịch hại duy nhất và được thu nhận trong điều kiện ghi hình tương đối đồng nhất, do đó chưa phản ánh đầy đủ sự biến thiên của môi trường canh tác ngoài đồng ruộng. Bên cạnh đó, việc đánh giá mô hình trong nghiên cứu chủ yếu dựa trên các chỉ số hiệu năng phát hiện và chi phí tính toán, trong khi các yếu tố liên quan đến tốc độ suy luận thực tế như độ trễ hay số khung hình trên giây trên phần cứng triển khai cụ thể chưa được xem xét. Ngoài ra, bài toán được thiết lập dưới dạng phát hiện một lớp đối tượng nhằm tập trung đánh giá khả năng định

vị và đếm đối tượng trong điều kiện mật độ cao, nên chưa mở rộng sang các kịch bản đa loài hoặc phân loại theo mức độ gây hại. Trong các nghiên cứu tiếp theo, hướng phát triển có thể tập trung vào mở rộng tập dữ liệu với nhiều điều kiện môi trường và đối tượng dịch hại khác nhau, đồng thời đánh giá khả năng triển khai mô hình trên các nền tảng phần cứng hạn chế tài nguyên và trong các kịch bản giám sát gần thời gian thực.

## TÀI LIỆU THAM KHẢO

- [1] S. M. Hue and M. Y. Low, "An Insight into Sweet Potato Weevils Management: A Review," 2015, *Hindawi Limited*. doi: 10.1155/2015/849560.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," May 2016, [Online]. Available: <http://arxiv.org/abs/1506.02640>
- [3] C. M. Schloemer, S. H. Graham, and K. S. Lawrence, "Sweetpotato pest challenges and management options," *J. Integr. Pest Manag.*, vol. 16, no. 1, 2025, doi: 10.1093/jipm/pmaf003.
- [4] Z. Khan, Y. Shen, and H. Liu, "ObjectDetection in Agriculture: A Comprehensive Review of Methods, Applications, Challenges, and Future Directions," Jul. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agriculture15131351.
- [5] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/computers13120336.
- [6] S. Khalid, H. M. Oqaibi, M. Aqib, and Y. Hafeez, "Small Pests Detection in Field Crops Using Deep Learning Object Detection," *Sustainability (Switzerland)*, vol. 15, no. 8, Apr. 2023, doi: 10.3390/su15086815.
- [7] M. Nikouei *et al.*, "Small object detection: A comprehensive survey on challenges, techniques and real-world applications," Sep. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.iswa.2025.200561.
- [8] B. Sekachev *et al.*, "opencv/cvat: v1.1.0," Aug. 2020, *Zenodo*. doi: 10.5281/zenodo.4009388.
- [9] M. Aljabri, M. AlAmir, M. AlGhamdi, M. Abdel-Mottaleb, and F. Collado-Mesa, "Towards a better understanding of annotation tools for medical imaging: a survey," *Multimed. Tools Appl.*, vol. 81, no. 18, pp. 25877–25911, Jul. 2022, doi: 10.1007/s11042-022-12100-1.
- [10] S. A. Magalhães *et al.*, "Evaluating the single-shot multibox detector and yolo deep learning models for the detection of tomatoes in a greenhouse," *Sensors*, vol. 21, no. 10, May 2021, doi: 10.3390/s21103569.
- [11] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review."
- [12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, [Online]. Available: <http://arxiv.org/abs/2004.10934>
- [13] Z.-Q. Zhao, P. Zheng, S. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1807.05511>

# ENHANCING NETWORK SECURITY BASED ON ANOMALY DETECTION USING DEEP LEARNING FOR INTELLIGENT IDS/IPS SYSTEMS

**Hà Trọng Thắng**

*Faculty of Informatics, East Asia University of Technology*

*Email: thanght@eaut.edu.vn*

## **ABSTRACT**

*Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) are critical security solutions designed to monitor and manage network activities in order to detect and respond to attacks or abnormal behaviors. An IDS primarily monitors network traffic and generates alerts upon detecting suspicious activities, whereas an IPS can proactively respond to threats by automatically blocking packets, terminating connections, disabling accounts, or isolating devices from the network.*

*However, traditional IDS/IPS systems are limited in their ability to detect complex anomalous patterns in network traffic and often respond ineffectively in dynamic and heterogeneous network environments.*

*This research proposes an intelligent IDS/IPS framework for LAN environments, integrating a Deep Learning-based Autoencoder to perform anomaly detection on network traffic features. As an unsupervised learning approach, it enables the identification of anomalies without requiring pre-labeled attack data, while an automated alert mechanism and real-time response capability enhance the effectiveness of threat detection, monitoring, and prevention in a more proactive and efficient manner.*

**Keywords:** *Autoencoder, Deep Learning, network monitoring, IDS/IPS, anomaly detection*

---

*Ngày nhận bài: 06/11/2025; Ngày sửa bài: 23/01/2026; Ngày duyệt đăng: 02/04/2026*

## **1. INTRODUCTION**

The rapid development of information technology has led to the emergence of a diverse ecosystem of interconnected devices, offering exceptional convenience to users. Within this context, computer networks play a foundational role, serving as critical infrastructure for global connectivity, data transmission, and information exchange.

The exponential growth in the

number of connected devices and the volume of transmitted data has spurred continuous research in computer networking - particularly in areas related to security and information protection. Today, computer networks are not merely tools for communication, but have become indispensable components across multiple domains such as business, education, healthcare, and media.

In Vietnam, cybersecurity has experienced several notable developments in recent years:

+ Number of cyberattacks: In 2024, more than 659,000 cyberattacks were recorded, with 46.15% of organizations and enterprises having experienced at least one incident [1].

+ Growing sophistication of threats: Cyberattacks have become increasingly sophisticated and complex, highlighting the urgent need to enhance the security of internal networks, including LANs [2]. Raising awareness and investing in robust cybersecurity measures are crucial for data protection, ensuring system stability, and maintaining uninterrupted operations of organizations and businesses.

Therefore, this research focuses on designing an intelligent, LAN-based AI-IDS/IPS system that leverages anomaly detection to enhance the security of internal networks.

## **2. RELATED WORK**

Intrusion Detection and Prevention Systems (IDS/IPS) play a critical role in monitoring network traffic, analyzing behavior, and protecting systems against security threats. IDS (Intrusion Detection System) is responsible for detecting abnormal or unauthorized behaviors, while IPS (Intrusion Prevention System) actively prevents such activities [3]. However, in increasingly complex network environments and with the constant evolution of attack techniques, traditional IDS/IPS systems still exhibit significant limitations [4].

Specifically, traditional IDS only performs monitoring, detection, and alerting when abnormal signs are observed, while IPS is designed to automatically prevent attacks by dropping packets, resetting connections, or

blocking IP addresses. Nevertheless, the effectiveness of IPS still largely depends on accurate detection based on predefined patterns or known signatures [5].

Furthermore, high rates of false positives and false negatives are common issues for both IDS and IPS. False positives consume system resources and affect operational efficiency, while false negatives risk overlooking actual attacks. Properly adjusting detection thresholds in traditional systems remains a major challenge, particularly in environments where traffic patterns and user behavior are continuously changing [6].

Additionally, signature-based IDS/IPS systems rely heavily on databases of known attack patterns. This dependency makes it difficult to detect novel (zero-day) attacks, sophisticated evasion techniques, or obfuscation strategies. These systems are not designed to learn from new data or adapt to emerging threats [7].

Lastly, traditional IDS/IPS systems mainly focus on detecting unauthorized access from external sources, while their ability to identify anomalies within internal networks remains limited [8].

In light of these limitations, this study proposes an integrated solution that combines Artificial Intelligence (AI) and Deep Learning to enhance control capabilities and improve decision-making in threat detection and prevention [9].

## **3. METHODOLOGY**

To effectively enhance cybersecurity, this study proposes an intelligent IDS/IPS system improved by integrating a Deep Learning-based Autoencoder model - an unsupervised learning technique widely used in anomaly detection - for identifying deviations in network traffic [10]. The system is capable of learning

normal network behavior and detecting anomalies through reconstruction error when processing connection sessions. Upon identifying an anomaly, the system triggers an automated alert mechanism via a real-time dashboard or initiates timely responses such as blocking packets from suspicious devices through Telnet communication with network equipment [11].

The combination of deep learning capabilities and real-time response enables the system to proactively detect threats at an early stage, minimize potential risks, and enhance the protection level of the internal network [12].

### 3.1. AI-enhanced Intelligent IDS/IPS System

#### 3.1.1. Overall System Architecture

The operational model of the AI-based IDS/IPS system includes key components such as network data collection, data preprocessing, model training, anomaly detection, and real-time response [13].

#### Network data collection

In this research, the *scan\_system* module was developed to automate data collection from the local network by leveraging Nmap for device scanning and Scapy for traffic analysis. The information collected includes the IP address, MAC address, number of packets, data volume, connection duration, and device status. This data is then stored in the *network\_device\_data.csv* file, which serves both real-time analysis and as input for training the Autoencoder deep learning model.

The structure of the *network\_device\_data.csv* file is illustrated in **Table 1**. Data collection is typically performed over several minutes to several hours to ensure that a sufficient quantity of data is obtained.

Table 1. Field descriptions in the *network\_device\_data.csv* file

Field name	Description
ip_address	IP address of the device
mac_address	Physical MAC address of the device
packets	Total number of packets sent/received during the connection session
bytes	Total amount of transmitted data (in bytes)
duration	Duration of the connection session (since first detection)
connection_time	Timestamp when the system scanned and recorded the device information
status	Status of the device (e.g., <i>Active</i> )

Data preprocessing and feature extraction

The collected data are preprocessed by normalizing numerical features such as *packets*, *bytes*, and *duration* using the MinMaxScaler technique, which transforms the raw values into a standard range [0, 1] according to Eq. (1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where,

-  $x$  is the original value, and  $x_{min}$ ,  $x_{max}$  are the minimum and maximum values of the feature in the dataset.

-  $x'$ : The normalized value of the feature, scaled to the range [0, 1]

Categorical features, such as *status*, are encoded using One-hot Encoding. As a result, the dataset is transformed into numerical vectors  $x \in \mathbb{R}^n$ , where  $n$  is the total number of features after encoding. These vectors are then fed into the Autoencoder deep learning model for training. This preprocessing step helps reduce noise and improve the model's accuracy.

Training and anomaly detection using Autoencoder

The Autoencoder model used in this research consists of three main layers: an input layer, two hidden layers (encoder/decoder), and an output layer. It is trained to reconstruct normal network traffic data, with reconstruction errors used to detect anomalies during actual operation.

Let  $x \in \mathbb{R}^n$  denote the input feature vector. The encoder compresses the input into a lower-dimensional latent representation  $\mathbf{z}$ :

$$\mathbf{z} = f_{\text{enc}}(x) = \sigma(W_e x + b_e) \quad (2)$$

where,

-  $f_{\text{enc}}$ : The encoder function that maps to

-  $W_e$ : The encoder weight matrix, learned during training

-  $b_e$ : The bias vector of the encoder layer

-  $\sigma(\cdot)$ : A nonlinear activation function (ReLU) that models nonlinear relationships between inputs and the latent representation

The decoder then reconstructs the input from this latent vector:

$$\hat{x} = f_{\text{dec}}(\mathbf{z}) = \sigma(W_d \mathbf{z} + b_d) \quad (3)$$

where,

-  $\hat{x}$ : The reconstructed output vector

-  $f_{\text{dec}}$ : The decoder function that maps back to the input space

-  $W_d$ : Weight matrix of the decoder

-  $b_d$ : The bias vector of the decoder layer

The network is trained to minimize the Mean Squared Error (MSE) between the input  $x$  and the reconstruction  $\hat{x}$ :

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (4)$$

where,

-  $\mathcal{L}_{MSE}$ : The Mean Squared Error loss value, measuring the average squared difference between input and reconstructed output

-  $n$ : The number of elements (dimensions) in the feature vector

-  $x_i$ : The  $i$ -th component of the original input vector

-  $\hat{x}_i$ : The  $i$ -th component of the reconstructed output vector

After training, the model is deployed in real-time monitoring. During inference, any input sample that yields a reconstruction error greater than a predefined threshold  $\theta$  is flagged as an anomaly:

If  $\mathcal{L}_{MSE}(x, \hat{x}) > \theta \Rightarrow$  Anomaly Detected

The threshold  $\theta$  is empirically determined based on the reconstruction error distribution of the normal training data.

❖ **The model architecture includes the following components**

- Input layer: This layer consists of 4 neurons corresponding to the normalized input features derived from real-world network data, including: *packets*, *bytes*, *duration*, and *status*. Among these, the quantitative features (*packets*, *bytes*, and *duration*) are normalized using the MinMaxScaler technique, whereas the categorical feature (*status*) is encoded using one-hot encoding.

- Encoder: The encoder comprises two consecutive hidden layers with 6 and 3 neurons, respectively, each utilizing the ReLU activation function. This component is responsible for reducing the input dimensionality and extracting essential features into a lower-dimensional latent representation.

- Latent space: Located between the encoder and the decoder, this layer contains 3 neurons that represent the compressed representation of device characteristics in a low-dimensional space.

- Decoder: The decoder mirrors the encoder, consisting of two layers with 3 and 6 neurons, respectively, also employing ReLU activation. This stage

reconstructs the original input data from the compressed latent representation.

- Output layer: This layer contains 4 neurons - equal to the input layer - and uses a linear activation function. Its objective is to produce reconstructed outputs that closely approximate the original inputs. The reconstruction error (difference between input and output) is used as the basis for anomaly detection.

- Training configuration: The model is trained over 100 epochs, with a batch size of 16, and a learning rate of 0.001, using the Adam optimizer and the Mean Squared Error (MSE) loss function.

- Mathematical description: Mathematically, the model learns a mapping function  $f: x \rightarrow \hat{x}$  such that the reconstruction error  $\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2$  is minimized for normal input data.

- Anomaly detection mechanism: During the inference phase, if a data sample  $x$  produces a reconstruction error exceeding a predefined threshold  $\theta$ , it is flagged as anomalous. The threshold  $\theta$  is determined based on the distribution of reconstruction errors on the normal training data using a percentile-based method, as shown in Eq. (5):

$$\theta = \text{percentile}(\mathcal{L}(x, \hat{x}), 95) \quad (5)$$

where the function  $\text{percentile}(\cdot, 95)$  returns the 95th percentile of the error distribution on normal data, and different threshold values may significantly affect the trade-off between false positives and false negatives.

#### ❖ Operating mechanism

An overview of the anomaly detection

model based on the Autoencoder is shown in **Figure 1**. The Autoencoder, a type of deep neural network, is capable of learning to encode and decode input data such that the reconstructed output closely matches the original input. For anomaly detection, the Autoencoder is trained in an unsupervised fashion exclusively on normal data. When it encounters anomalous data, the reconstruction error tends to be high, as the model is not familiar with the patterns present in abnormal samples [14].

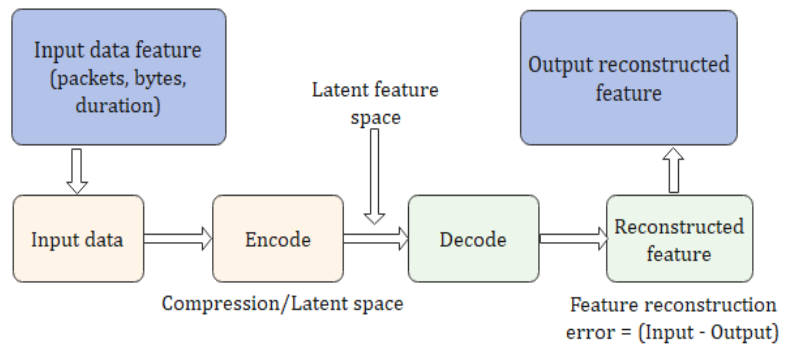


Figure 1. Autoencoder-based anomaly detection model

Specifically, the model is trained on sessions of normal network connections to learn a compressed representation of their features. During operation, new connection sessions are input into the model to calculate their reconstruction error; if this error exceeds a predefined threshold, the session is flagged as anomalous. The system is architected with two independent modules: *train\_model* and *anomaly\_detector*. This modular design allows the model to be updated periodically, flexibly, and without causing interruptions to the operational system.

#### d) Alerting and automated response

When an anomaly is detected, the AI-IDS/IPS system automatically logs the event, generates an alert on the web dashboard, and can proactively block

related traffic by interfacing with network devices via Telnet, thereby minimizing potential risks.

This mechanism ensures real-time anomaly detection with low latency, while effectively reducing the false positive rate through the use of a deep learning Autoencoder model implemented with the TensorFlow library [15].

3.1.2. Operational workflow of the AI-IDS/IPS system

The operational workflow of the AI-IDS/IPS system consists of several stages, including network data collection, data preprocessing, model training, anomaly detection, and real-time response. This process is illustrated in Figure 2.

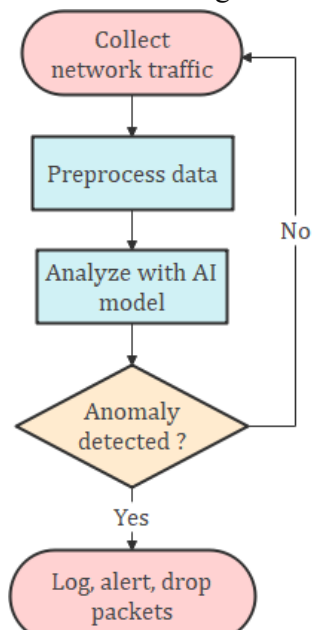


Figure 2. Operational workflow of the AI-IDS/IPS system

3.2. Deployment of the LAN model

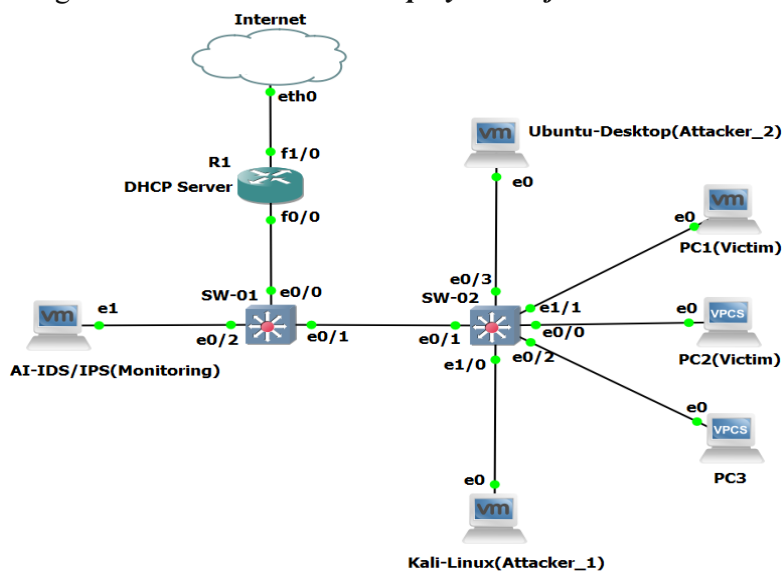


Figure 3. LAN network system model

In this research, the network model is simulated using GNS3 (Figure 3), where the AI-IDS/IPS server functions as the monitoring system. The Kali-Linux and Ubuntu-Desktop machines are used to perform various experimental attacks. Tools like *Nmap* and *hping3* are commonly used to simulate different types of network attacks, such as TCP SYN Flood and ICMP Flood, for security testing purposes.

❖ LAN network model

The LAN network model consists of the following components:

- Router (R1): Configured as a DHCP server with Internet access.
- Switches (SW-01, SW-02): Configured with SPAN (Switched Port Analyzer) to monitor and analyze network traffic by mirroring data from source ports to a destination port connected to the monitoring server (AI-IDS/IPS).
- Monitoring Server (AI-IDS/IPS) - Ubuntu Server: Performs network traffic monitoring, anomaly detection, and alert

generation.

- Clients (PC1, PC2): Target machines used in security testing scenarios.

- Attacker\_1 (Kali-Linux): Simulates different types of attacks.

- Attacker\_2 (Ubuntu-Desktop): Simulates additional attack scenarios.

❖ **System operation workflow**

In the LAN network model, the system comprises a Router (R1) configured as a DHCP server, providing IP address allocation and Internet access. Two switches (SW-01 and SW-02) operate in SPAN (Switched Port Analyzer) mode, mirroring network traffic to the AI-IDS/IPS monitoring server. The server

captures all mirrored traffic and performs real-time anomaly detection using a deep learning-based Autoencoder. Attacker machines (Kali-Linux and Ubuntu-Desktop) simulate cyberattacks - such as TCP SYN Flood and ICMP Flood. The monitoring system visualizes alerts via a web-based dashboard and can dynamically block malicious hosts by issuing Telnet commands to network devices. Client machines (PC1 and PC2) serve as attack targets, providing a realistic environment for testing and validating the model. This flexible detection and response mechanism allows the network to continue functioning normally while effectively mitigating both internal and external threats.

❖ **Operation of the monitoring system (AI-IDS/IPS)**

On the AI-IDS/IPS monitoring

server (Figure 4), the system provides a Web Dashboard interface developed using FastAPI, integrated with Tailwind CSS and WebSocket technology to display device information tables, alert statistics charts, and real-time updates. Administrators can monitor device statuses, review the history of detected anomalies, and view the response actions performed by the system.

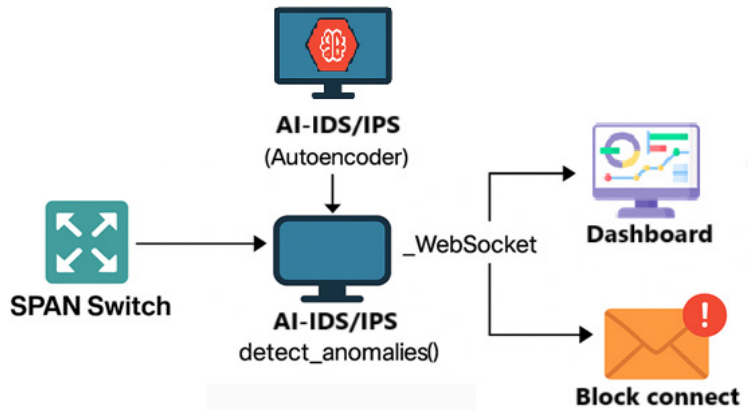


Figure 4. Operational model of the monitoring system (AI-IDS/IPS)

**4. EXPERIMENTAL RESULTS AND DISCUSSION**

**4.1. System testing**

**4.1.1. Data collection and AI model training**

The *scan\_system* module collects network device data over a predefined period of 30 seconds (Figure 5), with estimated scanning durations detailed in Table 2. The system determines device connection duration using a cache file (*device\_cache.json*), which records when each device is first detected and calculates its total time on the network. All device data is normalized and saved to *network\_device\_data.csv*. This module enables the system not only to monitor real-time traffic but also to build a comprehensive dataset for more effective AI model training and anomaly detection.

```
(notebook-env) root@ainac:/home/admin2025/notebook-env/bin/AI-NAC# python3 scan_system.py
||=> Starting network monitoring (Press Ctrl+C to stop)
||=> [17:30:08] Scanning with Nmap...
||=> Sniffing packets on interface 'ens37' for 30s...
OK ! Saved 8 records to network_device_data.csv
```

Figure 5. Device scanning and information collection interface

Estimated time for device scanning:

Table 2. Estimated time for device scanning

IP range (Subnet mask)	Number of IPs	Sequential scanning	Multithreaded scanning (scapy hoặc asyncio)
/30	4	~2s	<2s
/24	256	~1-2 minutes	~5-10 seconds

/22	1024	~8-10 minutes	~20-30 seconds
/16	65.536	>1 hour	~5-10 minutes

The statistical chart showing the ratio of normal/abnormal devices, along with a table displaying device information (IP address, MAC address, packet count, data volume, duration, and status), is illustrated in Figure 6 on the Web Dashboard.

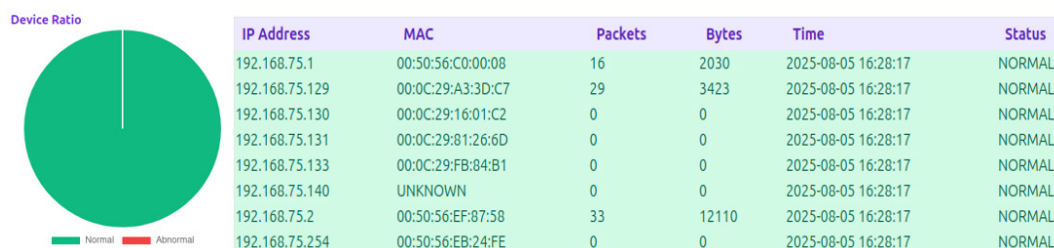


Figure 6. Statistical chart of normal and abnormal device status on the Dashboard

The AI-IDS/IPS web dashboard interface displays statistics including the total number of devices, the number of anomalous devices, and the last update

time. It also presents a chart visualizing the reconstruction error of each device across specified error ranges, as illustrated in Figure 7.

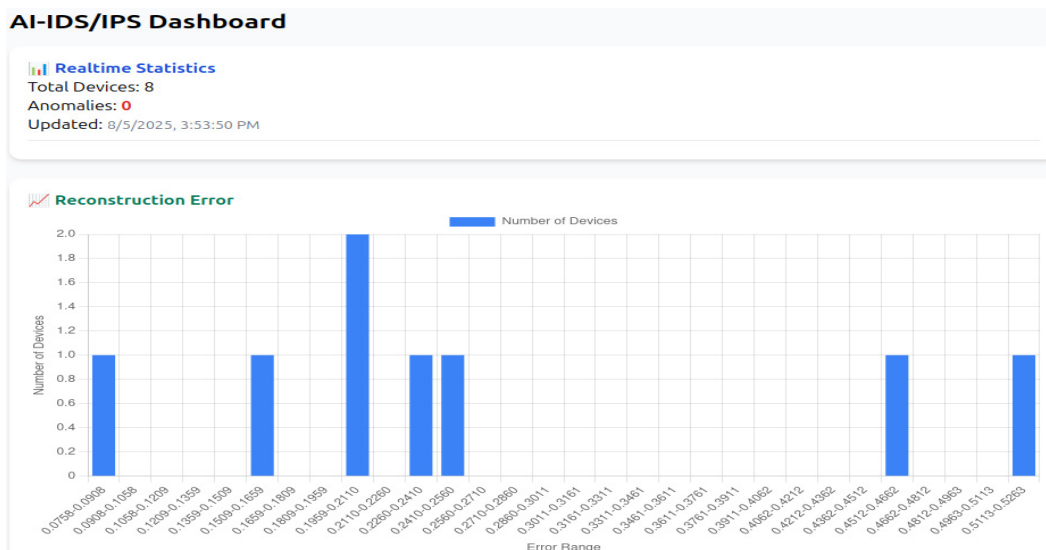


Figure 7. Device count statistics and reconstruction error chart

4.1.2. Attack experiment and response of the AI-IDS/IPS

During the performance evaluation of the AI-IDS/IPS system, real-world attack experiments were conducted to assess the model’s capability to detect and respond to common network threats. Two representative types of attacks and their respective tools were chosen to represent widely used attack techniques, including:

- ICMP Flood
- TCP SYN Flood

❖ Scenario 1: Denial of Service attack - ICMP Flood

In this attack scenario, the Ubuntu-Desktop machine (IP 192.168.75.130) uses the *hping3* tool to perform a Denial of Service (DoS) attack by flooding the target device with a large number of ICMP Echo Request (ping) packets. The target machine is PC1, with IP address 192.168.75.11. This attack may cause disruption or even crash the victim system.

Figure 8 shows the interface where the command is executed:

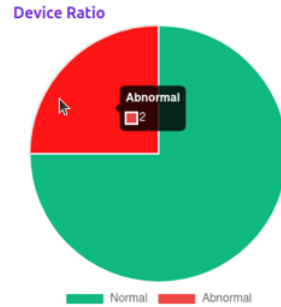
```
root@admin2025:/home/admin2025# hping3 -1 --flood 192.168.75.11
HPING 192.168.75.11 (ens33 192.168.75.11): icmp mode set, 28 headers + 0 data by tes
hping in flood mode, no replies will be shown
```

hping3 -1 --flood 192.168.75.11  
 Figure 8. Command Line Interface on the Ubuntu-Desktop PC

Response of the AI-IDS/IPS System:

The monitoring system detected a sudden spike in ICMP packets from a single source within a short timeframe. The AI model recognized this as an anomaly, triggering an immediate alert on the Web Dashboard. As illustrated

in Figure 9, the dashboard shows one anomalous device with IP address 192.168.75.130 - the attacking machine - sending 95,322 packets totalling 5,719,320 bytes to the target machine.



IP Address	MAC	Packets	Bytes	Time	Status
192.168.75.1	00:50:56:C0:00:08	7	2986	2025-08-05 21:11:49	NORMAL
192.168.75.11	00:0C:29:FB:84:B1	95322	5719320	2025-08-05 21:11:49	ANOMALY
192.168.75.129	00:0C:29:A3:3D:C7	8	3076	2025-08-05 21:11:49	NORMAL
192.168.75.130	00:0C:29:16:01:C2	95322	5719320	2025-08-05 21:11:49	ANOMALY
192.168.75.131	00:0C:29:81:26:6D	0	0	2025-08-05 21:11:49	NORMAL
192.168.75.140	UNKNOWN	0	0	2025-08-05 21:11:49	NORMAL
192.168.75.2	00:50:56:EF:87:58	1	90	2025-08-05 21:11:49	NORMAL
192.168.75.254	00:50:56:E1:3B:17	0	0	2025-08-05 21:11:49	NORMAL

Figure 9. Anomaly detection and alerts displayed on the AI-IDS/IPS Dashboard

❖ Scenario 2: Denial-of-Service Attack - TCP SYN Flood and ICMP Flood

In this scenario, a coordinated attack is simulated from two machines - Kali-Linux (IP 192.168.75.131) and Ubuntu-Desktop (IP 192.168.75.130) - simultaneously targeting the victim

machine, PC1 (IP 192.168.75.11). The Kali-Linux machine performs a ICMP Flood operation using *hping3* to overload the CPU and bandwidth of PC1, while the Ubuntu-Desktop machine conducts a TCP SYN Flood attack using *hping3* against the same target. Figure 10 displays the command execution interface and the results of the scan:

```
hping3 -1 --flood 192.168.75.11
```

```
(root@kali)-[~]
└─# hping3 -1 --flood 192.168.75.11
HPING 192.168.75.11 (eth0 192.168.75.11): icmp mode set, 28 headers + 0 data bytes
hping in flood mode, no replies will be shown
```

Figure 10. Probing and Information Gathering on PCI

Figure 11 shows the command execution interface from the Ubuntu-Desktop machine: `hping3 -S -p 80 --flood 192.168.75.11`

```
root@admin2025:/home/admin2025# hping3 -S -p 80 --flood 192.168.75.11
HPING 192.168.75.11 (ens33 192.168.75.11): S set, 40 headers + 0 data bytes
hping in flood mode, no replies will be shown
```

Figure 11. Command Line Interface on the Ubuntu-Desktop machine

**Response of the AI-IDS/IPS System:**

In Figures 12 and 13, the system detected a sudden spike in TCP SYN connection frequency across multiple ports, flagging it as an anomaly. An alert was triggered with the logged IP

address of the suspicious device, which is displayed on the Web dashboard. The information table lists two anomalous devices, accompanied by a reconstruction error chart of all monitored devices.

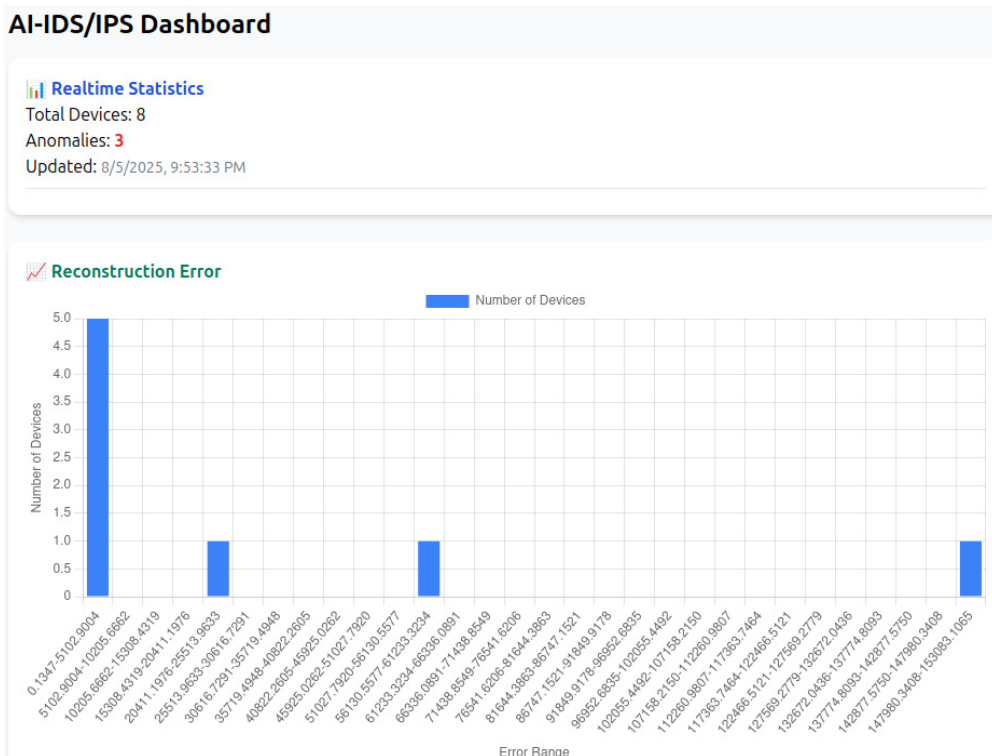
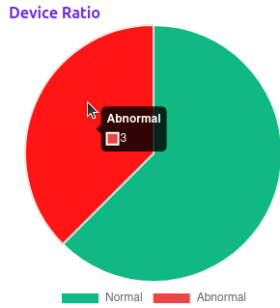


Figure 12. Statistics of detected anomalous devices and reconstruction error chart



IP Address	MAC	Packets	Bytes	Time	Status
192.168.75.1	00:50:56:C0:00:08	5	1673	2025-08-05 21:50:48	NORMAL
192.168.75.11	00:0C:29:FB:84:B1	38588	2315473	2025-08-05 21:50:48	ANOMALY
192.168.75.129	00:0C:29:A3:3D:C7	4	1598	2025-08-05 21:50:48	NORMAL
192.168.75.130	00:0C:29:16:01:C2	14362	861720	2025-08-05 21:50:48	ANOMALY
192.168.75.131	00:0C:29:81:26:6D	24227	1453828	2025-08-05 21:50:48	ANOMALY
192.168.75.140	UNKNOWN	0	0	2025-08-05 21:50:48	NORMAL
192.168.75.2	00:50:56:EF:87:58	0	0	2025-08-05 21:50:48	NORMAL
192.168.75.254	00:50:56:E1:3B:17	0	0	2025-08-05 21:50:48	NORMAL

Figure 13. AI-IDS/IPS anomaly detection and alerting on the Dashboard

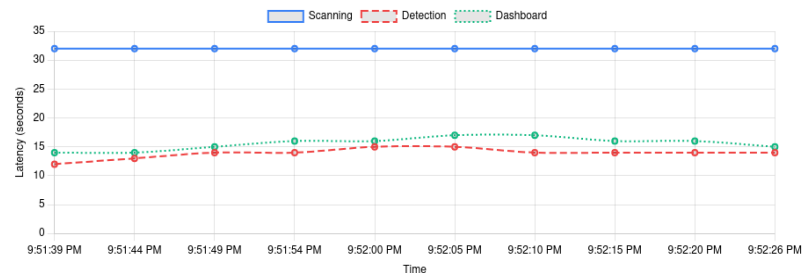
4.1.3. Evaluation of anomaly detection and AI response effectiveness

The performance evaluation of the AI-IDS/IPS system is divided into several key stages: network scanning latency (*scan\_system*), anomaly detection latency (using Autoencoder and MLP models), and the alert display latency via WebSocket on the dashboard. Measurements were conducted using the time library to record the start and end times of the *scan\_system()* function. Additionally, timestamps at each processing phase - including packet reception, anomaly detection, alert transmission via WebSocket, and device control response via Telnet - were logged through the system's logging mechanism. The Dashboard interface measures real-time display latency using the *onmessage()* event of WebSocket and

\* System Performance Monitoring (seconds)

- 1. System Scanning: 32
- 2. Anomaly Detection: 14
- 3. Dashboard Display: 15

\* System Performance Chart (seconds)



the timestamp at the DOM (Document Object Model) rendering time on the frontend. The measured results are aggregated and displayed directly on the dashboard to support real-time performance monitoring.

Specifically, as shown in Figure 14, system performance monitoring is presented with two axes: latency (in seconds) and time intervals measured every 5 seconds. Through experimental evaluation, the

average time from the start of the attack to the completion of the system's response ranges between 25 and 35 seconds. The anomaly detection and alerting process takes approximately 10-15 seconds. Following detection, the system sends an alert through WebSocket to the dashboard and initiates a Telnet connection from the AI-IDS/IPS device to the switch to block packets, which takes another 10-15 seconds.

Therefore, the AI-IDS/IPS system design (shown in Figure 3) effectively demonstrates the capability to detect anomalies and respond effectively within a short time frame.

Figure 14. Performance monitoring chart of system alerts on the Dashboard

However, in large-scale and wide-area network models with many devices, latency increases significantly in subnet ranges of /22 (1,024 IP addresses) and /16 (65,536 IP addresses) if the scanning scope is not subdivided. Additionally, AI inference on the GPU needs to be optimized, as it may cause CPU resource bottlenecks when handling multiple concurrent connections and tasks.

To reduce latency in large-scale network systems, several solutions should be implemented: subdividing subnets to reduce the processing load per segment; increasing the number of AI-IDS/IPS monitoring units deployed in local areas; and optimizing GPU inference processes to offload the CPU.

**4.2 Quantitative evaluation of model performance**

Autoencoder is an unsupervised deep learning model used to extract anomalous features from network data via reconstruction error. By applying a predefined threshold to this score, each session is automatically assigned a binary pseudo label (0: normal, 1: anomalous). These pseudo labels are then used to train a supervised classification model, specifically a Multi Layer Perceptron (MLP), which aims to approximate the anomaly definition induced by the Autoencoder.

The test dataset consists of 12,438 network sessions collected over a period of three days, covering three traffic types: Ping, ICMP Flood, and Port Scan. This dataset was assigned binary pseudo-labels (0: normal, 1: anomalous) in a preprocessing stage that applied an Autoencoder-based reconstruction-error threshold to each session. After training the MLP classifier on 80% of the dataset,

the model was applied to the remaining 20% to predict output labels. The MLP model outputs a probability (ranging from 0 to 1) representing the likelihood that a session is anomalous. This probability is then thresholded at 0.5 to assign binary labels (if the output > 0.5, the session is labeled as anomalous). The corresponding command is:  $y_{pred} = (mlp.predict(X_{test}) > 0.5).astype(int)$ .

Next, the confusion matrix is computed using the confusion\_matrix function from the sklearn.metrics library, as shown below:

```
from sklearn.metrics import
confusion_matrix
cm = confusion_matrix(y_test, y_pred)
```

$TN, FP, FN, TP = cm.ravel()$

The confusion matrix yields four key values - TP, FP, FN, and TN - which are presented in **Table 3**:

- TP (True Positive) = 1,812: Number of anomalous sessions correctly detected by the model.
- FP (False Positive) = 284: Number of normal sessions incorrectly identified as anomalous.
- FN (False Negative) = 188: Number of anomalous sessions missed by the model (predicted as normal).
- TN (True Negative) = 10,154: Number of normal sessions correctly identified by the model.

**Performance evaluation metrics:**

*Table 3. Performance evaluation metrics*

Metric	Meaning	Formula	Value
Precision	Precision of anomaly detection: The ratio of correctly predicted anomalies to the total number of predicted anomalies, reflecting the false alarm rate.		0,865

Metric	Meaning	Formula	Value
Recall	Recall (Detection Rate): The proportion of correctly detected anomalies out of all actual anomalies, reflecting the model's ability to avoid missing attacks.		0,906
Accuracy	Overall accuracy: The ratio of correctly classified sessions to the total number of sessions.		0,962
F1-score	Balanced metric (F1 Score): The harmonic mean of Precision and Recall, balancing the trade-off between false alarms and missed attacks.		0,885

The results indicate that the model achieved a high accuracy of 96.2% and a high Recall of 90.6%, effectively meeting the requirements for early anomaly detection.

The obtained Precision, Recall, Accuracy, and F1-score thus reflect how consistently the MLP reproduces the Autoencoder-based anomaly definition over the collected sessions, suggesting its potential suitability for near real-time network monitoring, particularly for DoS flood scenarios tested in the simulated LAN environment.

### 5. CONCLUSION

This research proposes an intelligent AI-IDS/IPS model that utilizes an Autoencoder - an unsupervised deep learning architecture - for real-time anomaly detection in network traffic. Through experimental evaluations involving common attack types such as TCP SYN Flood and ICMP Flood, the system demonstrated effective

capabilities in detecting anomalies and responding promptly by logging events, issuing alerts, and blocking traffic from malicious sources.

In terms of performance, the system exhibits a noticeable increase in processing latency within large-scale network environments, particularly when operating across wide subnets or under high device density conditions. Additionally, since the detection mechanism primarily relies on statistical features of network packets, it may face difficulties in identifying stealthy or long-duration attacks (commonly referred to as low and slow attacks).

Future research will focus on optimizing the system's processing infrastructure - especially by leveraging AI inference on GPUs to reduce the computational load on CPUs. Enhancements will also include the integration of Reinforcement Learning techniques and User Behavior Analytics (UBA) to improve the detection of high-level anomalous behaviors. Furthermore, plans include expanding the training dataset to accommodate additional attack patterns such as ARP Spoofing, Brute Force attacks and HTTP Floods, as well as deploying the AIIDS/IPS system under edge or hierarchical architectures to ensure scalability while maintaining high performance and responsiveness.

With these improvements, the system is expected to achieve greater accuracy, enhanced adaptability, and higher reliability, ultimately contributing to the development of a secure, proactive, and resilient network environment capable of addressing modern cybersecurity challenges.

---

### REFERENCES

- [1] K. An, “Hơn 659.000 vụ tấn công mạng cơ quan, doanh nghiệp năm 2024,” *Lao Động*, 2024. [Online]. Available: <https://laodong.vn/ban-doc/hon-659000-vu-tan-cong-mang-co-quan-doanh-nghiep-nam-2024-1439390.ldo>
- [2] A. L. Giri and S. Annamalai, “Intrusion detection system for local networks - A review study,” in *Proc. 2nd Int. Conf. on Advances in Computing, Innovation and Technology in Engineering (ICACITE)*, pp. 1388–1391, 2022.
- [3] A. S. Ashoor and S. D. Gore, “Difference between Intrusion Detection System (IDS) and Intrusion Prevention System (IPS),” *Communications in Computer and Information Science*, vol. 151, pp. 1–6, Springer, India, Jul. 2011.
- [4] M. M. Issa, M. Aljanabi, and H. M. Muhialdeen, “Systematic literature review on intrusion detection systems: Research trends, algorithms, methods, datasets, and limitations,” *Journal of Intelligent Systems*, vol. 33, no. 1, pp. 15–26, 2024.
- [5] H.-Y. Kwon, T. Kim, and M.-K. Lee, “Advanced intrusion detection combining signature-based and behavior-based detection methods,” *Electronics*, vol. 11, no. 6, pp. 2–4, 2022.
- [6] F. Wang and L. Yao, “A deep learning approach for intrusion detection system with reduced false positives,” *Journal of Network and Computer Applications*, vol. 155, pp. 2–10, 2020.
- [7] J. F. Canola Garcia and G. E. Taborda Blandon, “A deep learning-based intrusion detection and prevention system for detecting and preventing denial-of-service attacks,” *IEEE Access*, vol. 10, pp. 83050–83055, 2022.
- [8] S. Kumar, S. Gupta, and S. Arora, “Research trends in network-based intrusion detection systems: A review,” *IEEE Access*, vol. 9, pp. 157761–157774, 2021.
- [9] S. S. Raghavan, “A comprehensive study of Artificial Intelligence applications in Intrusion Detection and Prevention,” *Int. J. of AI Research and Development (IJAIRD)*, vol. 3, no. 1, pp. 16–37, Jan.–Jun. 2025.
- [10] V. N. Kumar, D. Siri, and A. Badhoutiya, “Anomaly detection in provenance data using autoencoder network,” in *Proc. IEEE Conf. on Intelligent Systems*, 2024. ISBN: 979-8-3303-6810-9.
- [11] H. Torabi, S. L. Mirtaheri, and S. Greco, “Practical autoencoder-based anomaly detection by using vector reconstruction error,” *Cybersecurity*, vol. 6, art. 1, 2023.
- [12] S. Ang, S. Huy, and M. Janarthanan, “Utilizing IDS and IPS to Improve Cybersecurity Monitoring Process,” *Journal of Cyber Security and Risk Auditing*, vol. 2025, no. 3, pp. 77–88, 2025.
- [13] A. Thakkar and R. Lohiya, “A survey on intrusion detection system: Feature selection, model, performance measures, application perspective, challenges, and future research directions,” *Artificial Intelligence Review*, vol. 55, no. 4, pp. 470–510, 2022.
- [14] Asif Ahmed Neloy, Maxime Turgeon, “A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs,” *Machine Learning with Applications* 17, 100572, July 2024
- [15] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn and TensorFlow 2*, 3rd ed., Birmingham, U.K.: Packt Publishing, 2019.

# TRANSFORMING BUSINESS EDUCATION IN IT: THE ROLE OF MIS IN SHAPING FUTURE LEADERS

**Nguyen Thi Thuy Nga**

*Faculty of Information Technology, East Asia University of Technology,  
Bacninh, Vietnam*

*Email: ngantt@eaut.edu.vn*

## **ABSTRACT**

*Integrating Management Information Systems (MIS) into business education is crucial for preparing future leaders to navigate technological advancements. The digital revolution, characterized by cloud computing, social media, and mobile systems, has fundamentally transformed business operations and educational delivery. MIS courses equip students with essential technological skills, fostering an environment where proficiency becomes second nature. Innovative tools like MyMISLab, video libraries, and decision-making simulations enhance student engagement by bridging theory and practice. Additionally, teaching ethics and security within MIS curricula is vital; case studies provide insights into the ethical and security challenges of the digital age, preparing students for real-world dilemmas. MIS education also helps students identify emerging job roles by developing non-routine cognitive skills such as critical thinking and problem-solving, essential for data-driven decision-making. Thus, integrating MIS into business curricula is necessary to ensure graduates are equipped to meet future business demands, emphasizing real-world applications and job preparedness..*

**Keywords:** *MIS, MyMISLab, Leaders, Transforming Business Education, IT, Cloud computing*

**Ngày nhận bài:** 18/11/2025; **Ngày sửa bài:** 09/01/2026; **Ngày duyệt đăng:** 02/04/2026

## **1. Introduction**

In today's rapidly evolving business landscape, characterized by relentless technological advancements and increasing digitalization, the importance of a robust Management Information Systems (MIS) education for aspiring business professionals cannot be overstated. This article delves into the critical role of MIS in modern business education, highlighting a unique approach that moves beyond theoretical concepts to cultivate practical skills and future job preparedness.

Traditionally, business education has focused on core disciplines like

finance, marketing, and operations. However, the pervasive influence of technology necessitates a paradigm shift, integrating MIS as a fundamental pillar of the curriculum (Gorgone et al., [1]). This integration is not merely about understanding technical jargon but about equipping students with the knowledge and skills to leverage technology for strategic decision-making, problem-solving, and innovation. As highlighted by Bai et al. [2], "The digital transformation of business demands a new generation of business professionals who are not only tech-savvy but also possess strong analytical and problem-solving skills,

along with an understanding of how technology can be used to create business value.” This underscores the need for a pedagogical shift that emphasizes real-world applications and hands-on learning experiences, aligning with the call for greater relevance and practicality in business education (AACSB, [3]).

The business world is experiencing a technological renaissance, fueled by advancements in artificial intelligence (AI), big data analytics, cloud computing, and the Internet of Things (IoT), among others. These technologies are fundamentally reshaping business models, operational processes, and customer interactions (Bharadwaj et al., [4]). This digital transformation necessitates a corresponding evolution in business education. The 2020 Future of Jobs report by the World Economic Forum highlights an increasing need for skills like analytical reasoning, advanced problem solving, as well as technology design and programming.

Researchers share this perspective, stressing that graduates must have not only technical expertise but also adaptability, innovation, and critical thinking abilities to succeed in a world driven by technology (DeNisi Murphy, [5]).

This article presents a forward-looking approach to integrating MIS into business education. It emphasizes real-world applications, future job preparedness, and innovative teaching tools that enhance student engagement and critical thinking.

**Real-World Applications:** Moving beyond theoretical frameworks, the focus is on equipping students with practical skills through case studies, simulations, and real-world projects, allowing them to apply MIS principles to solve contemporary business challenges.

This is consistent with Kolb’s [6] experiential learning theory, which underscores the significance of hands-on experiences and reflective observation in fostering effective learning. Recent studies highlight the effectiveness of project-based learning and industry collaborations in enhancing student engagement and developing practical skills in MIS education (Ribble Park, [7]).

**Future Job Preparedness:** The aim is to close the gap between academia and the workforce by aligning the curriculum with evolving technologies and industry trends, ensuring that graduates have the skills and expertise valued by employers in the digital era. This requires close collaboration between universities and industry partners to ensure curriculum relevance and provide students with valuable work-integrated learning opportunities (Hausman et al., [8]). The rapid pace of technological change necessitates a dynamic and responsive curriculum that incorporates the latest advancements and industry best practices (Ahmad et al., [9]).

**Innovative Teaching Tools:** The article examines the incorporation of advanced technologies and teaching methods, including gamification, virtual reality, and data analytics tools, to develop immersive and interactive learning experiences that promote critical thinking and problem-solving abilities. This approach resonates with the principles of active learning and constructivism, which emphasize student engagement and knowledge construction (Prince Felder, [10]). Recent research highlights the positive impact of gamified learning and virtual reality simulations on student motivation, engagement, and knowledge retention in MIS courses (Yildirim et al., [11]).

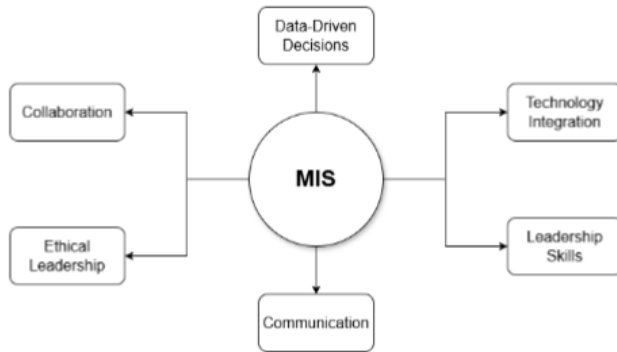


Figure 1. Architecture MIS System Overview

By embracing this innovative approach to MIS education, we can empower future business leaders to navigate the complexities of a technology-driven world, fostering a generation of proactive problem solvers and strategic thinkers equipped to thrive in the digital age in Figure 1. The diagram illustrates the key components of a MIS. At the center is MIS, which is interconnected with six essential elements. Data-Driven Decisions involve using data analytics to inform strategic choices. Technology Integration refers to the seamless incorporation of new technologies to enhance system efficiency. Leadership Skills are necessary for guiding teams and ensuring successful project outcomes. Communication is vital for effective information exchange within an organization. Ethical Leadership emphasizes the importance of integrity and ethical standards in management. Finally, Collaboration highlights the need for teamwork and cooperative efforts across different departments. Together, these components contribute to the effective functioning of an MIS.

This study adopts a qualitative approach based on a systematic review of existing literature combined with conceptual analysis. The objective of the paper is threefold: (i) to examine the

evolving role of Management Information Systems (MIS) in transforming business education; (ii) to analyze how MIS contributes to leadership development, ethical awareness, and technological adaptability; and (iii) to propose an integrated framework linking innovative learning tools, ethical and

security considerations, and future career preparedness.

## 2. The Digital Revolution in Education

The digital revolution has significantly transformed various sectors, including education. Key technologies such as cloud computing, social media, and mobile

systems are reshaping how educators teach and how students learn. This paper explores these technologies and discusses how Management Information Systems (MIS) courses prepare students for future technological changes.

### 2.1. Cloud Computing in Education

Cloud computing offers scalable and flexible resources, enabling educational institutions to store vast amounts of data and provide access to learning materials anywhere, anytime. According to Jones [12], cloud platforms like Google Classroom and Microsoft Azure facilitate collaborative learning environments, allowing for seamless interaction between students and teachers. The scalability of cloud services ensures that schools can efficiently manage resources without significant upfront investment.

### 2.2. Social Media's Educational Impact

Social media platforms provide innovative ways to engage students

and foster communication. Smith [13] highlight that platforms like Twitter and Facebook can be used for educational purposes, such as sharing resources, hosting discussions, and enhancing community building. Social media also allows for realtime feedback and greater interaction, making learning more dynamic and interactive.

### **2.3. Mobile Systems and Learning**

Mobile systems have transformed the way educational content is accessed. With the widespread use of smartphones and tablets, students now have the ability to engage with learning materials anytime, anywhere. Research by Alam [14] shows that mobile learning applications offer personalized learning experiences, catering to individual student needs and promoting self-paced learning. Apps like Duolingo and Khan Academy exemplify how mobile systems facilitate interactive and engaging education.

Preparing for Technological Changes: The Role of MIS Courses MIS courses equip students with the skills needed to navigate and leverage these technological advancements. They focus on the intersection of technology and business, preparing students to implement and manage information systems effectively.

### **2.4. Curriculum Focus**

MIS courses cover a range of topics, including database management, cybersecurity, and data analytics. Johnson and Kim [15] emphasize that understanding these areas is crucial for future professionals who will manage and protect digital infrastructures.

### **2.5. Hands-On Experience**

Students participate in hands-on projects that mirror real-world situations, enhancing their critical thinking and problem-solving abilities. This

experiential learning method equips students with the skills needed to address technological challenges after graduation.

### **2.6. Emphasis on Adaptability**

The dynamic nature of technology requires adaptability. MIS courses stress the importance of continuous learning and staying updated with emerging technologies. By fostering this mindset, courses enable students to stay adaptable and relevant in an ever-evolving digital environment.

In conclusion, the integration of cloud computing, social media, and mobile systems is reshaping education, offering new opportunities for teaching and learning. MIS courses play a pivotal role in preparing students for these technological changes by providing a comprehensive understanding of information systems and fostering essential skills. As the digital revolution continues, the ability to adapt and innovate will be crucial for success in the educational sector.

## **3. Innovative Learning Tools**

In recent years, the educational landscape has been transformed by innovative learning tools such as MyMISLab, video libraries, and decision-making simulations. These tools are designed to enhance student engagement and improve learning outcomes by providing interactive and immersive learning experiences. This paper explores these tools and their impacts, drawing on recent literature to highlight their effectiveness.

### **3.1. MyMISLab**

MyMISLab is an online platform that offers a range of resources for learning management information systems (MIS). It provides interactive tutorials, assessments, and personalized study plans. A study by Bossman et al. [16] found

that MyMISLab significantly improved student engagement and performance in MIS courses, as it allowed students to learn at their own pace and receive immediate feedback. This aligns with the constructivist learning theory, which emphasizes active student participation and self-directed learning (Mohammed et al. [17]).

### **3.2. Video Libraries**

Video libraries, such as those offered by Khan Academy and Coursera, provide a vast array of instructional videos covering numerous subjects. These resources cater to diverse learning styles and offer students the flexibility to learn anytime and anywhere. According to Serra et al. [18], the use of video libraries has been shown to increase student motivation and understanding of complex topics, as visual and auditory elements can enhance comprehension and retention. Furthermore, video resources support differentiated instruction, allowing educators to address the varying needs of their students (Cagape, [19]).

### **3.3. Decision-Making Simulations**

Decision-making simulations are dynamic scenarios that give students the opportunity to apply theoretical knowledge in real-world situations. These simulations are particularly effective in fields such as business, healthcare, and engineering. Research by Lin et al. [20] indicates that simulations improve critical thinking and problem-solving skills by providing a safe environment for experimentation and exploration. Students engage more deeply with the material when they can see the real-world implications of their decisions, leading to improved learning outcomes (Gredler, [21]).

### **3.4. Impact on Student Engagement**

Engagement is a crucial factor in

successful learning. Innovative tools like MyMISLab, video libraries, and simulations have been shown to increase student engagement by making learning more interactive and enjoyable. As noted by Johnson and Frances [22], students are more likely to participate in courses that incorporate these tools, resulting in higher attendance and lower dropout rates. Moreover, these tools foster a sense of autonomy and competence, which are key components of intrinsic motivation (Ryan et al. [23]).

### **3.5. Impact on Learning Outcomes**

The impact of innovative learning tools on learning outcomes has been widely documented. Studies consistently show that these tools contribute to better academic performance and deeper understanding of course material. For instance, Johnson and Frances [22] found that students using MyMISLab scored higher on assessments compared to those who relied solely on traditional textbooks. Similarly, Cagape et al. [19] reported that students who supplemented their studies with video libraries demonstrated improved comprehension and retention. Moreover, decision-making simulations have been found to enhance cognitive skills and knowledge application. Park et al. [24] highlighted that students who participated in simulations were better equipped to transfer theoretical knowledge to practical situations, a key indicator of successful learning.

### **3.6. Challenges and Considerations**

Despite their benefits, the implementation of innovative learning tools poses certain challenges. Access to technology and the digital divide remain significant barriers for some students (Smith et al., [25]). Additionally, educators must be adequately trained to integrate these tools effectively into their

curriculum (Csenturk, [26]). Therefore, institutions must ensure that both students and teachers have the necessary resources and support to maximize the benefits of these tools.

In conclusion, innovative learning tools such as MyMISLab, video libraries, and decision-making simulations play a vital role in enhancing student engagement and improving learning outcomes [27–30]. By providing interactive and personalized learning experiences, these tools cater to diverse learning preferences and promote active participation. While challenges exist, the potential benefits of these tools in shaping the future of education are undeniable. As technology continues to evolve, so too will the opportunities for creating more effective and engaging learning environments.

#### **4. Ethical and Security Considerations**

In today's digital age, the importance of teaching ethics and security in MIS cannot be overstated. As businesses increasingly rely on technology, professionals must navigate complex ethical dilemmas and security challenges. This paper explores the significance of integrating ethics and security into MIS curricula, supported by recent literature and illustrative case studies.

##### ***4.1. Importance of Teaching Ethics in MIS***

Ethics in MIS involves understanding the moral implications of technology use and data management. With the rapid advancement of technology, ethical considerations have become increasingly complex. Teaching ethics in MIS equips students with the ability to recognize and address moral issues, fostering responsible decision-making.

##### ***4.2. Ethical Decision-Making***

Ethical decision-making is a crucial skill for MIS professionals. According to Dellaportas et al. [31], students exposed to ethical training are better prepared to handle real-world situations involving privacy, intellectual property, and user consent. By understanding ethical frameworks, students can assess the implications of their actions and make informed choices that align with societal values.

Case Study: Cambridge Analytica The Cambridge Analytica scandal serves as a cautionary tale. The misuse of Facebook user data for political advertising highlighted significant ethical breaches in data handling. As noted by Hess [32], this case underscores the necessity for robust ethical guidelines and education in MIS programs to prevent similar occurrences.

##### ***4.3. Importance of Teaching Security in MIS***

Security is a fundamental aspect of MIS, focusing on safeguarding information systems from unauthorized access, misuse, disclosure, disruption, alteration, or destruction. With the growing complexity of cyber threats, the need for skilled security experts has significantly increased.

##### ***4.4. Developing Security Competence***

Integrating security education into MIS curricula is essential for developing competence in safeguarding information systems. According to Alnajim et al, [33], students trained in cybersecurity principles are more adept at identifying vulnerabilities and implementing effective security measures. This knowledge is critical in protecting sensitive data and maintaining organizational integrity.

Case Study: Target Data Breach

The Target data breach of 2013, which compromised 40 million credit and debit card accounts, exemplifies the devastating impact of inadequate security measures. The breach, costing the company millions, illustrates the importance of comprehensive security education. As highlighted by Meadows et al, [34], teaching students about such incidents emphasizes the **real-world** consequences of security lapses.

### ***4.5. Integrating Ethics and Security into MIS Curricula***

To address the challenges posed by ethical and security issues, educators must integrate these topics into the MIS curriculum effectively. This involves a combination of theoretical knowledge and practical application.

### ***4.6. Curriculum Design***

A well-rounded MIS curriculum should include courses on ethical theories, privacy laws, cybersecurity fundamentals, and risk management. Arnal [35] suggest incorporating case studies, simulations, and ethical dilemmas to enhance learning experiences. By engaging students in real-world scenarios, educators can foster critical thinking and problem-solving skills.

### ***4.7. Interdisciplinary Approach***

An interdisciplinary approach can enrich MIS education by incorporating perspectives from law, philosophy, and computer science. According to Camilleri et al, [36], this approach encourages students to view ethical and security issues from multiple angles, promoting a deeper understanding of the complexities involved. In conclusion, teaching ethics and security in MIS is imperative for preparing students to navigate the challenges of the digital

world. By integrating these topics into the curriculum, educators can equip future professionals with the skills and knowledge necessary to make ethical decisions and protect information systems. As technology continues to evolve, the role of ethics and security in MIS education will only grow in importance, shaping responsible and competent leaders in the field.

## **5. Preparing for Future Careers**

MIS play a crucial role in preparing students for emerging job roles in the rapidly evolving job market. As technology continues to advance, new opportunities arise that require a unique blend of skills. This paper explores how MIS helps students identify and prepare for these roles, emphasizing the importance of developing non-routine cognitive skills.

### ***5.1. Identifying Emerging Job Roles***

The integration of MIS in education provides students with insights into current and future trends in the job market. By studying MIS, students gain a comprehensive understanding of how technology drives business innovation and creates new career opportunities.

### ***5.2. Understanding Industry Trends***

MIS courses often include analyses of industry trends and technological advancements. According to Kim, Youngsik and Lee, Hojun [37], students who engage with these analyses can better anticipate emerging job roles and align their career paths accordingly. For instance, the rise of data analytics, cloud computing, and cybersecurity has led to new roles such as data scientists, cloud architects, and security analysts.

### ***5.3. Networking and Professional Connections***

MIS programs frequently offer

opportunities for networking with professionals in the field. These connections can provide valuable insights into emerging job roles and industry demands. Rutman, Jeremy [38] highlight the significance of internships and industry partnerships in helping students gain firsthand experience and understand the skills required for future roles.

#### **5.4. Preparing for Emerging Job Roles**

Preparing for the future job market requires more than just technical skills. MIS education emphasizes the development of a diverse skill set, including both technical and non-technical competencies.

#### **5.5. Technical Skills**

MIS programs equip students with essential technical skills in areas such as database management, systems analysis, and information security. These skills are foundational for many emerging roles. Johnson and Miller [39] note that proficiency in programming languages, data analysis tools, and cybersecurity measures is crucial for students aiming to excel in technology-driven careers.

#### **5.6. Non-Routine Cognitive Skills**

In addition to technical skills, MIS education stresses the importance of non-routine cognitive skills—those that involve creativity, problem-solving, and critical thinking. These skills are increasingly valuable in a job market where routine tasks are often automated.

#### **5.7. Problem-Solving and Critical Thinking**

The ability to solve complex problems and think critically is essential in navigating the challenges of new job roles. MIS courses often include case studies and project-based learning to

enhance these skills.

According to Arnal [35], students who engage in problem-solving activities develop a mindset that is adaptable and innovative, key traits for success in dynamic environments.

#### **5.8. Creativity and Innovation**

Creativity is another crucial non-routine cognitive skill fostered by MIS education. As businesses seek innovative solutions to remain competitive, the demand for creative thinkers has surged. Smith [13] emphasizes that MIS students trained to approach problems creatively can develop novel solutions that drive business growth and transformation.

#### **5.9. The Significance of Non-Routine Cognitive Skills**

Non-routine cognitive skills are crucial as they distinguish human abilities from automated systems. With the ongoing advancements in artificial intelligence and machine learning, the capacity to handle tasks requiring human judgment is becoming more valuable than ever.

#### **5.10. Adaptability in a Changing Job Market**

Adaptability is a key outcome of developing non-routine cognitive skills. The ability to learn and apply new skills quickly is essential in a job market characterized by constant change. Arnal [35] argue that adaptable individuals are better equipped to transition between roles and industries, making them more resilient in their careers.

#### **5.11. Enhanced Collaboration and Communication**

Non-routine cognitive skills also enhance collaboration and communication, crucial elements in many emerging job roles. MIS education often involves team projects and interdisciplinary work, which foster these

skills. According to Kim, Youngsik and Lee, Hojun [37], effective communication and collaboration are vital for roles that require cross-functional teamwork and interaction with diverse stakeholders.

In summary, MIS education is essential in readying students for future careers by enabling them to recognize and prepare for emerging job roles. By integrating both technical and non-routine cognitive skills, students are well-prepared to manage the complexities of today's job market. As technology advances, these skills will become increasingly important, highlighting the crucial role of MIS in shaping the workforce of tomorrow.

### 6. Conclusion

In this part, we conclude that the Imperative of Integrating MIS into Business curriculum is as such:

**Importance of MIS in Business Education:** Integrating MIS into business curriculum is crucial for preparing students to thrive in a technology driven world. As businesses increasingly rely on data and digital processes, understanding MIS becomes essential for future leaders. This paper reiterates the importance of MIS in education and calls educators to embrace technology in teaching.

**Equipping Students with Essential Skills:** MIS education provides students with technical and analytical skills necessary for modern business environments. According to Johnson and Miller [39], MIS courses teach students how to manage information systems, analyze data, and implement technology solutions. These skills are indispensable as companies seek employees who can leverage technology for strategic advantage.

**Enhancing Decision-Making Capabilities:** Understanding MIS also enhances decision-making capabilities.

By learning how to interpret data and use information systems, students can make informed decisions that drive business success. Kim, Youngsik and Lee, Hojun [37] emphasize that MIS education bridges the gap between technical expertise and strategic thinking, enabling students to become effective decision-makers in complex scenarios.

**The Role of Educators:** Educators are key in embedding MIS into business curricula. By adopting technology, they can develop dynamic learning environments that mirror the demands of the contemporary workplace.

**Innovative Teaching Methods:** Educators are encouraged to adopt innovative teaching methods that incorporate technology. This includes using digital tools, simulations, and case studies to provide hands-on experience. Rutman, Jeremy [38] suggest that interactive learning not only engages students but also enhances their understanding of complex concepts.

**Preparing Students for the Future:** By integrating technology into teaching, educators prepare students for future challenges. The rapid pace of technological advancement means that students must be adaptable and proficient in using new tools. Arnal [35] argue that educators who embrace technology foster a culture of continuous learning, equipping students to navigate an everchanging job market.

**Call to Action:** To remain relevant and effective, business curriculum must evolve with technological advancements. Educators are urged to: (1) Integrate MIS into Core Courses: Ensure that MIS is a fundamental component of business education, providing students with the skills and knowledge needed in today's digital economy. (2) Embrace Technology

in Teaching: Utilize digital resources and innovative teaching strategies to enhance learning experiences and prepare students for the future. (3) Foster Collaboration with Industry: Establish partnerships with businesses to keep curricula aligned with industry needs and trends.

Promote Lifelong Learning: Encourage students to view education as a continuous process, essential for staying current with technological changes.

This paper contributes to the literature by providing a structured conceptual synthesis of how Management Information Systems (MIS) transform business education in the digital era. By integrating perspectives on innovative learning tools, ethical and security education, and non-routine cognitive skill development, the study offers practical implications for curriculum design and academic–industry alignment. As technology continues to reshape industries, students equipped with MIS knowledge and skills will be better prepared to lead and innovate. Educators have a responsibility to embrace technology in their teaching practices, ensuring that their students are ready to meet the demands of the future workplace. By doing so, they will cultivate a new generation of business leaders capable of harnessing the power of information systems to drive success.

---

### References

[1] J. U. Kim and R. Kishore, “Do we fully understand information systems failure? an exploratory study of the cognitive schema of is professionals,” *Information systems frontiers*, vol. 21, no. 6, pp. 1385–1419, 2019.

[2] J. Heredia, M. Castillo-Vergara, C. Geldes, F. M. C. Gamarra, A. Flores, and

W. Heredia, “How do digital capabilities affect firm performance? the mediating role of technological capabilities in the “new normal”,” *Journal of Innovation & Knowledge*, vol. 7, no. 2, p. 100171, 2022.

[3] C. L. Beck Dudley, “Aacsb: Paving new ways in business education,” *Executive education after the pandemic: A vision for the future*, pp. 89–97, 2022.

[4] A. Bharadwaj, O. A. El Sawy, P. A. Pavlou, and N. v. Venkatraman, “Digital business strategy: toward a next generation of insights,” *MIS quarterly*, pp. 471–482, 2013.

[5] A. S. DeNisi and K. R. Murphy, “Performance appraisal and performance management: 100 years of progress?” *Journal of applied psychology*, vol. 102, no. 3, p. 421, 2017.

[6] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[7] J. Ball and M. Savin-Baden, “Postdigital learning for a changing higher education,” *Postdigital science and education*, vol. 4, no. 3, pp. 753–771, 2022.

[8] N. Hausman, “University innovation and local economic growth,” *Review of Economics and Statistics*, vol. 104, no. 4, pp. 718–735, 2022.

[9] S. F. Ahmad, M. K. Rahmat, M. S. Mubarik, M. M. Alam, and S. I. Hyder, “Artificial intelligence and its role in education,” *Sustainability*, vol. 13, no. 22, p. 12902, 2021.

[10] M. J. Prince and R. M. Felder, “Inductive teaching and learning methods: Definitions, comparisons, and research bases,” *Journal of engineering education*, vol. 95, no. 2, pp. 123–138, 2006.

[11] İ. Yıldırım and S. Şen, “The effects

of gamification on students' academic achievement: A meta-analysis study," *Interactive Learning Environments*, vol. 29, no. 8, pp. 1301–1318, 2021.

[12] F. Jonske, K. Osthues, A. Dada, E. Nasca, J. Fragemann, J. Alff, O. Bakumenko, M. Birnbach, M. Kondratenko, L. Reinike et al., "Designing and implementing an interactive cloud platform for teaching machine learning with medical data," in *Artificial Intelligence for Security: Enhancing Protection in a Changing World*. Springer, 2024, pp. 259–292.

[13] T. Smith and A. Short, "Needs affordance as a key factor in likelihood of problematic social media use: Validation, latent profile analysis and comparison of tiktok and facebook problematic use measures," *Addictive Behaviors*, vol. 129, p. 107259, 2022.

[14] A. Alam and A. Mohanty, "Educational technology: Exploring the convergence of technology and pedagogy through mobility, interactivity, ai, and learning tools," *Cogent Engineering*, vol. 10, no. 2, p. 2283282, 2023.

[15] K. F. Johnson, H. Kim, C. E. Molina, K. A. Thompson, S. Henry, and B. Zyromski, "School counseling prevention programming to address social determinants of mental health," *Journal of Counseling & Development*, vol. 101, no. 4, pp. 402–415, 2023.

[16] A. Bossman and S. K. Agyei, "Technology and instructor dimensions, e-learning satisfaction, and academic performance of distance students in ghana," *Heliyon*, vol. 8, no. 4, 2022.

[17] S. Mohammed and L. Kinyo, "Constructivist theory as a foundation for the utilization of digital technology in the lifelong learning process," *Turkish Online Journal of Distance Education*,

vol. 21, no. 4, pp. 90–109, 2020.

[18] L. Serra-Marín and N. Berbel-Gómez, "The impact of the use of technology on student engagement and motivation in the music classroom," *Journal of Music, Technology & Education*, vol. 14, no. 2-3, pp. 157–178, 2021.

[19] W. Cagape, H. Batao, F. Atillo, M. G. Tampus, C. But-ay, and J. Malnegrof, "Differentiated instruction practices: A disclosure of receiving teachers," *Differentiated Instruction Practices: A Disclosure of Receiving Teachers*, vol. 127, no. 1, pp. 21–21, 2023.

[20] P.-H. Lin, P.-S. Huang, Y. Geng, and Y.-M. Huang, "Enhancing problem-based learning with computational thinking concepts for nursing students in virtual simulation context: Bridging knowledge and practice," *Education and Information Technologies*, pp. 1–21, 2024.

[21] J. Gredler and D. Harland, "Online faculty's use of technology when advising doctoral capstone writers," *Australasian Journal of Educational Technology*, vol. 38, no. 3, pp. 43–57, 2022.

[22] F. V. Johnson, "Teachers' perceptions on the influence of kahoot! on the mathematics achievement of high school students," Ph.D. dissertation, The Chicago School of Professional Psychology, 2023.

[23] R. M. Ryan, J. J. Duineveld, S. I. Di Domenico, W. S. Ryan, B. A. Steward, and E. L. Bradshaw, "We know this much is (meta-analytically) true: A meta-review of metaanalytic findings evaluating self-determination theory." *Psychological Bulletin*, vol. 148, no. 11-12, p. 813, 2022.

[24] S. Park, H. J. Shin, H. Kwak, and H. J. Lee, "Effects of immersive

technology-based education for undergraduate nursing students: Systematic review and metaanalysis using the grading of recommendations, assessment, development, and evaluation (grade) approach,” *Journal of Medical Internet Research*, vol. 26, p. e57566, 2024.

[25] E. R. Smith, K. Hodgkin, and N. Young, “Student transitions to university in wales: A mixed-method study of the enablers and barriers of first-year engagement,” *International Journal of Educational Research Open*, vol. 3, p. 100216, 2022.

[26] B. Şentürk, B. A. Dünya, and M. C. Demir, “Training 21<sup>st</sup> century english language teachers in turkish context: Development of a technology-enhanced measurement curriculum,” *Journal of Measurement and Evaluation in Education and Psychology*, vol. 14, no. Özel Sayı, pp. 345–358, 2023.

[27] K. C. Laudon and J. P. Laudon, *Essentials of management information. New Jersey: Prentice-Hall. Mayasari, PDS, & Sadeli, D.(2016). The . . .*, 2003.

[28] K. Laudon and J. Laudon, *Management Information Systems: International Edition*, 11/E. Citeseer, 2009.

[29] W. Sause, J. Woodside, and S. Amiri, “Applying virtual reality to business simulations for e-learning,” in *EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE)*, 2016, pp. 1720–1724.

[30] L. Slusky, “Lessons from online teaching,” in *EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE)*, 2019, pp. 1742–1756.

[31] S. Dellaportas, L. A. Nguyen, and D. Luo, “Fact or fiction? elevating accounting students’ ethical sensitivity

and moral reasoning with ethics education interventions,” *Research Handbook on Accounting and Ethics*, pp. 64–82, 2023.

[32] J. L. Hess, D. Kim, and N. D. Fila, “Critical incidents in ways of experiencing ethical engineering practice,” *Studies in Engineering Education*, vol. 3, no. 2, pp. 1–30, 2023.

[33] A. M. Alnajim, S. Habib, M. Islam, H. S. AlRawashdeh, and M. Wasim, “Exploring cybersecurity education and training techniques: a comprehensive review of traditional, virtual reality, and augmented reality approaches,” *Symmetry*, vol. 15, no. 12, p. 2175, 2023.

[34] J. J. Meadows III, “Mandatory gamified security awareness training impacts on texas public middle school students: A qualitative study,” Ph.D. dissertation, Colorado Technical University, 2023.

[35] M. Arnal-Palacián and J. M. Johnson, “Analysis of interdisciplinary stem lessons generated by pre-service and inservice teachers in the united states,” *Tech. Rep.*, 2023.

[36] S. Camilleri and A. Bezzina, “Adopting the circle pedagogy-relatedness, autonomy and competence,” *Pastoral Care in Education*, vol. 41, no. 4, pp. 449–472, 2023.

[37] Y. Kim and H. Lee, “Investigating the effects of career education programs on high school students’ career development competencies in korea,” *Sustainability*, vol. 15, no. 18, p. 13970, 2023.

[38] J. Rutman, “Exploring the role of internships in personal and professional development,” 2023.

[39] A. Johnson and S. Miller, “Total learning architecture standards digital learning acquisition techniques.” *Advanced Distributed Learning Initiative*, 2023.

# SỬ DỤNG THUẬT TOÁN QP CỦA BÀI TOÁN TỐI ƯU TOÀN PHƯƠNG ĐỂ XÂY DỰNG BỘ ĐIỀU KHIỂN MPC CHO HỆ PHI TUYẾN

**Phan Xuân Minh**

Trường Đại học Quốc tế, Đại học Quốc gia Hà Nội

Email: minh.phanxuan@hust.edu.vn

## TÓM TẮT

Điều khiển dự báo dựa theo mô hình (Model Predictive Control – MPC) là phương pháp điều khiển được sử dụng rất rộng rãi trong thực tế, mang lại hiệu quả tin cậy. Nó ra đời từ rất sớm, được xuất phát từ bài toán thực tế, dưới tên gọi Thuật toán điều khiển tựa mô hình (Model Algorithm Control – MAC) và dần phát triển thành một công cụ ưa thích để giải quyết những bài toán có ràng buộc cho cả hệ tuyến tính và phi tuyến. Tuy vậy, do là phương pháp dựa vào mô hình, nên việc phân tích, đánh giá chất lượng ổn định hay bền vững của hệ, nhất là hệ phi tuyến, khi mô hình không dừng (non-autonomous), chứa các thành phần bất định, là không hề đơn giản. Có thể nói cho tới tận ngày nay bài toán đánh giá tính ổn định của hệ MPC phi tuyến dạng này vẫn còn là bài toán mở. Lý do chính là vì ở MPC phi tuyến người ta phải tìm nghiệm toàn cục của bài toán tối ưu phi tuyến không lồi. Xuất phát từ lý do đó mà bài báo này đã đưa ra một đề xuất xây dựng bộ điều khiển MPC cho hệ phi tuyến đủ đảm bảo được tính ổn định của hệ. Đề xuất này dựa chủ yếu vào tư tưởng xấp xỉ tuyến tính từng đoạn theo cửa sổ dự báo (receding linearization) cho hệ phi tuyến rồi sử dụng quy hoạch toàn phương (QP) cũng ở dạng từng đoạn, thay vì trực tiếp tìm nghiệm bài toán tối ưu phi tuyến không lồi. Một ví dụ của bài báo đã xác nhận tính hiệu quả của giải pháp đề xuất trên.

**Từ khóa:** Điều khiển MPC; Quy hoạch toàn phương; Tính ổn định của MPC.

## ABSTRACT

Model Predictive Control (MPC) is a widely used control method in practice, providing reliable results. It originated from a real-world problem, under the name Model Algorithm Control (MAC), and gradually developed into a preferred tool for solving constrained problems in both linear and nonlinear systems. However, because it is a model-based method, analyzing and evaluating the stability or robustness of the system, especially nonlinear systems, when the model is non-autonomous and contains uncertain components, is not simple. It can be said that even today, the problem of evaluating the stability of this type of nonlinear MPC system remains an open problem. The main reason is that in nonlinear MPC, one must find a global solution to the non-convex nonlinear optimization problem. Based on this reason, this paper proposes a design for an MPC controller for nonlinear systems that ensures system stability. This proposal is primarily based on the idea of segmented linear approximation using the prediction window (receding linearization) for nonlinear systems, and then using quadratic programming (QP) also in segmented form, instead of directly finding the solution to the non-convex nonlinear optimization problem. An example in the paper confirms the effectiveness of the proposed solution.

**Keywords:** MPC control; Quadratic programming; MPC stability.

**Ngày nhận bài:** 16/01/2026; **Ngày sửa bài:** 11/03/2026; **Ngày duyệt đăng:** 02/04/2026

## 1. Mở đầu

Xuất phát từ một bài toán của công nghiệp, năm 1982, mà kỹ thuật MAC (Model Algorithm control) [1] được ra đời. Tiếp theo đó là DMC (dynamic matrix control) [2] và GPC năm 1987 (Generalized Predictive Control) [3]. Tất cả chúng, tuy rằng vẫn tồn tại những ưu nhược điểm riêng biệt [2], song đã tạo tiền đề cho việc hình thành kỹ thuật MPC (Model Predictive Control) [4],[5],[6] với vô vàn ứng dụng thành công trong thực tế. Ưu điểm nổi trội của kỹ thuật này so với các phương pháp điều khiển phi tuyến khác, như thích nghi, bền vững, trượt... là nó xử lý được các bài toán điều khiển có ràng buộc về trạng thái và tín hiệu điều khiển [7]. Sở dĩ nó làm được điều này là vì bản chất của MPC là từng bước chuyển bài toán điều khiển phi tuyến thông thường về thành bài toán quy hoạch phi tuyến có ràng buộc rồi sử dụng các phương pháp tối ưu hóa để xử lý điều kiện ràng buộc đó [7], cùng với vô vàn công cụ tối ưu hóa đã có sẵn hỗ trợ việc cài đặt chúng lên thiết bị điều khiển số [8]. Tuy nhiên, cũng vì đã chuyển bài toán điều khiển phi tuyến có ràng buộc thành bài toán quy hoạch phi tuyến, tức là bài toán tối ưu có hàm mục tiêu phi tuyến và tập ràng buộc cũng được mô tả bởi các bất phương trình phi tuyến, nên vấn đề khảo sát tính ổn định của hệ thu được gặp khá nhiều khó khăn [9], do còn thiếu những công cụ tìm nghiệm toàn cục cho bài toán quy hoạch phi tuyến không lồi [10].

Nhằm khắc phục nhược điểm trên, bài báo định hướng sẽ cùng với việc dịch cửa sổ dự báo của MPC để chuyển dần từng bước mô hình phi tuyến không dừng thành tuyến tính tham số hằng (hệ LTI) trong cửa sổ dự báo hiện tại. Nhờ đó, thay vì gặp phải bài toán quy hoạch

phi tuyến không lồi, ta sẽ có bài toán quy hoạch toàn phương (lồi), giúp ta luôn tìm được nghiệm toàn cục khi tập ràng buộc cũng là một tập lồi nhờ các phương pháp số giải bài toán QP (quadratic programming).

## 2. Nội dung chính

### 2.1. Phát biểu bài toán và các giả thiết kèm theo

Nội dung bài báo liên quan tới nhiệm vụ thiết kế bộ điều khiển bám quỹ đạo có ràng buộc, cho hệ phi tuyến không dừng, mô tả bởi mô hình rời rạc dạng affine

$$(1) \quad \begin{cases} \underline{x}_{k+1} = f(\underline{x}_k, k) + B(\underline{x}_k, k)\underline{u}_k \\ \underline{y}_k = C(\underline{x}_k, k)\underline{x}_k \end{cases}$$

trong đó  $\underline{x}_k \in \mathbb{R}^n$  là vector của  $n$  biến trạng thái ở thời điểm rời rạc  $k$ ,  $\underline{u}_k \in \mathbb{R}^m$  là vector của  $m$  các tín hiệu điều khiển ở cùng thời điểm  $k$ ,  $\underline{y}_k \in \mathbb{R}^l$  là vector của các tín hiệu đầu ra,  $B(\underline{x}_k, k)$  là ma trận điều khiển, có kiểu  $n \times m$  và  $C(\underline{x}_k, k)$  là ma trận đầu ra, có kiểu  $l \times n$ . Như vậy hệ (1) này có bậc  $n$  và MIMO (nhiều đầu vào–nhiều đầu ra) với  $m$  đầu vào và  $l$  đầu ra.

Bộ điều khiển phải thiết kế cần thỏa mãn:

Làm cho đầu ra  $\underline{y}_k$  bám theo được quỹ đạo mẫu  $\underline{r}_k$  có thể chỉ cần cho trước từng đoạn  $[k, k + N_r(k)]$  với  $N_r(k) \in \mathbb{N}$  cũng có thể thay đổi theo  $k$ , và thỏa mãn  $\underline{y}_k - \underline{r}_k \rightarrow \underline{0}_l$  khi  $k \rightarrow \infty$ , trong đó  $\underline{0}_l$  là ký hiệu của vector có tất cả các phần tử bằng 0, và tất nhiên cũng phải có số chiều tương ứng là  $l$ .

Vector các tín hiệu điều khiển  $\underline{u}_k$  luôn thuộc  $U$  đóng và giới nội (tập compact) cho trước.

Để bài toán ở trên có nghiệm, ta giả thiết:

$B(\underline{x}_k, k) \neq \mathbf{0}_{n \times m}, \forall \underline{x}_k, k$  để hệ là điều khiển được tại  $\underline{x}_k$  ở thời điểm rời rạc  $k$ ,

trong đó ký hiệu  $\mathbf{0}_{n \times m}$  là chỉ ma trận kiểu  $n \times m$  có tất cả các phần tử bằng 0.

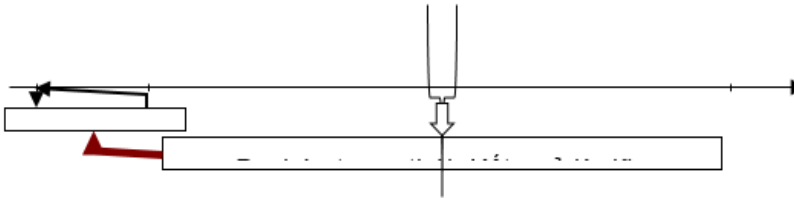
$C(\underline{x}_k, k) \neq \mathbf{0}_{l \times n}, \forall \underline{x}_k, k$  để tránh trường hợp hệ rơi vào chế độ động học không tại  $\underline{x}_k$  ở thời điểm rời rạc  $k$ .

Vector hàm  $f(\underline{x}_k, k)$  thỏa mãn

$$\frac{\|f(\underline{x}_k, k)\|}{\|\underline{x}_k\|} < \infty, \forall \underline{x}_k, k \quad (2)$$

tức là  $f(\underline{x}_k, k)$  biến đổi không nhanh hơn  $\underline{x}_k$  với mọi  $k$ . Ta có thể thấy tất cả những vector hàm  $f(\underline{x}_k, k)$  Lipschitz toàn cục đều thỏa mãn điều kiện (2) này.

**2.2. Phương pháp điều khiển MPC truyền thống**



*Hình 1: Nguyên tắc hoạt động của bộ điều khiển dự báo*

Một cách chi tiết thì nó gồm các bước sau:

Chọn trước các khoảng dự báo  $N_u$  cho tín hiệu điều khiển  $\underline{u}_k$  và  $N_x$  cho trạng thái  $\underline{x}_k$ , trong đó  $N_u \geq N_x$ . Thông thường người ta hay chọn  $N_u = N_x = N$  với  $N$  không được lớn hơn khoảng  $N_r(k)$  mà quỹ đạo mẫu  $\underline{r}_k$  sẽ được xác định trước từng đoạn.

Xây dựng mô hình dự báo cho tín hiệu đầu ra  $y_i, i = k+1, \dots, k+N$  tính từ thời điểm rời rạc  $k$  hiện tại. Thông thường người ta có thể sử dụng ngay mô hình rời rạc (1) của hệ để làm mô hình dự báo.

Thực hiện lần lượt cho  $k = 0, 1, \dots$  các bước sau:

Đo trạng thái  $\underline{x}_k$ .

Sử dụng mô hình dự báo để xác định các tín hiệu đầu ra tương lai  $y_i$  trong cửa sổ dự báo hiện tại  $i = k+1, \dots, k+N$

. Những tín hiệu đầu ra tương lai này sẽ là các vector hàm phụ thuộc  $\underline{u}_j, j = k, k+1, \dots, k+N-1$ . Để đơn giản ta sẽ sử dụng ký hiệu  $y_i(\underline{u}_k)$  với  $\underline{u}_k = \text{col}(\underline{u}_k, \underline{u}_{k+1}, \dots, \underline{u}_{k+N-1})$  để chỉ sự phụ thuộc đó. Lúc này  $\underline{u}_k$  sẽ là vector trong không gian  $mN$  chiều.

Chọn hàm mục tiêu đánh giá sai lệch bám

$$J_k(\underline{u}_k) = h_k(\underline{u}_k) + \sum_{j=1}^N g_k(y_{k+j}(\underline{u}_k), r_{k+j}) \quad (3)$$

trong đó  $h_k(\underline{u}_k)$  phải là hàm xác định dương và  $g_k(y_{k+j}(\underline{u}_k), r_{k+j})$  phải ít nhất là hàm bán xác định dương của các sai lệch

$\frac{y_{k+j}(\underline{u}_k) - r_{k+j}}{}$ . Thông thường người ta hay chọn hàm mục tiêu dạng toàn phương như sau:

$$J_k(\underline{u}_k) = \underline{u}_k^T \mathbf{R}_k \underline{u}_k + \sum_{j=1}^N \left[ \left( y_{k+j}(\underline{u}_k) - r_{k+j} \right)^T \mathbf{Q}_k \left( y_{k+j}(\underline{u}_k) - r_{k+j} \right) \right] \quad (4)$$

thay cho (3), với  $\mathbf{R}_k, \mathbf{Q}_k$  là các ma trận đối xứng xác định dương chọn trước. Ở đây ta đã sử dụng ký hiệu  $\mathbf{R}_k$  thay vì  $R_k$  như vẫn làm với ma trận có các phần tử là số thực, vì  $\mathbf{R}_k$  thực chất là ma trận khối (các phần tử của nó cũng là ma trận).

Như vậy, nếu ta sử dụng ký hiệu gộp (vector và ma trận)

$$\begin{aligned} \underline{e}_k &= \text{col}(e_{k+1}, \dots, e_{k+N}) \\ &= \text{col}(y_{k+1}(\underline{u}_k) - r_{k+1}, \dots, y_{k+N}(\underline{u}_k) - r_{k+N}) \\ \mathbf{Q}_k &= \text{diag}(\mathbf{Q}_k, \dots, \mathbf{Q}_k) = \begin{pmatrix} \mathbf{Q}_k & \dots & \mathbf{0}_{l \times l} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{l \times l} & \dots & \mathbf{Q}_k \end{pmatrix} \end{aligned} \quad (5)$$

trong đó  $\mathbf{0}_{l \times l}$  là ma trận kiểu  $l \times l$  có tất cả các phần tử bằng 0, thì hàm mục tiêu (4) sẽ trở thành dạng gần giống toàn phương chính tắc

$$J_k(\underline{\mathbf{u}}_k) = \underline{\mathbf{u}}_k^T \mathbf{R}_k \underline{\mathbf{u}}_k + \mathbf{e}_k^T \mathbf{Q}_k \mathbf{e}_k. \quad (6)$$

Ở đây ta nói rằng (6) gần giống toàn phương, chứ chưa phải toàn phương theo biến  $\underline{\mathbf{u}}_k$ , vì còn quan hệ  $y_{k+i}(\underline{\mathbf{u}}_k) - r_{k+i}$  của các phần tử trong  $\mathbf{e}_k$  không ở dạng tuyến tính theo  $\underline{\mathbf{u}}_k$ .

Tìm nghiệm bài toán tối ưu hóa có ràng buộc

$$\underline{\mathbf{u}}_k^* = \arg \min_{\underline{\mathbf{u}}_k \in U} J_k(\underline{\mathbf{u}}_k) \quad (7)$$

Đưa  $\underline{\mathbf{u}}_k = (I_m, \mathbf{0}_{m \times m}, \dots, \mathbf{0}_{m \times m}) \underline{\mathbf{u}}_k$  vào điều khiển hệ (1), trong đó  $I_m$  là ma trận đơn vị kiểu  $m \times m$ .

Như vậy, vòng lặp 3 với  $k=0,1, \dots$  chỉ kết thúc khi quá trình điều khiển kết thúc. Về phương pháp MPC truyền thống này ta có một vài ý kiến đánh giá chất lượng mang lại của nó như sau:

Khi hệ là phi tuyến dạng (1), để có được đầu ra  $y_{k+i}$  trong cửa sổ dự báo hiện tại, tức là với  $i=1, \dots, N$ , ta cần dự báo được trạng thái  $\underline{\mathbf{x}}_{k+i}$  và chúng là những vector hàm phụ thuộc  $\underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \dots, \underline{\mathbf{u}}_{k+N_y-1}$ . Sự phụ thuộc này mang tính phi tuyến khá cao. Cửa sổ dự báo  $N$  được chọn càng lớn, tính phi tuyến càng cao. Điều này ta có thể thấy thông qua phép khai triển vector trạng thái  $\underline{\mathbf{x}}_{k+i}$  trong của sổ dự báo hiện tại  $[k+1, k+N]$

$$\begin{aligned} \underline{\mathbf{x}}_{k+1} &= f(\underline{\mathbf{x}}_k, k) + B(\underline{\mathbf{x}}_k, k) \underline{\mathbf{u}}_k \\ \underline{\mathbf{x}}_{k+2} &= f(\underline{\mathbf{x}}_{k+1}, k+1) + B(\underline{\mathbf{x}}_{k+1}, k+1) \underline{\mathbf{u}}_{k+1} \\ &= f(f(\underline{\mathbf{x}}_k, k) + B(\underline{\mathbf{x}}_k, k) \underline{\mathbf{u}}_k, k+1) + B(f(\underline{\mathbf{x}}_k, k) + B(\underline{\mathbf{x}}_k, k) \underline{\mathbf{u}}_k, k+1) \underline{\mathbf{u}}_{k+1} \\ &\doteq \varphi_2(k, k+1, \underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_{k+1}) \\ \underline{\mathbf{x}}_{k+3} &= f(\underline{\mathbf{x}}_{k+2}, k+2) + B(\underline{\mathbf{x}}_{k+2}, k+2) \underline{\mathbf{u}}_{k+2} \\ &= f(\varphi_2(k, k+1, \underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_{k+1}), k+2) + \\ &\quad + B(\varphi_2(k, k+1, \underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_{k+1}), k+2) \underline{\mathbf{u}}_{k+2} \\ &\doteq \varphi_3(k, k+1, k+2, \underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_{k+1}, \underline{\mathbf{u}}_{k+2}) \\ &\vdots \\ \underline{\mathbf{x}}_{k+i} &\doteq \varphi_i(k, k+1, k+2, \dots, k+i-1, \underline{\mathbf{x}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_{k+1}, \underline{\mathbf{u}}_{k+2}, \dots, \underline{\mathbf{u}}_{k+i-1}) \\ &\vdots \end{aligned}$$

trong đó,  $\doteq$  chỉ phép gán, tức là ký hiệu của một quy ước. Tuy nhiên, từ dạng hàm mục tiêu cho ở công thức (4)

cũng dễ thấy rằng khi  $N$  được chọn càng nhỏ, khả năng bám tín hiệu đặt  $r_k$  của bộ điều khiển càng thấp.

Tính phi tuyến cao của kết quả dự báo  $y_{k+i}$  dẫn tới chất lượng rất xấu của hàm mục tiêu (3) hoặc (4), làm cho nó có tính lồi mạnh với nhiều điểm cực tiểu địa phương, trong khi tất cả các phương pháp tối ưu hóa hiện có đều không có khả năng xác định được ít nhất một điểm cực tiểu toàn cục của bài toán tối ưu lồi (non-convex).

Hiện nay, bài toán đánh giá chất lượng bám  $e_k = y_k - r_k$  khi  $k \rightarrow \infty$  của MPC phi tuyến vẫn đang mở (chưa có kết quả rõ ràng). Để cải thiện chất lượng bám một cách định tính, người ta thường sử dụng các giải pháp trung gian như:

Lựa chọn hợp lý các hàm  $h_k(\underline{\mathbf{u}}_k)$  và  $g_k(y_{k+j}, r_{k+j})$  cho hàm mục tiêu (3) hoặc bổ sung thêm hàm phạt cho  $J_k(\underline{\mathbf{u}}_k)$  để tạo được dãy  $\|e_k\|$  đơn điệu giảm. Tuy nhiên công việc này chưa tổng quát và chỉ mang tính đơn lẻ với từng hệ cụ thể. Ngoài ra, việc có được tính đơn điệu giảm của  $\|e_k\|$  cũng chưa đủ để khẳng định sẽ có  $\|e_k\| \rightarrow 0$ .

Xây dựng được một mô hình dự báo phi tuyến phù hợp, vừa đảm bảo là nó chứa đựng được đầy đủ đặc tính động học

của hệ (1), vừa có khả năng điều khiển ổn định tiệm cận được, ít nhất là với bộ điều khiển MPC. Hiện tại đây cũng là một bài toán mở và chưa có một gợi ý tổng quát nào cho việc xác định mô hình dự báo đó.

Khắc phục các nhược điểm cơ bản trên của MPC phi tuyến truyền thống cũng chính là mục đích của bài

báo này

## 2.2. Phương pháp điều khiển MPC đề xuất cho hệ phi tuyến (1)

Bản chất phương pháp đề xuất cho MPC phi tuyến (1) của bài báo được xuất phát từ kết quả điều khiển MPC đã có cho hệ tuyến tính tham số hằng (hệ LTI), rằng khi hàm mục tiêu có dạng toàn phương, điều kiện ràng buộc  $U$  là tập lồi, thì bài toán tối ưu hóa (7) sẽ là bài toán tối ưu lồi, nên nghiệm toàn cục của nó dễ dàng tìm được nhờ các phương pháp tối ưu hóa hiện có, cũng như khi hệ LTI có chứa  $m$  thành phần tích phân thì mọi bộ điều khiển làm nó ổn định cũng sẽ là nó bám theo được tín hiệu đặt khả vi  $m$  lần.

Xuất phát từ mô hình (1) thì cùng với giả thiết (2), nó sẽ viết lại được thành

$$\begin{cases} \underline{x}_{k+1} = A(\underline{x}_k, k)\underline{x}_k + B(\underline{x}_k, k)\underline{u}_k \\ \underline{y}_k = C(\underline{x}_k, k)\underline{x}_k \end{cases} \quad (8)$$

trong đó  $A(\underline{x}_k, k) = \frac{f(\underline{x}_k, k)}{\|\underline{x}_k\|^2} \underline{x}_k^T$

là ma trận kiểu  $n \times n$  có tất cả các phần tử là hàm bị chặn theo  $\underline{x}_k$  và  $k$ . Như vậy, tại thời điểm  $k$  hiện tại, và do  $\underline{x}_k$  là đã biết đo được, nên (8) trở thành tuyến tính không dừng

$$\begin{cases} \underline{x}_{k+1} = A_k \underline{x}_k + B_k \underline{u}_k \\ \underline{y}_k = C_k \underline{x}_k \end{cases} \quad (9)$$

với  $A_k = A(\underline{x}_k, k)$ ,  $B_k = B(\underline{x}_k, k)$ ,  $C_k = C(\underline{x}_k, k)$  (10)

là những ma trận rõ (certain). Nhược điểm duy nhất của mô hình (9) này, so với hệ LTI, là các ma trận tham số  $A_k$ ,  $B_k$ ,  $C_k$  phụ thuộc thời gian.

Ta sẽ sử dụng (9) để xây dựng mô hình dự báo với  $m$  khâu tích phân bổ sung. Ký hiệu  $v_k = \underline{u}_k - \underline{u}_{k-1}$  thì do tại thời điểm  $k$  hiện tại đã có  $\underline{u}_{k-1}$ , nên để tìm  $\underline{u}_k$  ta chỉ cần tìm  $v_k$ , tức là  $v_k$  bây giờ sẽ là vector tín hiệu điều khiển của mô hình dự báo. Với ký hiệu này, thì tại thời điểm  $k$  mô hình (9) trở thành

$$\begin{cases} \begin{pmatrix} \underline{x}_{k+1} \\ \underline{u}_k \end{pmatrix} = \begin{pmatrix} A_k & B_k \\ \mathbf{0}_{n \times n} & I_{m \times m} \end{pmatrix} \begin{pmatrix} \underline{x}_k \\ \underline{u}_{k-1} \end{pmatrix} + \begin{pmatrix} B_k \\ I_{m \times m} \end{pmatrix} v_k \\ \underline{y}_k = (C_k \quad \mathbf{0}_{l \times m}) \begin{pmatrix} \underline{x}_k \\ \underline{u}_{k-1} \end{pmatrix} \end{cases}$$

hay  $\begin{cases} \underline{z}_{k+1} = A_k \underline{z}_k + B_k v_k \\ \underline{y}_k = C_k \underline{z}_k \end{cases} \quad (11)$

trong đó

$$\underline{z}_k = \begin{pmatrix} \underline{x}_k \\ \underline{u}_{k-1} \end{pmatrix}, A_k = \begin{pmatrix} A_k & B_k \\ \mathbf{0}_{n \times n} & I_{m \times m} \end{pmatrix},$$

$$B_k = \begin{pmatrix} B_k \\ I_{m \times m} \end{pmatrix}, C_k = (C_k \quad \mathbf{0}_{l \times m}) \quad (12)$$

Rõ ràng, tại thời điểm  $k$  hiện tại, mô hình dự báo (11) trên mang đầy đủ đặc tính động học của hệ (1) ban đầu. Hơn nữa nó còn có  $m$  khâu tích phân trong đó, vì  $\det(\lambda I_{(n+m) \times (n+m)} - A_k) = \det(\lambda I_{n \times n} - A_k)(\lambda - 1)^m = 0$

có nghiệm  $\lambda = 1$  bội  $m$ .

Sử dụng mô hình dự báo (11) này, trong đó các ma trận  $A_k$ ,  $B_k$ ,  $C_k$  được giữ cố định trong toàn khoảng dự báo  $[k+1, k+N]$  hiện tại, tức là chúng không còn phụ thuộc  $k+i$  trong cửa sổ dự báo đó, ta được

$$\begin{aligned} \underline{z}_{k+1} &= A_k \underline{z}_k + B_k v_k \\ \underline{z}_{k+2} &= A_k (A_k \underline{z}_k + B_k v_k) + B_k v_{k+1} \\ &\vdots \\ &= A_k^2 \underline{z}_k + A_k B_k v_k + B_k v_{k+1} \\ \underline{z}_{k+i} &= A_k^i \underline{z}_k + A_k^{i-1} B_k v_k + \dots \\ &\quad + A_k B_k v_{k+i-2} + B_k v_{k+i-1} \\ &\vdots \\ \underline{z}_{k+N} &= A_k^N \underline{z}_k + A_k^{N-1} B_k v_k + \dots \\ &\quad + A_k B_k v_{k+N-2} + B_k v_{k+N-1} \end{aligned}$$

Vậy, từ (11) ta có vector tín hiệu đầu ra cho mọi  $i = 1, \dots, N$

$$\underline{y}_{k+i} = C_k A_k^i \underline{z}_k + C_k A_k^{i-1} B_k v_k + \dots + C_k A_k B_k v_{k+i-2} + C_k B_k v_{k+i-1} \quad (13)$$

Từ đây, khi viết chung lại (13) cho tất cả các chỉ số  $i = 1, 2, \dots, N$  được

$$\begin{pmatrix} \underline{y}_{k+1} \\ \vdots \\ \underline{y}_{k+N} \end{pmatrix} = \begin{pmatrix} C_k A_k \\ \vdots \\ C_k A_k^N \end{pmatrix} \underline{z}_k + \begin{pmatrix} C_k B_k & \mathbf{0}_{l \times m} & \dots & \mathbf{0}_{l \times m} \\ \vdots & \vdots & \ddots & \vdots \\ C_k A_k^{N-1} B_k & C_k A_k^{N-2} B_k & \dots & C_k B_k \end{pmatrix} \begin{pmatrix} v_k \\ \vdots \\ v_{k+N-1} \end{pmatrix}$$

Do đó, với các ký hiệu

$$\begin{aligned} \underline{y} &= \text{col}(y_{k+1}, \dots, y_{k+N}), \quad \underline{v} = \text{col}(v_k, \dots, v_{k+N-1}), \\ \underline{r} &= \text{col}(r_{k+1}, \dots, r_{k+N}), \\ Q &= \begin{pmatrix} Q_k & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & Q_k \end{pmatrix}, \quad R = \begin{pmatrix} R_k & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & R_k \end{pmatrix}, \quad D = \begin{pmatrix} C_k A_k \\ \vdots \\ C_k A_k^N \end{pmatrix}, \\ F &= \begin{pmatrix} C_k B_k & \mathbf{0}_{l \times m} & \dots & \mathbf{0}_{l \times m} \\ \vdots & \vdots & \ddots & \vdots \\ C_k A_k^{N-1} B_k & C_k A_k^{N-2} B_k & \dots & C_k B_k \end{pmatrix} \end{aligned} \quad (14)$$

thì hàm mục tiêu (4) trở thành

$$J_k(\underline{v}) = \underline{v}^T (F^T Q F + R) \underline{v} - 2(\underline{r} - D \underline{z}_k)^T Q F \underline{v} \quad (15)$$

và là hàm toàn phương. Từ đó, nếu ràng buộc  $U$  còn là tập lồi, điều thường gặp trong thực tế khi các tín hiệu điều khiển luôn thuộc một khoảng đóng xác định cho trước, chẳng hạn như

$$\underline{\alpha}_k \leq \underline{u}_k \leq \underline{\beta}_k$$

với  $\underline{\alpha}_k, \underline{\beta}_k$  là hai vector và phép so sánh vector được thực hiện cho các phần tử của chúng, thì bài toán tối ưu phi tuyến (7) bây giờ trở thành bài toán tối ưu lồi, có hàm mục tiêu là toàn phương, tập ràng buộc được mô tả bởi các bất đẳng thức tuyến tính (bài toán QP)

$$\underline{v}^* = \arg \min_{\underline{v} \in V_k} J_k(\underline{v}) \quad (16)$$

trong đó tập ràng buộc  $V_k$  được sinh ra từ  $U$  bằng cách dịch nó một khoảng  $-\underline{u}_{k-1}$  trong không gian điều khiển  $\mathbb{R}^m$ , cũng là một tập lồi, mô tả bởi các bất đẳng thức tuyến tính. Cuối cùng, ta suy ra được từ  $\underline{v}^*$

$$\underline{u}_k = \underline{u}_{k-1} + (I_m, \mathbf{0}_{m \times m}, \dots, \mathbf{0}_{m \times m}) \underline{v}^* \quad (17)$$

và đó là tín hiệu điều khiển cần tìm ở thời điểm  $k$  hiện tại (hình 1).

### 2.3. Thuật toán điều khiển

Để tiện cho việc cài đặt phương pháp đề xuất, sau đây ta sẽ đúc kết lại nội dung của nó đã trình bày ở trên dưới dạng thuật toán gồm các bước sau:

Khởi tạo: Chọn trước khoảng dự báo  $N$ . Khoảng dự báo được chọn này không

được nhỏ hơn khoảng biết trước của tín hiệu đặt  $r_k$ , nếu nó cũng chỉ được cho trước từng đoạn. Ở trường hợp đã có  $r_k$  với mọi  $k \geq 0$  thì  $N \geq 2$  là tùy chọn.

Gán  $k = 0$ . Tùy gán  $\underline{u}_{-1}$ .

Thực hiện với  $k = 0, 1, \dots$  các bước sau:

Đo trạng thái  $\underline{x}_k$ . Xác định  $\underline{z}_k, A_k, B_k, C_k$  theo (12).

Chọn hai ma trận đối xứng xác định dương  $R_k, Q_k$  rồi thiết lập  $R, Q$  theo (14). Trong trường hợp chúng được chọn không phụ thuộc  $k$  thì bước này sẽ được chuyển vào phần khởi tạo.

Xây dựng các ma trận  $D, F$  và vector  $\underline{r}$  theo (14). Xác định tập ràng buộc  $V_k$  từ  $U$  và  $\underline{u}_{k-1}$ .

Tìm nghiệm tối ưu  $\underline{v}^*$  của bài toán tối ưu toàn phương (16) với ràng buộc lồi, trong đó  $J_k(\underline{v})$  cho tại (15). Có thể thấy, nếu

$$\underline{v}^* = (F^T Q F + R)^{-1} F^T Q (\underline{r} - D \underline{z}_k) \in V_k \quad (18)$$

thì nó là nghiệm toàn cục của bài toán tối ưu (16), ngược lại  $\underline{v}^*$  chỉ có thể nằm trên biên của tập lồi  $V_k$  và tìm được nhờ thuật toán truyền thống của Karush-Kuhn-Tucker (KKT) [11]. Thuật toán này cũng đã được tích hợp trong MatLab dưới tên lệnh `fmincon` rất tiện sử dụng [12],[13].

Tính  $\underline{u}_k$  theo (17) và đưa vào điều khiển hệ (1).

Có thể thấy là hoàn toàn không khó khăn gì khi ta mở rộng thuật toán trên cho cả những hệ (1) affine nhưng có dạng tổng quát hơn ở đầu ra như sau:

$$\begin{cases} \underline{x}_{k+1} = f(\underline{x}_k, k) + B(\underline{x}_k, k) \underline{u}_k \\ \underline{y}_k = g(\underline{x}_k, k) \end{cases} \quad (19)$$

bằng cách đưa nó về dạng (1) với

$$C(\underline{x}_k, k) = \frac{g(\underline{x}_k, k)}{\|\underline{x}_k\|^2} \underline{x}_k^T.$$

### 3. Minh họa bằng ví dụ số

Để minh họa chất lượng của bộ điều khiển đề xuất, có so sánh với bộ điều

khuyến MPC phi tuyến truyền thống, sau đây ta sẽ cài đặt cả hai phương pháp để đưa đầu ra  $y_k$  của hệ

$$\underline{x}_{k+1} = \begin{pmatrix} x_{k_2} - x_{k_3} + u_k \cos x_{k_1} \\ 2x_{k_3} - x_{k_2} \\ x_{k_1} \sin x_{k_2} + x_{k_3} + u_k \end{pmatrix} \quad (20)$$

$$y_k = |x_{k_1} - x_{k_2}| x_{k_1} + 2x_{k_2}$$

với vector trạng thái  $\underline{x}_k = (x_{k_1}, x_{k_2}, x_{k_3})^T$ , tín hiệu đầu vào (tín hiệu điều khiển)  $u_k$ , bám theo được tín hiệu đặt  $r_k = 2, \forall k$  và  $u_k$  thỏa mãn điều kiện ràng buộc dạng khoảng  $|u_k| \leq 0.2$ . Hàm mục tiêu (3) tương ứng với chất lượng bám đặt ra này là

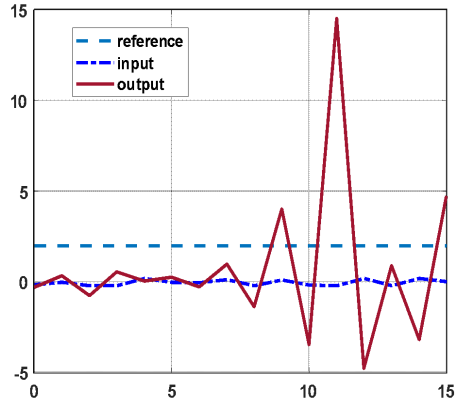
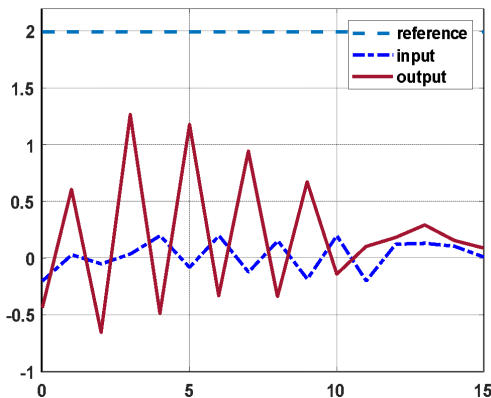
$$J_k(\underline{u}_k) = \|\underline{u}_k\|^2 + q \sum_{j=1}^N (y_{k+j}(\underline{u}_k) - r_{k+j})^2$$

trong đó  $N \geq 1$  là độ rộng của sổ dự báo tùy chọn,

$\underline{u}_k = (u_k, u_{k+1}, \dots, u_{k+N-1})^T$  và hệ số  $q = 5$  cho trước.

Trước tiên ta cài đặt thuật toán truyền thống như mô tả ở mục 2.2 thành chương trình pxm1.m viết trên MatLab, có mã nguồn như dưới đây, sẽ thu được kết quả như ở hình 2, cho hai trường hợp  $N = 10$  và  $N = 20$  khác nhau. Có thể thấy ở cả hai trường hợp này đầu ra của hệ không bám theo được tín hiệu đặt cho trước. Thậm chí là khi  $N$  được chọn càng lớn, chất lượng càng xấu.:

- a) Khi  $N = 10$
- b) Khi  $N = 20$



**Hình 2:** Kết quả điều khiển với phương pháp truyền thống

```
% pxm1.m: Conventional method
global N x r q
ncon,'algorithm','sqp');
M=16; N=10; w=2; r=w*ones(N,1);
q=5*ones(N,1);
ub=0.2*ones(N,1); lb=-ub; u=0;
x=[0;0;0]; up=[]; yp=[]; t=[];
for i=1:M
    t=[t;i-1]; p0=u*ones(N,1);
    p=fmincon(@
myobj,p0,[],[],[],[],lb,ub,[],op);
    u=p(1); [x,y]=mysys(u,x);
    up=[up;u]; yp=[yp;y]; % for printing
end
plot(t,w*ones(M,1),t,up,t,yp);
legend('reference','input','output');
function J=myobj(u) % objective
function
global N x r q
J=0; xk=x;
for i=1:N
    [xk,yk]=mysys(u(i),xk);
    J=J+q(i)*(r(i)-yk)^2+u(i)^2;
end
end
function [x,y]=mysys(u,x) % system
dynamic
x=[x(2)-x(3)+u*cos(x(1));2*x(3)-
x(2);x(1)*sin(x(2))+x(3)+u];
y=abs(x(1)-x(2))*x(1)+2*x(1);
end
```

Tiếp theo, ta sẽ sử dụng phương pháp đề xuất mà ở đó hệ (20) đã được viết lại thành dạng tương đương

$$\underline{x}_{k+1} = \begin{pmatrix} x_{k_2} - x_{k_3} + u_k \cos x_{k_1} \\ 2x_{k_3} - x_{k_2} \\ x_{k_1} \sin x_{k_2} + x_{k_3} + u_k \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 2 \\ \sin x_{k_2} & 0 & 1 \end{pmatrix} \underline{x}_k + \begin{pmatrix} \cos x_{k_1} \\ 0 \\ 1 \end{pmatrix} u_k$$

$$y_k = \begin{vmatrix} x_{k_1} - x_{k_2} \\ x_{k_1} + 2x_{k_2} \end{vmatrix} = \begin{pmatrix} x_{k_1} - x_{k_2} \\ 2, 0 \end{pmatrix} \underline{x}_k$$

tức là nó có

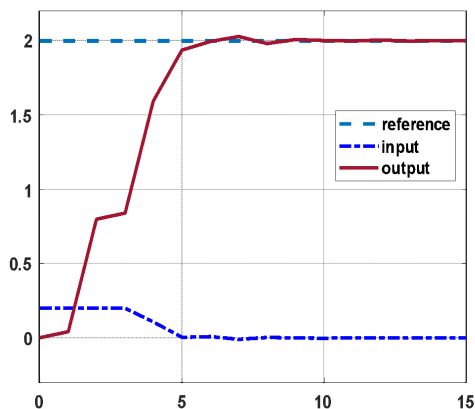
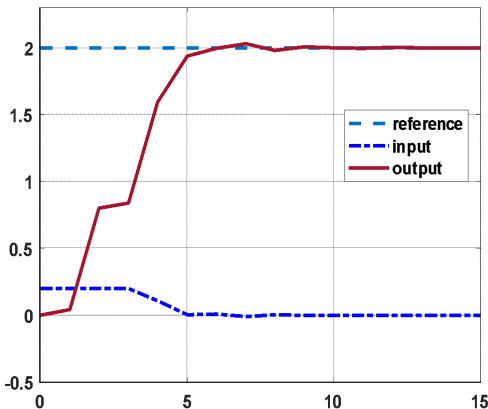
$$A_k = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 2 \\ \sin x_{k_2} & 0 & 1 \end{pmatrix}, B_k = \begin{pmatrix} \cos x_{k_1} \\ 0 \\ 1 \end{pmatrix},$$

$$C_k = \begin{pmatrix} x_{k_1} - x_{k_2} \\ 2, 0 \end{pmatrix}$$

Cài đặt thuật toán đề xuất đã giới thiệu ở mục 2.3 thành chương trình pxm2.m viết trên MatLab có mã nguồn chi tiết như ở dưới, ta thu được kết quả cho ở hình 3. Kết quả này cho thấy bộ điều khiển đề xuất đã đáp ứng được yêu cầu chất lượng bám đặt ra cho cả hai trường hợp  $N = 10$  và  $N = 20$ . Hơn nữa nó còn có tốc tính toán nhanh hơn nhiều so với phương pháp truyền thống do chỉ cần tìm nghiệm của bài toán QP thay vì của bài toán quy hoạch phi tuyến.

a) Khi  $N = 10$

b) Khi  $N = 20$



end

Hình 3: Kết quả điều khiển với phương pháp đề xuất

```
% pxm2.m: Proposed algorithm
N=10; w=2; Q=5*eye(N); R=eye(N);
x=[0;0;0]; u=[];y=[]; a=0.2;
op=optimoptions(@
fmincon,'algorithm','sqp');
t = [ ] ;
r=w*ones(N,1);u0=0;
M=16; O=[1 zeros(1,N-1)];
v=0; %O=[1,0,...,0]
for k=1:M
    A=[0 1 -1;0 -1 2;sin(x(2)) 0
1];B=[cos(x(1));0;1];
    C=[abs(x(1)-x(2)) 2 0];
    E=[];F=zeros(N,N);
    Ah=[A B;0 0 0 1]; Bh=[B;1]; Ch=[C
0];
    for i=1:N
        E=[E;Ch*Ah^i];
        for j=1:i F(i,j)=Ch*Ah^(i-j)*Bh;
    end
    end
    t(k)=k-1; y(k)=C*x; d=E*[x;u0];
    G=F'*Q*F+R; b=F'*Q*(r-d);
    lb=-inf*ones(N,1); ub=-lb; lb(1)=
a-u0; ub(1)=a-u0;
    p0=v*ones(N,1);
    p=fmincon(@(p) p'*G*p-2*b'*p,p
0,[],[],[],lb,ub,[],op);
    du=O*p; u(k)=u0+du;
    x=A*x+B*u(k); u0=u(k);
end
```

```
plot(t,w*ones(length(t),1),t,u,t,y);  
legend('reference','input','output');
```

Một nhận xét thêm ở phương pháp đề xuất này là mặc dù khi  $N$  được chọn càng lớn, sai lệch giữa mô hình gốc (1) ban đầu với mô hình tuyến tính (9) càng cao do ở đó, trong cửa sổ dự báo hiện tại, ta đã giữ nguyên các ma trận tham số  $A_k, B_k, C_k$ , tức là đã gán  $A_k = A_{k+i}, B_k = B_{k+i}, C_k = C_{k+i}$  với mọi  $0 \leq i \leq N-1$ , song ảnh hưởng của sai lệch mô hình này tới chất lượng điều khiển là không đáng kể, vì kết quả  $u_k$  thu được cũng chỉ là phần tử đầu tiên trong dãy

$\mathbf{u}_k = (u_k, u_{k+1}, \dots, u_{k+N-1})$  của cửa sổ dự báo hiện tại.

#### 4. Kết luận

Chất lượng bám tín hiệu đặt theo nghĩa  $\|e_k\| \rightarrow 0$  của phương pháp đề xuất đã được kiểm chứng bằng ví dụ mô phỏng. Chất lượng này có được là do ta đã sử dụng sự dịch chuyển của cửa sổ dự báo để chuyển từng bước bài toán quy hoạch phi tuyến không lồi thành bài toán quy hoạch toàn phương có ràng buộc là tập lồi, rồi tìm nghiệm toàn cục nhờ phương pháp QP trong cửa sổ dự báo đó, cũng như đã bổ sung thêm được  $m$  khâu tích phân vào mô hình dự báo. Tuy nhiên còn thiếu lời chứng minh lý thuyết một cách chặt chẽ cho chất lượng thu được này và nó sẽ là nội dung nghiên cứu tiếp theo của tác giả.

---

#### Tài liệu tham khảo

1. Ramine Rouhani, Raman K. Mehra (1982). *Model algorithmic control (MAC): Basic theoretical properties*. Automatica Volume 18, Issue 4, 401-414.
2. P. Lundström; J.H. Lee; M. Morari; S. Skogestad (1995). *Limitations of*

*dynamic matrix control*. Computers & Chemical Engineering, Volume 19, Issue 4, 409-421

3. D.W. Clarke; C. Mohtadi; P.S. Tuffs (1987). *Generalized predictive control. The basic algorithm*. Automatica Volume 23, Issue 2, 137-148

4. Rawlings JB (2000). *Tutorial overview of model predictive control*. IEEE Control Sys 20(3):38-52

5. Camacho; Bordons (2004). *Model predictive control*. Springer Verlag

6. Morari Manfred (2025). *Model Predictive Control: The Genesis of an Idea - Histories of Control*. IEEE Control Systems. 45 (4): 86-88

7. Nguyễn Doãn Phước (2016). *Tối ưu hóa và điều khiển tối ưu*, Nhà xuất bản Bách khoa Hà Nội.

8. Rawlings James B.; Mayne David Q. and Diehl, Moritz M. (2017). *Model Predictive Control: Theory, Computation, and Design (2nd Ed.)*. Nob Hill Publishing

9. Mayne David Q.; Rawlings James B.; Rao Christopher V.; Scokaert Pierre O. M. (2000). *Constrained model predictive control: Stability and Optimality*. Automatica. 36 (6): pp. 789-814.

10. A. L. Peressini; F. E. Sullivan and J. J. Uhl, Jr. (1988). *The Mathematics of Nonlinear Programming*. Springer-Verlag, New York.

11. Nocedal, J. and Wright, S.J. (1996). *Numerical Optimization*. Springer-New York.

12. S. J. Chapman (2004). *MATLAB Programming for Engineers*. Thomson, 2004

13. A. Gilat (2004). *MATLAB: An introduction with Applications*. John Wiley and Sons.

# ROBUST STABILITY ANALYSIS AND REAL-TIME SPEED TRACKING CONTROL OF STEPPER MOTORS VIA MATLAB/OPC-PLC INTEGRATION WITH SMITH PREDICTOR

Nguyen Quang Binh<sup>1</sup>, Hoang Van Quyet<sup>1</sup>, Le Van Tuan<sup>1\*</sup>

<sup>1</sup> Control, Automation in Production and Improvement of Technology Institute (CAPITI), 89B Ly Nam De, Hoan Kiem, Hanoi, Vietnam

\*Email: visaosang89@gmail.com

## ABSTRACT

*This paper presents a robust real-time speed tracking control approach for a two-phase stepper motor based on MATLAB/Simulink–OPC–PLC integration with Smith predictor delay compensation. In OPC-based industrial systems, distributed communication and computation processes introduce significant time delays that may degrade closed-loop stability and tracking performance. An equivalent-delay modeling framework is established to represent MATLAB computation, OPC communication, PLC execution, and sensor feedback delays as a unified pure delay. The experimentally measured total delay is approximately 0.40s. A stability-oriented analysis of the Smith predictor is conducted using small-gain conditions and phase-margin-based delay tolerance criteria. The nominal delay-free system exhibits a gain margin of 9.6dB and a phase margin of 52.3°, with a complementary sensitivity peak, ensuring robust stability under 10–15% model uncertainty and practical delay variations. Semi-real-time experiments were performed under three feedback configurations and repeated five times for consistency verification. Results show that the Smith predictor significantly improves tracking performance compared to conventional PID control. The model-based feedback configuration achieves the best performance, with a settling time of 1.5s, zero overshoot, and a steady-state error of 1.83%. The proposed approach provides a practical and computationally efficient solution for delay compensation in PLC-based industrial control systems.*

**Keywords:** Robust stability, Smith predictor, distributed time delay, OPC–PLC integration, real-time industrial control.

Ngày nhận bài: 06/02/2026; Ngày sửa bài: 12/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. INTRODUCTION

Programmable Logic Controllers (PLCs) are widely used in industrial process control due to their robustness, flexibility, and suitability for harsh operating environments. PLC-based systems offer reliable solutions for controlling a wide range of electromechanical devices, including electric drives, conveyors, and robotic subsystems. However, most industrial PLCs are typically equipped with basic control structures such as on-off control or conventional PID controllers, which may be insufficient for achieving high-performance control in systems affected by communication and computation delays.

To overcome these limitations, MATLAB/Simulink is often employed as a high-level design and analysis platform, enabling the implementation of advanced control algorithms while PLCs handle low-level execution. The integration of MATLAB/Simulink with PLCs is commonly realized through OPC (Object Linking and Embedding for Process Control) servers, which provide a standardized and flexible communication framework. KEPServerEX, in particular, supports OPC UA and is widely adopted for secure and scalable data exchange between industrial devices and software applications.

Several studies have explored OPC-based control architectures in different application domains. In [1], OPC servers and image processing techniques were used for fruit classification on conveyor systems. The authors of [2] demonstrated PLC-KEPServerEX-MATLAB integration for speed control of DC motors, while [3] applied OPC-based

adaptive fuzzy PID control to temperature regulation problems. Educational-oriented OPC-PLC control platforms were reported in [4], and PID model prediction for PLC implementation was investigated in [5]. These studies confirm the feasibility and practicality of OPC-based control systems for industrial and educational purposes.

Nevertheless, a common limitation of the above works is that time delays inherent in OPC-PLC-based systems are not explicitly modeled or rigorously analyzed. In practice, delays arise from multiple sources, including data transmission through OPC servers, computation time in PLCs and MATLAB/Simulink, sensor signal processing, and actuator response. The combined effect of these delays forms a distributed delay structure, which can significantly degrade control performance and may even destabilize the closed-loop system if not properly addressed.

The Smith predictor is a classical and effective approach for compensating pure time delays in control systems. Its application has been reported in various tracking and industrial control problems, including systems with communication and image-processing delays [10]. However, most existing implementations apply Smith compensation in a heuristic manner, without providing formal guarantees on stability and robustness when model mismatch or delay uncertainty is present. This limitation becomes particularly critical in OPC-PLC environments, where delay characteristics may vary depending on network load, sampling periods, and computational burden.

Motivated by these observations, this paper focuses on robust real-time speed tracking control of stepper motors in OPC–PLC-based systems. Unlike servo motors, stepper motors are widely used in industrial automation due to their simplicity, cost-effectiveness, and precise positioning capability, but they are also sensitive to delay-induced oscillations and performance degradation. Therefore, ensuring stability and robustness in delayed control loops is of practical importance. The main contributions of this paper can be summarized as follows:

1. An equivalent-delay modeling framework is established for MATLAB/Simulink–OPC–PLC integrated control systems, explicitly capturing distributed communication, computation, execution, and sensing delays and representing them as a unified pure-delay element suitable for predictor design.

2. A stability-oriented analysis of the Smith predictor is developed for the considered industrial architecture. Small-gain robustness conditions and phase-margin-based delay tolerance criteria are derived and interpreted in the context of OPC-based PLC control loops.

3. A comparative semi-real-time experimental validation is conducted under three different feedback configurations. The results reveal that model-based feedback within the Smith structure significantly enhances robustness and tracking performance in PLC implementations, providing practical insight for industrial deployment and engineering education laboratories.

The remainder of this paper is organized as follows. Section 2 describes the structure of the OPC–PLC-based

embedded control system and introduces the Smith predictor. Section 3 presents the stability and robustness analysis of the Smith delay compensation scheme. Section 4 discusses the identification of the system model and evaluates the control performance through semi-real-time experiments. Finally, Section 5 concludes the paper and outlines directions for future research.

## **2. STRUCTURE OF EMBEDDED CONTROL SYSTEM USING OPC-PLC**

### *2.1. Embedded system control structure using OPC-PLC*

The system works based on OPC-PLC with the structure of a 4-layer SCADA system [11], shown in Figure 1. The monitoring layer acts as a central control station. Parameters and control signals from algorithms are displayed, and data storage is managed by SCADA software. Here, Simulink is used to design the interface, establish OPC connections, and output control signals to PLC. The core of the OPC Server layer is formed based on the OPC Server platform, serving to extract data on the OPC bus. This study uses OPC Server KEPServer Ex. The connection between PC and OPC Server is performed through the process of linking and embedding OPC objects. At the embedded control layer, the control program based on algorithms is designed and executed on PLC. The connection between PLC and OPC Server is via Modbus RTU industrial communication. The field layer is a microstep controller that receives control pulses from PLC to control the stepper motors. The rotary encoder sensor feeds back the measured speed to the PLC.

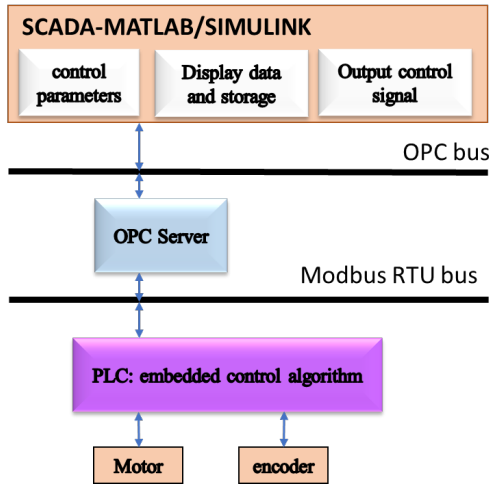


Figure 1. 4-layer embedded diagram configuration of the control system.

The control structure in Figure 1 shows the presence of delays in several components, including communication delays for reading/writing data via OPC server, calculation delays of the controller (PLC, driver), communication delays between the sensor (encoder) and the controller, delays between the controller and the actuator (motor), and calculation delays on Simulink. The paper considers the combined delay in the form of pure delay, which can be determined experimentally.

2.2. Smith Predictor

The Smith predictor proposed by O.J.M. Smith works by using a predictive model of the control object to estimate and predict the future of the system. Instead of reacting directly to the output, the system uses information from the predictive model to adjust the output. This helps to minimize the effect of delay and improve the overall performance of the control system. The Smith predictor

uses the predictive model of the delayed object to restructure the control loop, thereby eliminating the effect of pure delay.

Consider the control object with a transfer function , where is the gain, is the time constant, and is the delay.

$$G(s) = \frac{K}{Ts + 1} e^{-\tau s} \tag{1}$$

The Smith delay compensation structure is of the form (2), and the Smith predictor is represented in Figure 2 [6].

$$Y(s) = \frac{\hat{G}(s)}{1 + \hat{G}(s)C(s)} R(s) \tag{2}$$

Where, is the controller, is the desired signal, is the output, is the predictive model. When the model is asymptote , the system will almost eliminate the effect of the delay element.

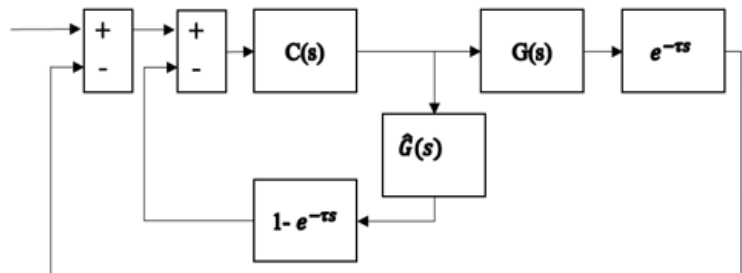


Figure 2. Structural diagram depicting the Smith predictor.

It is observed that the Smith controller contains an object model in the feedback circuit, so it is necessary to find the transfer function (without delay) of the system relatively accurately. Delay makes the system less stable, so the analysis of the stability of the system with delay is a very important task in control systems.

3. STABILITY ANALYSIS OF SMITH DELAY COMPENSATION SYSTEM

3.1. Assumption

A1: Consider the controller C(s)

as a linear controller (e.g., P, PI, PID) designed so that the pair  $[C, \hat{G}]$  is stable (i.e., the nominal closed-loop system with the model  $\hat{G}$  is stable).

A2:  $T_0(s)$  is complementary sensitivity of the nominal system (no delay in Smith's internal circuit).

$$T_0(s) = \frac{C(s)\hat{G}(s)}{1 + C(s)\hat{G}(s)}$$

A3: Every  $\| * \|_\infty$  standard is an infinite frequency standard:  $\|F\|_\infty = |F(j\omega)|$ .

### 3.2. Mathematical basis of stability analysis

**Theorem 1** (On equivalent stability when the model is correct): *If the model is completely correct, meaning  $G(s) = \hat{G}(s)e^{-\tau s}$ , then the actual closed-loop system using the Smith predictor has the same characteristic polynomial as the system  $[C(s), \hat{G}(s)]$  without delay, specifically, the poles are the solutions of the characteristic equation  $1 + C(s)\hat{G}(s) = 0$ . The internal stability of the Smith closed-loop system is equivalent to the stability of the pair  $[C, \hat{G}]$  without delay.*

**Proof.** According to assumption A1, the control signal  $U(s)$  generated by controller  $C(s)$  ( $P$  controller) from the error  $E(s)$ :

$$U(s) = C(s)E(s) \quad (3)$$

We have an actual object:

$$Y(s) = G(s)U(s) = \hat{G}(s)e^{-\tau s}U(s) \quad (4)$$

The model in the predictor generates two quantities: Model output without delay (5) and model output with delay (6)

$$\tilde{Y}(s) = \hat{G}(s)U(s) \quad (5)$$

$$\hat{Y}(s) = \hat{G}(s)e^{-\tau s}U(s) \quad (6)$$

The Smith predictor determines the current estimate (predicted current output):

$$Y_{est}(s) = \tilde{Y}(s) + [Y(s) - \hat{Y}(s)] \quad (7)$$

When the model is correct:

$$Y(s) = \hat{G}(s)e^{-\tau s}U(s) = \hat{Y}(s) \quad (8)$$

Therefore:

$$Y(s) - \hat{Y}(s) = 0; Y_{est}(s) = \tilde{Y}(s) = \hat{G}(s)U(s) \quad (9)$$

The error given to the controller is:

$$E(s) = R(s) - Y_{est}(s) = R(s) - \hat{G}(s)U(s) \quad (10)$$

So:

$$U(s) = C(s) (R(s) - \hat{G}(s)U(s)) \\ \Rightarrow [1 + C(s)\hat{G}(s)]U(s) = C(s)R(s) \quad (11)$$

From there:

$$\frac{Y(s)}{R(s)} = \frac{\hat{G}(s)e^{-\tau s}C(s)}{1 + C(s)\hat{G}(s)} \quad (12)$$

The poles of the system are the solutions of the equation  $1 + C(s)\hat{G}(s) = 0$  (the numerator only adds the factor  $e^{-\tau s}$  without changing the characteristic polynomial).

Theorem 1 has been proven.

✘

**Theorem 2** (On the stability under factored errors, small-gain condition): *Suppose the model error  $G(s) = \hat{G}(s)(1 + W_m(s)\Delta(s))e^{-\tau s}$  with acceptable uncertainty  $\Delta(s)$  and  $\|\Delta\|_\infty \leq 1$ . If  $\|W_m T_0\|_\infty < 1$ , If then the Smith closed-loop system is stable against all  $\Delta$  with  $\|\Delta\|_\infty \leq 1$ .*

**Proof.** From the general Smith expression (similar to Theorem 1), but with a factorization error, we have:

$$Y(s) = \hat{G}(s)(1 + W_m(s)\Delta(s))e^{-\tau s}U(s) \quad (13)$$

By the predictor formula:

$$Y_{est} = \hat{G}U + (Y - \hat{G}e^{-\tau s}U) = \hat{G}U + \hat{G}e^{-\tau s}W_m\Delta U \quad (14)$$

We have the error:

$$E = R - Y_{est} = R - \hat{G}U - \hat{G}e^{-\tau s}W_m\Delta U \quad (15)$$

Therefore:

$$U = CE = C(R - \hat{G}U - \hat{G}e^{-\tau s}W_m\Delta U) \quad (16)$$

$$(1 + C\hat{G})U + C\hat{G}e^{-\tau s}W_m\Delta U = CR \quad (17)$$

Multiply both sides of (17) by  $(1 + C\hat{G})^{-1}$  (It is valid because of the assumption A1):

$$U + \frac{C\hat{G}}{1 + C\hat{G}}e^{-\tau s}W_m\Delta U = \frac{C}{1 + C\hat{G}}R \quad (18)$$

According to assumption A2, we have the expression:

$$(I + T_0e^{-\tau s}W_m\Delta)U = U_0; \quad U_0 := \frac{C}{1 + C\hat{G}}R \quad (19)$$

Thus, the condition for  $U$  (and the whole system) to be well-defined and stable for all  $\Delta$  is that the matrix or linear combination  $I + T_0e^{-\tau s}W_m\Delta$  is invertible for all  $\Delta$  in the norm set. By the small-gain theorem for linear feedback interfaces, a sufficient condition for all  $\Delta$  with  $\|\Delta\|_\infty \leq 1$  not to lose invertibility is:

$$\|T_0e^{-\tau s}W_m\|_\infty < 1 \quad (20)$$

Since  $e^{-\tau s}$  has amplitude on the imaginary axis ( $|e^{-j\omega\tau}| = 1$ ). This is equivalent to:

$$\|T_0W_m\|_\infty < 1 \quad (21)$$

The above condition is sufficient for stability for the given factor error.

Theorem 2 has been proven.

✘

**Remark 1:** The above condition is a standard small-gain condition widely used in robust control. It can be directly checked by frequency plotting (we can draw  $|W_mT_0|$  and check  $|W_mT_0| < 1$  for all  $\omega$ ), and is directly related to criteria such as peak  $|T_0|$  around the bandwidth (if the peak is small, the error tolerance is larger).

**Theorem 3** (On robustness to delay error - phase/magnitude condition): Suppose the real object has a delay different from the delay  $\Delta\tau$  used in

the predictor, i.e.,  $G(s) = \hat{G}(s)e^{-(\tau+\Delta\tau)s}$ . Let  $\delta(j\omega) = e^{-j\omega\Delta\tau} - 1$ . If  $|\delta(j\omega)| < 1$ , If then the Smith closed-loop system is stable to delay error  $\Delta\tau$ . Since  $|\delta(j\omega)| = |e^{-j\omega\Delta\tau} - 1| = 2\sin\left(\frac{\omega\Delta\tau}{2}\right) \leq \{2, \omega|\Delta\tau|\}$  a simple and conservative condition is that  $|T_0(j\omega)| * \{2, \omega|\Delta\tau|\} < 1$ .

In addition, a practical approximation at the cutoff bandwidth  $\omega_c$  gives us the interval condition:  $\omega_c|\Delta\tau| < \varphi_m$  where is the phase margin of the pair  $[C, \hat{G}]$ .

**Proof.** We arrange:

$$G(s) = \hat{G}(s)e^{-\tau s}e^{-\Delta\tau s} = \hat{G}(s)e^{-\tau s}(1 + (e^{-\Delta\tau s} - 1)) \quad (22)$$

Set:  $\delta(s) = e^{-\Delta\tau s} - 1$

Returning to Smith's equation (similar to that in Theorem 2), we get, after similar arrangements:

$$[I + T_0e^{-\tau s}\delta(s)]U = U_0 \quad (23)$$

Since  $e^{-\tau s}$  has units of amplitude, the small-gain condition for  $I + T_0e^{-\tau s}\delta$  to be invertible for all frequencies is.

$$\|T_0\delta\|_\infty < 1 \quad (24)$$

It means:

$$|T_0(j\omega)| * \delta(j\omega) < 1 \quad (25)$$

Using trigonometric formula:

$$|e^{-j\omega\Delta\tau} - 1| = 2 \left| \sin\left(\frac{\omega\Delta\tau}{2}\right) \right|$$

and the inequality  $|\sin \sin x| \leq \min\{1, |x|\}$  we have  $|\delta(j\omega)| \leq \min\{2, \omega|\Delta\tau|\}$ , which leads to the above conservation condition.

Regarding the phase margin relationship: The additional delay  $\Delta\tau$  causes an additional phase lag  $\omega\Delta\tau$ . In order not to exceed the existing phase margin  $\varphi_m$  at the cutoff frequency  $\omega_c$  (when  $|\tau(j\omega_c)| = 1$ ), needs  $\omega_c\Delta\tau < \varphi_m$ . This is a practical approximation condition when  $|T_0|$  reaches level 1 at  $\omega_c$ .

Theorem 3 has been proven.

✘

**Remark 2:** Theorem 1 explains that if the identification model is accurate,

the Smith predictor eliminates the delay from the characteristic polynomial, thus making it easy to control and stable according to the design for  $[C, \hat{G}]$ . This is the basis for explaining why Case 2 in the simulation gives good results. Theorem 2 gives the robust condition (small-gain). To keep the system stable when the model is not completely accurate (factor error), it is necessary to reduce the peak  $|T_0|$  around the bandwidth or reduce the relative error  $W_m$ . Empirically, if the identification is good (small  $W_m$ ) and the design of  $C$  is reasonable (no large peak in  $T_0$ ), the Smith predictor is still stable. Theorem 3 gives the condition with the delay error. If the delay error is small, satisfying the inequality with  $\min\{2, \omega|\Delta\tau|\}$  the system remains stable, the practical approximation is  $\omega_c\Delta\tau < \varphi_m$ .

### 3.3. Identification of the object model

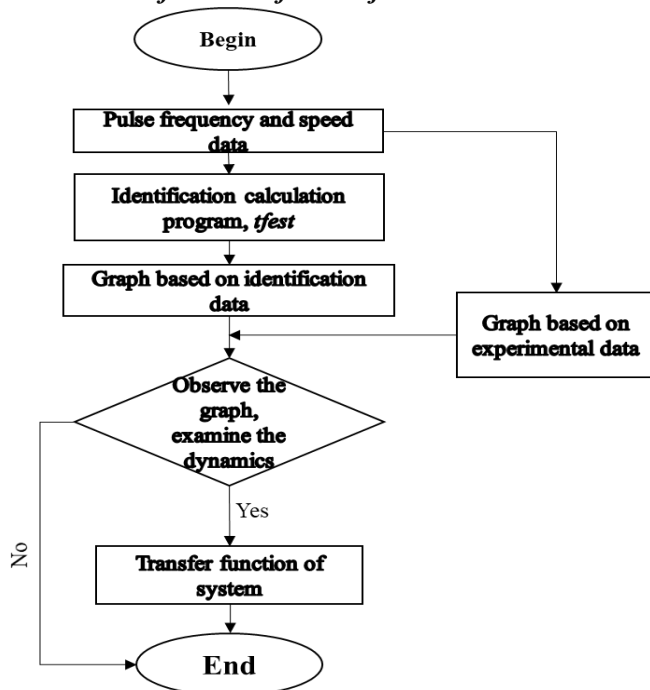


Figure 3. Flowchart of the system transfer function identification algorithm.

Next, to design the Smith delay compensator, the object model  $\hat{G}(s)$  is identified using MATLAB Identification Toolbox.

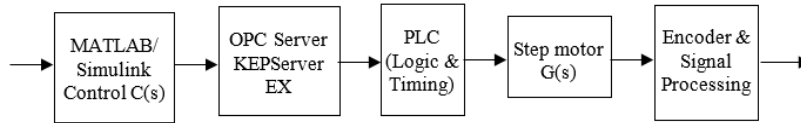
In the continuous time domain, the transfer function  $G(s)$  with a delay component describing the system's dynamics is a polynomial function of the following form:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_1 s + 1} e^{-\tau s}$$

$n \geq m$  (27)

To design a highly efficient controller, feedback system modeling is very important, especially with the Smith predictor. Studies [7-9] have proposed various system identification methods such as frequency analysis, spectrum analysis, recursive least squares method, maximum reliability method, state variable filter-based method, or existing toolbox-based method (CONTSID-Garnier Toolbox, 2008; CAPTAIN - Young, 2009). This study makes full use of MATLAB's identification toolbox with the *tfest* function. The model identification algorithm is implemented as shown in Figure 3. The system is activated with the input being the control pulse frequency, and the output being the motor speed. Figure 4 depicts the distributed delay structure in the MATLAB/Simulink-OPC Server-PLC-based real-time control system. The overall delay is modeled as an equivalent pure

delay resulting from computation, communication, execution, and sensing processes.



**Figure 4.** Distributed delay structure in the MATLAB/Simulink–OPC Server–PLC-based real-time control system.

To quantify the distributed delay in the MATLAB/Simulink–OPC–PLC architecture, each delay component was experimentally measured using timestamp-based logging and cycle-time monitoring within the PLC and Simulink environments, concretely:

$\tau_1$ : MATLAB/Simulink computation delay (0.05s);  $\tau_2$ : OPC read/write communication delay (0.15s);  $\tau_3$ : PLC execution and actuator interface delay (0.12s);  $\tau_4$ : Sensor measurement and feedback delay (0.8s). The sum of the equivalent delay:  $\tau = \tau_1 + \tau_2 + \tau_3 + \tau_4 = 0.4s$

The identified transfer function has the following form:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{0.7255}{s^2 + 15.19s + 77.4} e^{-0.4s} \quad (28)$$

**4. SIMULATION RESULTS**

**4.1. Semi-natural simulation model**

The authors use MATLAB’s OPC Toolbox library as an OPC Client to connect to the OPC Server (KEPServer Ex). This library consists of two main blocks, OPC Read and OPC Write, to read/write data to the Server. The sampling frequency for transmitting/receiving data between the MATLAB/Simulink Client and the OPC Server is 0.1s.

The OPC sampling period of 0.1s reflects a trade-off between communication load

and control performance. Experimental results indicate that the identified total delay (approximately 0.4 s) remains

within the stability margin derived in Section 3.3. Reducing the OPC sampling time further improves

tracking performance but increases network and PLC computational load, which may limit scalability in large-scale industrial systems.

Table 1 lists the tagged objects for linking and embedding OLE. For the experiment, the authors use PLC FX3U, Omron CWZ6C-1000 encoder, 2-phase stepper motor with DM422C driver.

**Table 1.** Tag definitions in KEPServerEx.

Variables	Modbus address	Data type	Describe
X0	X000	Boolean	Phase A encoder
X1	X001	Boolean	Phase B encoder
X5	X005	Boolean	Bit allows speed measurement
X7	X007	Boolean	Bit reset
D0	D0000	DWord	Save the pulse value
D10	D0010	DWord	Save the pulse frequency value

**4.2. Evaluate the quality of the system**

Before implementing the Smith predictor, the nominal closed-loop system (without delay) was analyzed in the frequency domain to evaluate robustness margins. Using MATLAB frequency response analysis, the following results were obtained: gain margin (GM): 9.6 dB, phase margin (PM): 52.3°, and crossover frequency ( $\omega_c$ ): 3.1 rad/s. These margins indicate that the nominal

delay-free system possesses sufficient robustness prior to delay compensation.

Using PID App Tuner of MATLAB/Simulink to simulate the design of a PID controller for the identification model. The coefficients

$$(K_p; K_i; K_d) = (27,65; 243,85; 0)$$

give the following system quality: Rise time 0.5 seconds; settling time 0.9 seconds and overshoot 1.61%.

Next, we experiment with 3 different cases:

Case 1: Feedback from the encoder in the feedback circuit, using the Smith predictor.

Case 2: Feedback from the identification in the feedback circuit, using the Smith predictor.

Case 3: Feedback from the encoder in the feedback circuit and no Smith predictor.

Figures 5, 6, and Table 2 show that in real-time control of the speed tracking system, to achieve a speed of 40 (rpm) when the feedback signal is the recognition model with the Smith predictor, it will give a better response. If we consider the raw signal from the encoder as a stable boundary state, there is almost no overshoot, the transition time is 1.5 seconds, and it starts to fluctuate around the set value with a steady-state error of 1.83% (0.7332 rpm); the error integral is 0.37 rpm<sup>2</sup>.s (after 1.5 seconds). Then the stable pulse frequency reaches 4267Hz.

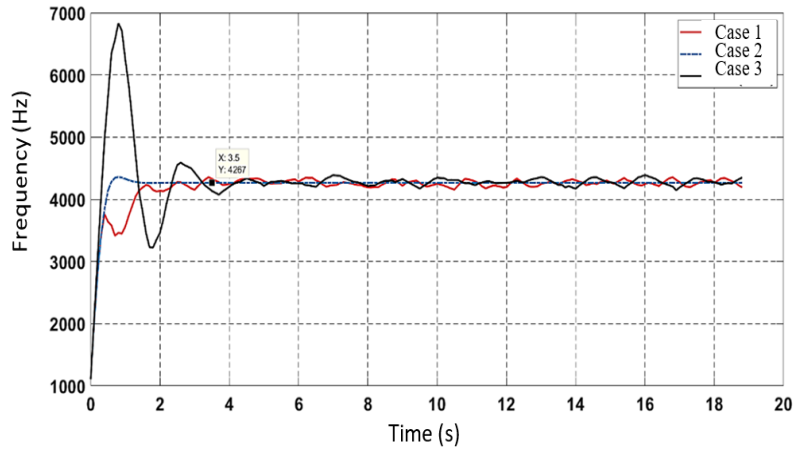


Figure 4. Control signal of the system.

The complementary sensitivity function  $T(s)$  of the nominal system was computed. The peak value was found to be:  $\|T(j\omega)\|_{\infty} = 1.38$ . Considering that the identified model uncertainty level is approximately 10–15% in gain and time constant variation, the small-gain condition:  $\|T(j\omega)\Delta(j\omega)\|_{\infty} < 1$ , is satisfied since:  $1.38 \times 0.15 = 0.207 < 1$ . From Theorem 3, delay robustness depends on the available phase margin. The maximum tolerable additional delay can be approximated as:

$$\Delta\tau_{max} \approx \frac{\phi_m}{\omega_c}$$

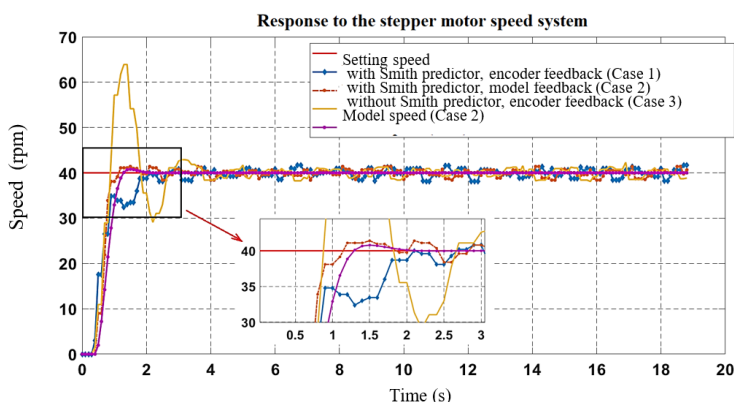
Substituting:

$$\Delta\tau_{max} \approx \frac{52.3^\circ \times \frac{\pi}{180}}{3.1} \approx 0.29 \text{ s}$$

Since the experimentally observed delay variation was within  $\pm 0.08$ s, the system operates well inside the derived delay stability bound. This confirms that the closed-loop system remains robustly stable under the considered model uncertainty.

Each experimental case was repeated five times under identical operating conditions. This confirms that the reported results are consistent and reproducible within the considered PLC-based control environment. Table

2 shows that, in real-time control of the speed tracking system, to achieve a speed of 40 (rpm) when the feedback signal is the recognition model with Smith predictor (Case 2), it will give the best response compared to Case 1 (encoder + Smith) and Case 3 (no Smith). Concretely: Settling time is reduced to 1.5s compared to 4.6s (Case 3); overshoot is 0% compared to 59.75% (Case 3); the lowest settling error is 1.83%.



**Figure 5.** System response at setting speed =40 rpm.

**Table 2.** System performance in comparison.

Scenario	Settling time (s)	Overshoot (%)	Steady-state error (%)	ISE (rpm <sup>2</sup> .s) (after 1.5s)	RMSE (rpm) (after 1.5s)
Case 1	2.1	0	2.47	1068.2	2.40
Case 2	1.5	0	1.83	927.9	0.22
Case 3	4.6	59.75	3.07	1316.4	0.85

In addition, we also tested the robustness for the case of the recognition model having a deviation ( $K$  decreased by 10%,  $T$  increased by 15%). Through simulation, the tracking quality remains stable, but the settling error increases by ~1%. This shows that the method is relatively sustainable against model errors.

Thus, when the system is recognized relatively accurately, and the feedback signal in the feedback circuit from the model using the Smith predictor gives the

best speed tracking quality of the stepper motor, as well as a stable control signal.

Compared with advanced delay compensation strategies such as Model Predictive Control (MPC), which require online optimization and higher computational resources, the Smith predictor provides a computationally efficient and easily implementable solution. This makes it particularly suitable for PLC-based industrial systems where memory and processing capability are limited.

**5. CONCLUSION**

This paper has presented a robust real-time speed tracking control approach for a two-phase stepper motor based on MATLAB/Simulink–OPC–PLC integration with Smith predictor delay compensation. An equivalent-delay modeling framework was developed to represent distributed communication and computation delays as a unified pure delay. Stability analysis showed that

when the model is accurate, the Smith predictor eliminates the delay from the characteristic equation. Furthermore, robustness against model uncertainty and delay mismatch was verified using small-gain conditions and phase-margin-based delay tolerance analysis. Frequency-domain verification confirmed a gain margin of 9.6dB and a phase margin of 52.3°, with a complementary sensitivity peak, ensuring robust stability under 10–15% model uncertainty. The

experimentally measured total delay of 0.40s remains within the derived delay tolerance bound. Semi-real-time experimental results demonstrate that the Smith predictor significantly improves speed tracking performance compared to conventional PID control. The configuration using model-based feedback within the Smith structure achieved the best performance, with a settling time of 1.5s, zero overshoot, and 1.83% steady-state error. The proposed approach provides a practical bridge between classical delay compensation theory and industrial PLC-based control implementation. It is particularly suitable for small-scale industrial automation systems and engineering education laboratories. Future work will focus on long-term validation under load disturbances and extension toward adaptive delay estimation strategies.

## REFERENCES

1. Onur Ayan, D. Zeynep Demirez, H. Kagan Kiziloz, Gizem Inci, Seckin Isleyen, Semih Ergin. The Detection of Spoiled Fruits on a Conveyor Belt Using Image Processing Techniques and OPC Server Software, *International Journal of Computational and Experimental Science and Engineering (IJCESSEN)* Vol. 4-No.1 (2018) pp. 11-15.
2. K.N Bagal, C.B.Kadu, P.S.Vikhe. PLC-Based Real-Time Process Control using SCADA and MATLAB. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Doi: 10.1109/ICCUBEA.2018.8697491.
3. Omur Aydogmus. A web-based educational tool using programmable logic controller-connected MATLAB-OPC server, *International Journal of Electrical Engineering Education*, Doi: 10.1177/0020720915571491.
4. Qingbao Huang, Qianzhong She, Xiaofeng Lin, Adaptive Fuzzy PID Temperature Control System Based on OPC and Modbus/TCP Protocol, 2010 2nd International Asia Conference on Infonnatics in Control, Automation and Robotics, Doi: 10.1109/CAR.2010.5456559.
5. Erwani Merry Sartika, T. Rudi Sarjono, Diki Dwi Saputra. Prediction of PID control model on PLC, *TELKOMNIKA*, Vol. 17, No.1, February 2019, pp.529-536. Doi: 10.12928/TELKOMNIKA.v17i1.11589.
6. K.Warwick, D.Rees. *Industrial digital control systems* (1988). Peter Peregrinus Ltd, London, United Kingdom.
7. Lennart Ljung. Experiments with Identification of Continuous Time Models, *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 6-8, 2009.
8. Lennart Ljung. On the estimation of transfer functions. 1985 International Federation of Automatic Control, Vol. 21, No. 6, pp. 677-696, 1985.
9. Kyeong Ha Lee, Seung Guk Baek, Jong Yoon Choi, Jae Jun Park, Hyouk Ryeol Choi, Hyungpil Moon, and Ja Choon Koo. System Identification of Hydraulic System Using RLS Algorithm for Feedforward Position Control.
10. Vũ Quốc Huy, Nguyễn Văn Đức, Chu Đức Chinh, Nguyễn Thị Lê Na. "Predictor control PI-Smith in the tracking system with time delay in image processing and communication", *Journal of Military Science and Technology*, FEE-2021, pp. 37-43, 2021.
11. Vũ Quốc Huy, Nguyễn Thị Lê Na, "A flexible SCADA architecture for auto-inspection of the electric motorcycles based on OPC server and PLC", *Journal of Military Science and Technology*, No. 75, pp. 65-73, 2021.

# ĐIỀU KHIỂN THÍCH NGHI PHI MÔ HÌNH CFDL-MFAC BÙ TRỄ CHO MỘT LỚP HỆ PHI TUYẾN SISO RỜI RẠC CÓ TRỄ ĐẦU VÀO BIẾN THIÊN

Lê Bá Yến<sup>1</sup>, Vũ Quốc Huy<sup>1\*</sup>

<sup>1</sup> Viện Tự động hóa (CAPITI), 89B Lý Nam Đế, Hoàn Kiếm, Hà Nội

\*Email liên hệ: maihuynv@gmail.com

## TÓM TẮT

Bài báo này nghiên cứu bài toán điều khiển thích nghi phi mô hình cho một lớp hệ phi tuyến SISO rời rạc có trễ đầu vào biến thiên theo thời gian. Trên cơ sở phương pháp tuyến tính hóa động dạng nén (Compact Form Dynamic Linearization – CFDL), bài báo thiết lập một mô hình dữ liệu mới có xét đến trễ đầu vào, trong đó mối quan hệ giữa sai phân đầu ra và sai phân tín hiệu điều khiển được biểu diễn thông qua một đại lượng gọi là giả đạo hàm riêng phụ thuộc trễ (Delay-Pseudo Partial Derivative – Delay-PPD). Bài báo chứng minh sự tồn tại của Delay-PPD và từ đó tổng hợp được luật điều khiển thích nghi phi mô hình CFDL-MFAC có xét đến trễ đầu vào. Điều kiện đủ để đảm bảo tính bị chặn của tín hiệu vòng kín và tính hội tụ của sai số bám cũng được phân tích. Các kết quả mô phỏng trên một đối tượng phi tuyến có trễ đầu vào biến thiên cho thấy bộ điều khiển đề xuất cải thiện rõ rệt chất lượng bám so với các trường hợp bỏ qua trễ hoặc giả thiết trễ cố định sai. Điều này cho thấy hiệu quả của việc đưa trực tiếp thông tin trễ vào mô hình CFDL và cấu trúc điều khiển MFAC.

Từ khóa: Điều khiển thích nghi phi mô hình; CFDL; giả đạo hàm riêng; trễ thay đổi; DDC.

## ABSTRACT

This paper studies the problem of non-model adaptive control for a class of discrete SISO nonlinear systems with time-varying input delays. Based on the Compact Form Dynamic Linearization (CFDL) method, the paper establishes a new data model that considers input delays, in which the relationship between the output difference and the control signal difference is expressed through a quantity called the delay-dependent partial derivative (Delay-Pseudo Partial Derivative – Delay-PPD). The paper proves the existence of Delay-PPD and from it synthesizes a non-model adaptive control law, CFDL-MFAC, that considers input delays. Sufficient conditions to ensure the boundedness of the closed-loop signal and the convergence of the tracking error are also analyzed. Simulation results on a nonlinear system with variable input delays show that the proposed controller significantly improves tracking quality compared to cases that ignore delays or assume fixed delays. This demonstrates the effectiveness of directly incorporating delay information into the CFDL model and the MFAC control structure.

**Keywords:** Model-free adaptive control; CFDL; pseudo-partial derivative; variable delay; DDC.

Ngày nhận bài: 08/02/2026; Ngày sửa bài: 02/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. ĐẶT VẤN ĐỀ

Trong nhiều hệ thống điều khiển thực tế như robot, hệ servo cơ điện, hệ điều khiển qua mạng và các hệ cơ điện tử phức tạp, việc xây dựng mô hình động lực học chính xác của đối tượng thường gặp nhiều khó khăn do ảnh hưởng của phi tuyến mạnh, bất định tham số và nhiễu. Vì vậy, các phương pháp điều khiển dựa trên dữ liệu (Data-Driven Control – DDC) hoặc điều khiển phi mô hình (Model-Free Control – MFC) đã thu hút sự quan tâm đáng kể trong những năm gần đây. Một trong những hướng tiếp cận tiêu biểu là MFC do Fliess và Join đề xuất, trong đó bộ điều khiển được thiết kế trực tiếp từ dữ liệu vào-ra của hệ thống mà không cần mô hình toán học chi tiết của đối tượng [1].

Trên cơ sở đó, phương pháp điều khiển thích nghi phi mô hình (Model-Free Adaptive Control – MFAC) đã được phát triển như một khung lý thuyết điều khiển thích nghi cho các hệ phi tuyến rời rạc. Ý tưởng cốt lõi của MFAC là kỹ thuật tuyến tính hóa động, trong đó động lực học phi tuyến của hệ được biểu diễn dưới dạng quan hệ tuyến tính cục bộ giữa sai phân đầu ra và sai phân tín hiệu điều khiển thông qua đại lượng giả đạo hàm riêng (Pseudo Partial Derivative – PPD) [5]. Nhiều nghiên cứu tiếp theo đã phát triển các dạng tuyến tính hóa động khác nhau như dạng nén (Compact-Form Dynamic Linearization – CFDL), cho phép thiết kế bộ điều khiển thích nghi phi mô hình với cấu trúc đơn giản và hiệu quả cho các hệ phi tuyến rời rạc [8]. Những năm gần đây, MFAC tiếp tục được mở rộng và ứng dụng trong nhiều bài toán điều khiển thực tế như hệ điều khiển qua mạng có mất dữ liệu, hệ tự hành và các hệ phi tuyến phức tạp [6], [10]. Một số nghiên cứu cũng đề xuất các cải tiến trong quá trình cập nhật PPD nhằm nâng cao tốc độ hội tụ và khả

năng thích nghi của thuật toán, chẳng hạn như MFAC với hệ số quên biến thiên hoặc các dạng tuyến tính hóa động cải tiến [9]. Ngoài ra, các công trình tổng quan gần đây cho thấy MFAC đang trở thành một hướng nghiên cứu quan trọng trong điều khiển dựa trên dữ liệu cho các hệ phi tuyến hiện đại [4].

Trễ đầu vào là hiện tượng phổ biến trong nhiều hệ thống điều khiển thực tế do quá trình truyền thông, xử lý tín hiệu hoặc động học cơ cấu chấp hành. Trong nhiều trường hợp, trễ đầu vào có thể biến thiên theo thời gian, làm cho việc phân tích và thiết kế bộ điều khiển trở nên phức tạp hơn. Các hệ có trễ đầu vào đã được nghiên cứu rộng rãi trong lý thuyết điều khiển và nhiều kết quả quan trọng đã được phát triển trong phân tích ổn định cũng như thiết kế điều khiển cho các hệ dạng này [2], [3]. Tuy nhiên, phần lớn các nghiên cứu hiện nay về MFAC chủ yếu xét các hệ không có trễ đầu vào hoặc trễ cố định đơn giản. Khi trễ đầu vào tồn tại hoặc biến thiên theo thời gian, quan hệ giữa sai phân đầu ra và sai phân tín hiệu điều khiển không còn tuân theo dạng CFDL truyền thống. Nếu ảnh hưởng của trễ bị bỏ qua hoặc giả thiết trễ không chính xác, chất lượng điều khiển có thể bị suy giảm đáng kể.

Xuất phát từ vấn đề trên, bài báo này nghiên cứu mở rộng phương pháp CFDL-MFAC cho lớp hệ phi tuyến SISO rời rạc có trễ đầu vào biến thiên theo thời gian. Trên cơ sở kỹ thuật tuyến tính hóa động dạng nén, bài báo xây dựng một mô hình dữ liệu có xét đến trễ đầu vào và từ đó tổng hợp luật điều khiển thích nghi phi mô hình tương ứng. Các đóng góp chính của bài báo được tóm tắt như sau:

Thứ nhất, bài báo mở rộng mô hình tuyến tính hóa động dạng nén CFDL truyền thống sang trường hợp hệ phi tuyến SISO rời rạc có trễ đầu vào biến thiên

theo thời gian, từ đó thiết lập mô hình dữ liệu mới dưới dạng Delay-CFDL. So với các nghiên cứu MFAC thông thường vốn chủ yếu xét tín hiệu điều khiển tác động tức thời hoặc trễ cố định, mô hình này phản ánh trực tiếp ảnh hưởng của trễ đầu vào lên quan hệ giữa sai phân đầu ra và sai phân điều khiển.

Thứ hai, bài báo đưa ra khái niệm giả đạo hàm riêng phụ thuộc trễ (Delay-PPD) và chứng minh sự tồn tại của đại lượng này. Đây là điểm mở rộng lý thuyết quan trọng so với CFDL-MFAC cổ điển, vì trong trường hợp có trễ đầu vào biến thiên, đại lượng PPD truyền thống không còn mô tả đúng quan hệ dữ liệu giữa đầu vào và đầu ra của hệ.

Thứ ba, trên cơ sở mô hình Delay-CFDL, bài báo tổng hợp luật điều khiển Delay-CFDL-MFAC và phân tích điều kiện để hệ kín duy trì tính bị chặn cũng như khả năng hội tụ sai số bám. Kết quả mô phỏng cho thấy khi trễ đầu vào thay đổi theo thời gian, việc đưa trực tiếp thông tin trễ vào mô hình CFDL và cấu trúc điều khiển giúp cải thiện chất lượng bám so với các trường hợp bỏ qua trễ hoặc giả thiết trễ cố định.

## 2. MÔ HÌNH DELAY-CFDL VÀ ĐỊNH LÝ TỒN TẠI DELAY-PPD

### 2.1. Lớp đối tượng nghiên cứu và giả thiết

Xét hệ phi tuyến SISO rời rạc có trễ đầu vào:

$$y(k+1) = f(y_k, u_{k-d(k)}) \quad (1)$$

Trong đó:

$$y_k = [y(k), y(k-1), \dots, y(k-n_y+1)] \quad (2)$$

$$u_{k-d(k)} = [u(k-d(k)), u(k-d(k)-1), \dots, u(k-d(k)-n_u+1)] \quad (3)$$

Trễ đầu vào chưa biết, có thể biến

thiên theo thời gian,  $d_{max}$  là trễ lớn nhất:

$$d(k) \in D := \{0, 1, 2, \dots, d_{max}\} \quad (4)$$

Ký hiệu sai phân:

$$\Delta y(k+1) = y(k+1) - y(k) \quad (5)$$

$$\Delta u(k) = u(k) - u(k-1)$$

**Giả thiết:**

**(A1)** - (Đặc tính trơn theo biến điều khiển tác động): Với mọi  $k, f$  khả vi theo đối số  $u(k-d(k))$  và đạo hàm riêng theo đối số đó liên tục trên miền làm việc.

**(A2)** - (Điều kiện Lipschitz tổng quát theo sai phân điều khiển trễ): Tồn tại  $L > 0$  sao cho

$$|\Delta y(k+1)| \leq L |\Delta u(k-d(k))|, \forall k \quad (6)$$

**(A3)** - (Trễ bị chặn và biến thiên chậm):

$$d(k) \in D, \quad |d(k+1) - d(k)| \leq 1 \quad (7)$$

### 2.2. Định lý về sự tồn tại Delay-PPD

**Phát biểu:**

Nếu hệ (1) thỏa mãn giả thiết A1, A2 thì tồn tại một đại lượng  $\phi(k)$  (được gọi là Delay-PPD) sao cho:

$$\Delta y(k+1) = \phi(k) \Delta u(k-d(k)) \quad (8)$$

Và bị chặn:

$$|\phi(k)| \leq L, \quad \forall k \quad (9)$$

**Chứng minh:**

a) *Chứng minh tồn tại  $\phi(k)$*

Viết sai phân đầu ra dưới dạng chênh lệch hai giá trị của  $f$ . Từ (1) ta có:

$$y(k+1) = f(y_k, u_{k-d(k)}) \quad (10)$$

Tương tự:

$$y(k) = f(y_{k-1}, u_{k-1-d(k-1)}) \quad (11)$$

Suy ra:

$$\Delta y(k+1) = f(y_k, u_{k-d(k)}) - f(y_{k-1}, u_{k-1-d(k-1)}) \quad (12)$$

Tách phần biến thiên do riêng biến  $u(k-d(k))$ . Xây dựng một véc-tơ

điều khiển lai  $\tilde{u}_{k-d(k)}$  bằng cách thay phân tử đầu tiên của  $u_{k-d(k)}$  từ thành  $u(k-d(k)-1)$ , còn các phân tử khác giữ nguyên ta được:

$$\begin{aligned} \tilde{u}_{k-d(k)} &:= [u(k-d(k)-1), \\ &u(k-d(k)-1), \\ &\dots, u(k-d(k)-n_u+1)] \end{aligned} \quad (13)$$

Ở đây chỉ cần đảm bảo  $\tilde{u}_{k-d(k)}$  và  $u_{k-d(k)}$  khác nhau đúng một biến là  $u(k-d(k)-1)$ . Khi đó ta cộng-trừ  $f(y_k, \tilde{u}_{k-d(k)})$  trong (12) nhận được:

$$\Delta y(k+1) = A(k) + r(k) \quad (14)$$

Trong đó:

$$A(k) := f(y_k, u_{k-d(k)}) - f(y_k, \tilde{u}_{k-d(k)}) \quad (15)$$

$$r(k) := f(y_k, \tilde{u}_{k-d(k)}) - f(y_{k-1}, u_{k-1-d(k-1)}) \quad (16)$$

$A(k)$  là phần chênh lệch do thay đổi riêng  $u(k-d(k))$ ; còn  $r(k)$  gom mọi thay đổi còn lại (thay đổi  $y$ , các thành phần điều khiển khác và cả việc trễ có thể đổi).

Áp dụng định lý giá trị trung bình cho  $A(k)$ . Theo giả thiết A1, ta có hàm  $g(\alpha)$

$$g(\alpha) = f(y_k, \tilde{u}_{k-d(k)} + \alpha(u_{k-d(k)} - \tilde{u}_{k-d(k)})), \alpha \in [0,1] \quad (17)$$

khả vi theo  $\alpha$ . Vì  $u_{k-d(k)} - \tilde{u}_{k-d(k)}$  chỉ có đúng một phân tử khác 0 tại vị trí  $u(k-d(k))$ , nên theo định lý giá trị trung bình tồn tại  $\alpha_k \in (0,1)$  sao cho:

$$A(k) = g(1) - g(0) = g'(\alpha_k)$$

Mà  $g'(\alpha)$  chính là đạo hàm riêng theo  $u(k-d(k))$  nhân với  $\Delta u(k-d(k))$ .

Do đó tồn tại một điểm trung gian  $\xi_k$  trên đoạn  $[u(k-d(k)-1), u(k-d(k))]$  sao cho:

$$A(k) = \frac{\partial f}{\partial u(k-d(k))} \Big|_{\xi_k} \Delta u(k-d(k)) \quad (19)$$

Đặt:

$$\phi_0(k) := \frac{\partial f}{\partial u(k-d(k))} \Big|_{\xi_k} \quad (20)$$

Khi đó:

$$A(k) = \phi_0(k) \Delta u(k-d(k)) \quad (21)$$

Hấp thụ  $r(k)$  vào hệ số nhân  $\Delta u(k-d(k))$ , từ (14) và (21):

$$\Delta y(k+1) = \phi_0(k) \Delta u(k-d(k)) + r(k) \quad (22)$$

Ta xét hai trường hợp:

\* Nếu  $\Delta u(k-d(k)) \neq 0$ :

Định nghĩa:

$$\phi(k) := \phi_0(k) + \frac{r(k)}{\Delta u(k-d(k))} \quad (23)$$

Thế (23) vào (22) suy ra ngay:

$$\Delta y(k+1) = \phi(k) \Delta u(k-d(k)) \quad (24)$$

\* Nếu  $\Delta u(k-d(k)) = 0$ :

Từ giả thiết A2, ta có  $|\Delta y(k+1)| \leq 0 \Rightarrow \Delta y(k+1) = 0$ . Khi đó chọn  $\phi(k) = 0$  thì (24) vẫn đúng. Vậy (8) đúng với mọi  $k$ .

b) Chứng minh  $\phi(k)$  bị chặn

Nếu  $\Delta u(k-d(k)) \neq 0$ , từ (8) và giả thiết A2:

$$|\phi(k)| = \frac{\Delta y(k+1)}{\Delta u(k-d(k))} \leq L \frac{|\Delta u(k-d(k))|}{|\Delta u(k-d(k))|} = L \quad (25)$$

Nếu  $\Delta u(k-d(k)) = 0$ , ta chọn  $\phi(k) = 0$  nên cũng thỏa  $|\phi(k)| \leq L$ .

Suy ra (9) đúng.

Định lý 1 được chứng minh.

□

**Nhận xét.** Định lý 1 cho phép mô tả hệ (1) bằng mô hình dữ liệu Delay-CFDL dạng (8) tại mỗi thời điểm lấy mẫu mà không cần mô hình động lực học tường minh. Đây là nền tảng để ước lượng Delay-PPD và tổng hợp điều khiển bù trễ.

### 3. ƯỚC LƯỢNG TRỄ ĐẦU VÀO

Vì  $d(k)$  chưa biết, ta xét tập ứng viên  $l \in D$ . Với mỗi  $l$ , giả sử tạm thời trễ bằng  $l$ , mô hình Delay-CFDL dự báo:

$$\Delta y(i+1) \approx \hat{\phi}_l(i) \Delta u(i-l) \quad (26)$$

Dùng luật cập nhật chuẩn hóa (trung thích MFAC):

$$\begin{aligned} \hat{\phi}_l(i) &= \hat{\phi}_l(i-1) + \frac{\eta \Delta u(i-l)}{\mu + \Delta u^2(i-l)} (\Delta y(i+1) \\ &- \hat{\phi}_l(i-1) \Delta u(i-l)) \end{aligned} \quad (27)$$

với  $\eta \in (0,1], \mu > 0$ .  $\eta$  là hệ số cập nhật của PPD,  $\eta$  là tham số điều chỉnh trong luật cập nhật.

Trên cửa sổ  $W$  mẫu gần nhất ( $W$  là kích thước cửa sổ dữ liệu), định nghĩa chỉ số lỗi:

$$J_l(k) = \sum_{i=k-W+1}^k \frac{(\Delta y(i+1) - \phi_l(i)\Delta u(i-l))^2}{\epsilon + \Delta u^2(i-l)}, \quad \epsilon > 0 \quad (28)$$

Chọn:

$$\hat{d}_{raw}(k) = \arg \arg J_l(k) \quad (29)$$

Dùng lọc đa số trên  $N_h$  bước ( $N_h$  là độ dài cửa sổ làm tròn):

$$\hat{d}(k) = \text{mode}\{\hat{d}_{raw}(k), \hat{d}_{raw}(k-1), \dots, \hat{d}_{raw}(k-N_h+1)\} \quad (30)$$

Chỉ cập nhật nếu  $J_l(k) \leq J_{th}$ , ngược lại giữ  $\hat{d}(k) = \hat{d}(k-1)$ .

#### 4. LUẬT ĐIỀU KHIỂN CFDL-MFAC BÙ TRỄ: DELAY-CFDL-MFAC

Mục tiêu bám  $y_d(k)$ , sai số (9):

$$e(k) = y_d(k) - y(k) \quad (31)$$

Do điều khiển tác động lên đầu ra sau bước, xét hàm mục tiêu (32), là tham số điều khiển.

$$J_u(k) = e^2(k + \hat{d}(k) + 1) + \lambda \Delta u^2(k), \quad \lambda > 0 \quad (32)$$

Từ Delay-CFDL, với trễ  $\hat{d}(k)$ :

$$\Delta y(k + \hat{d}(k) + 1) \approx \hat{\phi}_{\hat{d}(k)}(k)\Delta u(k) \quad (33)$$

Suy ra xấp xỉ (34):

$$y(k + \hat{d}(k) + 1) \approx y(k + \hat{d}(k)) + \hat{\phi}_{\hat{d}(k)}\Delta u(k) \quad (34)$$

Đặt:

$$\underline{e}(k) := y_d(k + \hat{d}(k) + 1) - y(k + \hat{d}(k)) \quad (35)$$

Thay (34) vào (32), lấy đạo hàm theo  $\Delta u(k)$  và giải điều kiện tối ưu, thu được (36),  $\rho$  là một tham số điều khiển.

$$\Delta u(k) = \rho \frac{\hat{\phi}_{\hat{d}(k)}(k)}{\lambda + \hat{\phi}_{\hat{d}(k)}^2(k)} \underline{e}(k), \rho \in (0,1] \quad (36)$$

Ta tổng hợp được luật điều khiển dạng truy hồi (37):

$$u(k) = u(k-1) + \rho \frac{\hat{\phi}_{\hat{d}(k)}(k)}{\lambda + \hat{\phi}_{\hat{d}(k)}^2(k)} (y_d(k + \hat{d}(k) + 1) - y(k + \hat{d}(k))) \quad (37)$$

#### 5. PHÂN TÍCH TOÁN HỌC ỔN ĐỊNH HỆ KÍN

Từ (27) và giả thiết A2,  $\phi(k)$  bị chặn bởi  $L$ . Với việc chọn  $\eta > 0$  và cơ chế đặt lại như [7] (nếu  $|\Delta u(i-l)| < \epsilon_u$  hoặc  $|\hat{\phi}_l(i)| < \epsilon_\phi$  thì  $\hat{\phi}_l(i) = \hat{\phi}_0$ ).

Suy ra tồn tại  $M_\phi > 0$  sao cho:

$$|\hat{\phi}_l(k)| \leq M_\phi, \forall l \in D, \forall k \quad (38)$$

Giả sử tại thời điểm  $k$ :  $\hat{d}(k) = d(k)$ : và  $\hat{\phi}_{\hat{d}(k)}(k) = \phi(k + d(k))$  (xấp xỉ lý tưởng). Khi đó từ (34)–(37) suy ra (39):

$$e(k + d(k) + 1) = \left(1 - \rho \frac{\phi^2}{\lambda + \phi^2}\right) \underline{e}(k) \quad (39)$$

với  $\phi = \phi(k + d(k))$ .

Do  $\lambda > 0, \rho \in (0,1]$  nên:

$$0 < \left|1 - \rho \frac{\phi^2}{\lambda + \phi^2}\right| < 1$$

Suy ra sai số tương lai giảm theo hệ số co.

Khi  $\hat{d}(k)$  được chọn bởi (28) – (30), với giả thiết A3 và điều kiện kích thích đủ trên cửa sổ, tức  $\Delta u(\cdot)$  không đồng nhất 0 trong  $W$  mẫu,  $\hat{d}(k)$  sẽ đứng sau một khoảng quá độ quanh các thời điểm trễ đổi. Khi đó, trong các khoảng thời gian mà giá trị trễ được lựa chọn đúng, sai số bám thỏa tính chất co theo (39)–(40); còn trong các khoảng quá độ khi trễ thay đổi hoặc bị lựa chọn sai tạm thời, sai số vẫn bị chặn bởi các chặn hữu hạn của tín hiệu vòng kín. Vì vậy, sai số bám của hệ kín là bị chặn và hội tụ về một lân cận

nhỏ của gốc.

Như vậy, so với CFDL-MFAC truyền thống, kết quả ổn định thu được trong bài báo cho thấy cấu trúc điều khiển đề xuất vẫn duy trì được tính chất ổn định đầu vào-đầu ra trong điều kiện có trễ đầu vào biến thiên, với cái giá phải trả là sai số bám chỉ hội tụ về một lân cận nhỏ thay vì triệt tiêu hoàn toàn trong mọi thời điểm.

## 6. MÔ PHỎNG SỐ VÀ BÌNH LUẬN

### 6.1. Kịch bản mô phỏng

Để đánh giá hiệu quả của phương pháp đề xuất, bài báo thực hiện mô phỏng trên một đối tượng phi tuyến rời rạc có trễ đầu vào biến thiên theo thời gian. Trễ đầu vào được giả thiết thay đổi theo từng đoạn thời gian nhằm mô phỏng các tình huống trễ biến thiên trong thực tế. Các chỉ tiêu đánh giá gồm sai số RMS, chỉ tiêu tích phân IAE, năng lượng điều khiển  $E_u$  và độ quá chỉnh. Ba bộ điều khiển được so sánh gồm: C1: CFDL-MFAC bỏ qua trễ đầu vào ( $d = 0$ ); C2: CFDL-MFAC với giả thiết trễ cố định ( $d = 1$ ); C3: Bộ điều khiển đề xuất, trong đó thông tin trễ biến thiên được đưa trực tiếp vào cấu trúc điều khiển Delay-CFDL-MFAC.

### 6.2. Tham số mô phỏng

$$y(k+1) = 0.6y(k) + 0.2\sin(y(k)) + \frac{0.8u(k-d(k))}{1+u^2(k-d(k))} \quad (41)$$

Trễ biến thiên có dạng (42):

$$d(k) = \begin{cases} 1: & 1 \leq k < 250 \\ 3: & 250 \leq k < 500 \\ 2: & 500 \leq k \leq 800 \end{cases} \quad (42)$$

Tín hiệu mong muốn và tham số mô phỏng có dạng (43), (44), (45):

$$d(k) = \begin{cases} 0: & 1 \leq k < 80 \\ 1: & 80 \leq k < 220 \\ 0.6: & 220 \leq k \leq 380 \\ 1.2: & 380 \leq k \leq 560 \end{cases} \quad (43)$$

$$W = 40, N_h = 5, \eta = 0.75, \mu = 10^{-3}, \epsilon = 10^{-3} \quad (44)$$

$$\lambda = 0.5, \rho = 0.8, \hat{\phi}_l(0) = 0.5, u(k) = 0 \quad (k \leq 0) \quad (45)$$

$$-0.15 \sin \sin (0.04(k - 560)) : k \geq 560 \quad (43)$$

$$W = 40, N_h = 5, \eta = 0.75, \mu = 10^{-3}, \epsilon = 10^{-3} \quad (44)$$

$$\lambda = 0.5, \rho = 0.8, \hat{\phi}_l(0) = 0.5, u(k) = 0 \quad (k \leq 0) \quad (45)$$

Ngoài ra, để tránh tín hiệu điều khiển tăng quá lớn trong mô phỏng số, tín hiệu điều khiển được giới hạn bởi:

$$|\Delta u(k)| \leq \Delta u_{max}, |u(k)| \leq u_{max}.$$

Ngưỡng  $J_{th}$  của hàm đánh giá được chọn theo thực nghiệm. Trong các mô phỏng của bài báo, giá trị  $J_{th} = 0.05$ , được sử dụng khi áp dụng dạng chuẩn hóa trong (28). Trong bài báo này, mô phỏng chính được thực hiện cho ba trường hợp: Bỏ qua trễ đầu vào, sử dụng trễ cố định và bù đúng trễ biến thiên theo thời gian. Các tham số liên quan đến cơ chế chọn trễ trực tuyến như  $d_{max}, W, N_h, \epsilon, J_{th}$  chỉ được sử dụng trong các thử nghiệm phụ, không phải là một phần của mô phỏng chính.

### 6.3. Kết quả mô phỏng

Bảng 1. So sánh hiệu năng tổng thể.

Kịch bản	RMS	IAE		Độ quá chỉnh
C1: Bỏ qua trễ	0.5690	401.95	0.5623	0.8981
C2: Trễ cố định	0.1499	82.83	1.2258	0.4529
C3: Thuật toán bù trễ đề xuất	<b>0.1427</b>	<b>60.09</b>	1.2646	0.4529

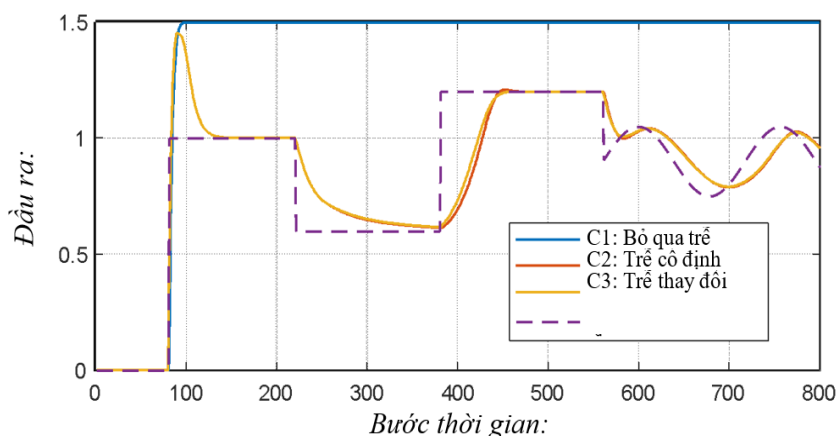
Theo Bảng 1, có thể thấy rằng bộ điều khiển bỏ qua trễ (C1) cho chất lượng điều khiển kém nhất. Khi trễ đầu vào được xét đến dưới dạng cố định (C2), sai số bám giảm đáng kể. Tuy nhiên, bộ điều khiển đề xuất (C3) cho kết quả tốt nhất về cả chỉ tiêu RMS và IAE.

Bảng 2. Kết quả theo từng đoạn trễ

Đoạn trễ	Lượng trễ	RMS C1	RMS C2	RMS C3
1	1	0.4814	0.1517	0.1517

2	3	0.6819	0.2063	<b>0.1905</b>
3	2	0.5316	0.0731	<b>0.0701</b>

Theo Bảng 2, trong đoạn đầu khi trễ thực bằng 1, bộ điều khiển C2 và C3 cho kết quả gần tương đương do giá trị trễ giả thiết trùng với trễ thực của hệ. Tuy nhiên, khi trễ đầu vào thay đổi sang các giá trị khác, sai số bám của C2 tăng lên đáng kể, trong khi bộ điều khiển đề xuất vẫn duy trì chất lượng bám tốt hơn.



**Hình 1.** So sánh đáp ứng đầu ra của hệ với ba bộ điều khiển.

Hình 1 thể hiện đáp ứng đầu ra của hệ đối với ba chiến lược điều khiển. Có thể thấy bộ điều khiển C1, trong đó ảnh hưởng của trễ đầu vào bị bỏ qua, cho sai số bám lớn và độ quá chỉnh cao. Điều này cho thấy giả thiết không có trễ là không phù hợp đối với hệ có trễ đầu vào. Khi trễ đầu vào được xét đến dưới dạng cố định (C2), chất lượng điều khiển được cải thiện đáng kể so với C1. Tuy nhiên, do trễ thực của hệ thay đổi theo thời gian nên bộ điều khiển này chỉ cho kết quả tốt trong đoạn đầu khi . Ngược lại, bộ điều khiển đề xuất (C3), trong đó thông tin trễ biến thiên được đưa trực tiếp vào cấu trúc điều khiển Delay-CFDL-MFAC, cho đáp ứng bám chính xác hơn trên toàn bộ miền mô phỏng. Sai số bám giảm rõ rệt khi trễ đầu vào thay đổi sang các giá trị khác nhau,

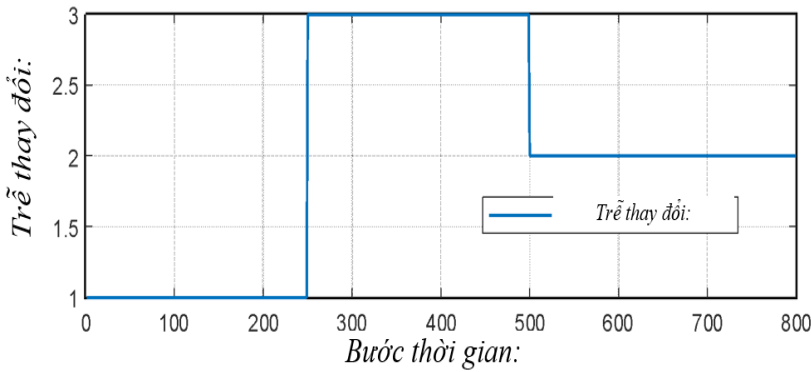
thể hiện ưu thế của phương pháp đề xuất.

Hình 2 biểu diễn tín hiệu điều khiển tương ứng của ba bộ điều khiển. Có thể nhận thấy rằng tín hiệu điều khiển của phương pháp C1 dao động mạnh hơn do bộ điều khiển không xét đến trễ đầu vào của hệ. Trong trường hợp C2 và C3, tín hiệu điều khiển ổn định hơn do cấu trúc điều khiển đã xét đến ảnh hưởng của trễ. Mặc dù năng lượng điều khiển của

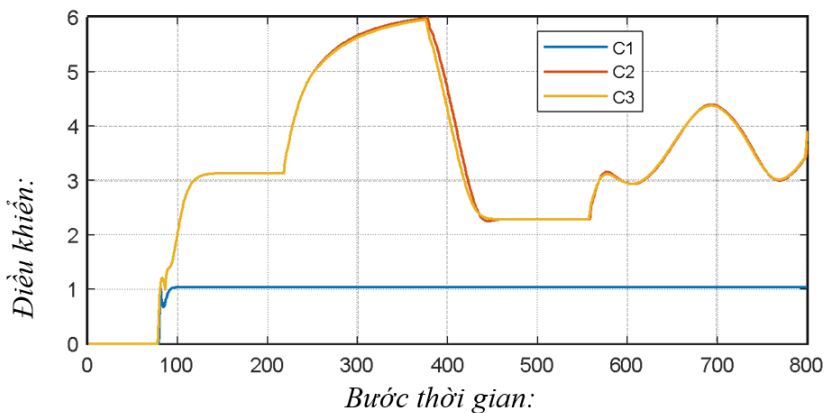
phương pháp đề xuất (C3) tăng nhẹ so với trường hợp trễ cố định (C2), bộ điều khiển này đạt được chất lượng bám tốt hơn đáng kể. Điều này cho thấy việc đưa thông tin trễ vào mô hình dự báo của bộ

điều khiển giúp cải thiện hiệu quả điều khiển mà không làm tăng đáng kể biên độ tín hiệu điều khiển.

Hình 3 mô tả sự thay đổi của trễ đầu vào theo thời gian trong kịch bản mô phỏng. Trễ đầu vào được giả thiết thay đổi theo ba đoạn với các giá trị lần lượt là , và . Kịch bản này nhằm mô phỏng các tình huống thực tế trong đó trễ đầu vào có thể thay đổi do các yếu tố như độ trễ truyền thông, thời gian xử lý hoặc đặc tính động học của cơ cấu chấp hành. Việc sử dụng trễ biến thiên cho phép đánh giá khả năng thích nghi của bộ điều khiển đề xuất khi điều kiện làm việc của hệ thay đổi. Kết quả mô phỏng cho thấy bộ điều khiển Delay-CFDL-MFAC duy trì chất lượng bám tốt ngay cả khi trễ đầu vào thay đổi theo thời gian.



**Hình 2.** So sánh tín hiệu điều khiển của ba phương pháp.



**Hình 3.** Trễ đầu vào biến thiên theo thời gian được sử dụng trong mô phỏng.

**Nhận xét:** Từ các kết quả trên có thể rút ra một số nhận xét như sau: Khi bỏ qua ảnh hưởng của trễ đầu vào, chất lượng điều khiển suy giảm đáng kể. Việc giả thiết trễ cố định chỉ cho kết quả tốt khi giá trị trễ thực gần với giá trị giả thiết. Bộ điều khiển Delay-CFDL-MFAC cho kết quả tốt nhất khi trễ đầu vào biến thiên theo thời gian, thể hiện ở sai số bám nhỏ hơn và khả năng thích nghi tốt hơn với sự thay đổi của trễ.

## 7. KẾT LUẬN

Bài báo đã nghiên cứu bài toán điều khiển thích nghi phi mô hình cho lớp hệ phi tuyến SISO rời rạc có trễ đầu vào biến thiên theo thời gian. Trên cơ sở phương pháp tuyến tính hóa động dạng nén CFDL, bài báo đã xây dựng một mô hình dữ liệu mới có xét đến trễ đầu vào

và đưa ra khái niệm Delay-PPD. Sự tồn tại của đại lượng này đã được chứng minh, từ đó cho phép tổng hợp luật điều khiển CFDL-MFAC bù trễ cho hệ có trễ đầu vào biến thiên. Kết quả mô phỏng cho thấy khi trễ đầu vào thay đổi theo thời gian, bộ điều khiển đề xuất cho chất lượng bám tốt hơn rõ rệt so với trường hợp bỏ qua trễ hoặc giả thiết trễ cố định sai. Điều này khẳng định tính cần thiết của việc đưa trực tiếp thông tin trễ vào mô hình CFDL và cấu trúc điều khiển MFAC. Trong các nghiên cứu tiếp theo, hướng mở rộng có thể là phát triển các cơ chế giám sát hoặc nhận dạng trễ trực tuyến nhằm cung cấp thông tin trễ cho bộ điều khiển trong các hệ thống thực tế.

## TÀI LIỆU THAM KHẢO

- [1] Fliess, M. (2013). Model-free control. International Journal of Control. <https://doi.org/10.1080/00207179.2013.810345>
- [2] Jean-Pierre Richard, Time-delay systems: an overview of some recent advances and open problems, Automatica, Volume 39, Issue 10, 2003, Pages 1667-1694, ISSN 0005-

1098, [https://doi.org/10.1016/S0005-1098\(03\)00167-5](https://doi.org/10.1016/S0005-1098(03)00167-5).

[3] E. Fridman, *Introduction to Time-Delay Systems: Analysis and Control*, Springer, 2014.

[4] Zhang C, Cen C, Huang J. An Overview of Model-Free Adaptive Control for the Wheeled Mobile Robot. *World Electric Vehicle Journal*. 2024; 15(9):396. <https://doi.org/10.3390/wevj15090396>

[5] Z.-S. Hou, S.-T. Jin, *Model-Free Adaptive Control: Theory and Applications*, CRC Press, London, UK, 2013.

[6] Z. S. Hou and X. H. Bu, “Model-free adaptive control with data dropouts”, *Expert Systems with Applications*, vol. 38, no. 8, pp. 10709–10717, 2011.

[7] Trần Ngọc Bình, Vũ Quốc Huy, “Tuyến tính hóa động dạng rút gọn CFDL và điều khiển thích nghi phi mô

hình MFAC trong hệ SISO rời rạc”, *Tạp chí Nghiên cứu KH-CN quân sự*, số 98, 2024, 15-22.

[8] Z. Hou and Y. Zhu, “Controller-Dynamic-Linearization-Based Model Free Adaptive Control for Discrete-Time Nonlinear Systems,” in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2301-2309, Nov. 2013, doi: 10.1109/TII.2013.2257806.

[9] Hoang-Anh Pham, Dirk Söffker, “Modified Model-Free Adaptive Control Using Compact-Form Dynamic Linearization Technique,” *IFAC World Congress*, 2020.

[10] Yulei Liao, Tingpeng Du, Quanquan Jiang, “Model-free adaptive control method with variable forgetting factor for unmanned surface vehicle control,” *Applied Ocean Research*, Volume 93, 2019, 101945, ISSN 0141-1187, <https://doi.org/10.1016/j.apor.2019.101945>.

# SIMULATION OF FIBER DEPOSITION ON 3D PRINTED MOLDED FIBER SCREEN USING MULTI-SPHERE DISCRETE ELEMENT METHOD

**K. Q. Le<sup>1,2</sup>, J. W. Chew<sup>1,3</sup>, M. F. Leyva-Mendivil<sup>1,4</sup>, T. H. Thai<sup>5</sup>, F. Duan<sup>1,\*</sup>**

<sup>1</sup>HP-NTU Digital Manufacturing Corporate Lab, Nanyang Technological University, 639798, Singapore

<sup>2</sup>Vietnam Japan University, Vietnam National University, Hanoi, Vietnam

<sup>3</sup>School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 639798, Singapore

<sup>4</sup>3D Lab, HP Labs, HP Inc., Palo Alto, California 94304, USA

<sup>5</sup>Asia East University of Technology, Vietnam

<sup>6</sup>School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798, Singapore

\* Corresponding author. E-mail address: feiduan@ntu.edu.sg

## ABSTRACT

*In line with the sustainable development goal, molded fiber products play important roles in reducing plastic-based packaging. To fabricate molded fiber products, besides using conventional meshing tools, 3D printing is employed to manufacture the molded fiber screen. 3D printing technique allows printing molded fiber screens with complex geometry, flexible in pore size and shape. The 3D printed molded fiber screens are in the process of investigation to improve the de-watering efficiency, fiber collection, mechanical strength, etc. In addition, the fiber distribution on the screen is also necessary to assess the quality of the screen. Besides using experimental methods to capture the fiber distribution on screen, simulation also offers using tools to assess the uniformity of the fiber. In this study, the multi-sphere model was employed to simulate the fibers. The interaction of the fibers was able to mimic by employing the discrete element method. The fiber distribution was captured and compared to the experiment. The simulation results were able to reveal the fiber deposition layer upon layer and explain the formation of uneven thickness in the tilted area of the molded fiber screen.*

**Keywords:** 3D printing, molded fiber screen, multi-sphere model, discrete element method.

*Ngày nhận bài: 20/02/2026; Ngày sửa bài: 12/03/2026; Ngày duyệt đăng: 02/04/2026*

### 1. INTRODUCTION

In the past several decades, plastic-based products have been widely used in the packaging industry due to their high performance, low cost, light weight, and easy processing [1]. However, most of the plastic products are non-biodegradable, causing plastic pollution, which threatens even human health [2]. It is an alarm for scientists and human beings to find an alternative solution to replace plastic-based packaging products. In recent years, molded fiber products have attracted attention due to their renewability, recyclability, sustainability, and biodegradability [3]. Hence, molded fiber products become a strong candidate to reduce the plastic waste from the packaging industry.

In conventional manufacturing, meshing tools were employed to fabricate molded fiber products [4]. However, it faces the limitation in fabricating complex geometry. It is known that pore size, shape, tilted angle of the pore axis, pore arrangement, and surface structure directly affect the de-watering speed of the molded fiber screen. However, it is challenging to arrange the meshing tools to generate these changes in the pore size, shape, and tilted angle of the pore axis by the conventional method. Hence, in recent years, 3D printing has been adopted to fabricate molded fiber screen [5]. It allows the fabrication of complex geometries, suitable for complex packaging. In addition, the pore size and shape can be varied without any hindrance [6].

The 3D printing molded fiber screen is still in progress to design the novelty screen to enhance the de-watering speed, fiber collection, and increase its strength. In fact, the molded fiber products experience the unevenness of

their thickness at different locations, especially at the tilted surface. There is a need to understand the cause of this issue, and it can be revealed by capturing the fiber deposition on the screen.

Some propose using an experimental setup to capture the fiber deposition on the screen. However, due to the condensation of fiber inside the slurry, it brings great challenges to assess the fiber depositing on the screen. Hence, the modelling technique is a useful tool to capture the fiber deposition on the screen.

In the prior study, Lagomarsino, Pagonabarraga, and Lowe simulated an elastic filament in the fluid on a microscopic scale where the bending of the filament depended on the applied forces [7]. In addition, Delmotte, Climent, and Plouraboué employed the bead model to simulate a flexible fiber at low Reynolds number [8]. The model provided a flexible framework without adding artificial forces, and it is suitable for the simulation of micro-organisms, membranes, and micro-mechanics. Besides, Hosseini and Tafreshi solved a computational fluid dynamics (CFD) code with the discrete element method (DEM) to simulate the fiber collected on the filtration surface [9]. Moreover, Wang, Zhao, Wang, He, and Zheng employed the Lattice-Boltzmann method to simulate the two-phase model for the filtration process [10].

In this study, the multi-sphere model was employed to simulate the pulp fiber made from wood. Using a multi-sphere model allows for generating fibers in different geometries and different orientations to mimic the experiment as closely as possible. This study aims to adopt the simulation results to reveal the defects obtained in the experiment.

2. METHODOLOGY

In the DEM model, each fiber was simulated as a clump of particles. The particle within the same clump has no interaction; however, the interaction of particles belonging to different clumps was considered. The displacement and rotation of each clump were calculated based on the following equations [11]:

$$m_i \frac{dU_i}{dt} = m_i \mathbf{g} + \sum_j \mathbf{F}_{i,j} + \mathbf{F}_{i,f}, \quad (1)$$

$$\mathbf{I}_i \frac{d\boldsymbol{\omega}_i}{dt} + \boldsymbol{\omega}_i \times (\mathbf{I}_i \cdot \boldsymbol{\omega}_i) = \sum_j \mathbf{M}_{i,j} + \mathbf{M}_{i,f}, \quad (2)$$

where  $m_i$  is the mass of the clump  $i$ ,  $U_i$  is the velocity of the clump  $i$ ,  $\mathbf{g}$  is gravitational acceleration,  $F_{i,j}$  is the inter-particle contact force between sub-particles in clump  $i$  and  $j$ , and  $F_{i,f}$  is the fluid force on clump  $i$ .  $\mathbf{I}_i$  is the inertia tensor of clump  $i$ ,  $\boldsymbol{\omega}_i$  is the rotation velocity of clump  $i$ ,  $\mathbf{M}_{i,j}$  is the moment due to particle-particle interaction of those belong to clump  $j$  interacting on clump  $i$ , and  $\mathbf{M}_{i,f}$  is the moment of clump  $i$  due to the fluid force. In this simulation, a natural drying process due to the gravity force was considered, and the fluid force was not considered in the simulation result.

The fibers were simulated in three different orientation including horizontal, vertical, and titled angle to represent the fiber mixture inside the slurry as depicted in Figure 1.

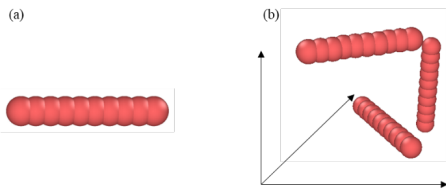


Figure 1. The morphology of (a) single fiber and (b) initial fiber orientation in the model

The simulation domain  $x \times y \times z = 10 \times 10 \times 22 \text{ mm}^3$  as shown in Figure 2. The simulated domain is sufficient

to represent the 3D molded fiber screen where the local area of the actual screen morphologies was fully captured. The same volume of fiber was set in the simulation. The material properties used in the simulation are listed in Table 1.

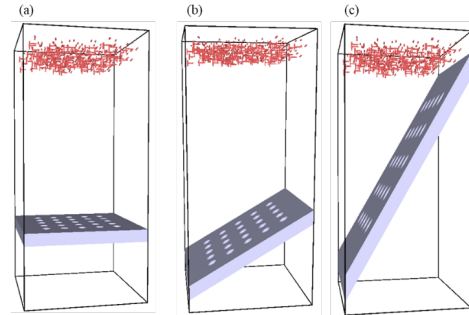


Figure 2. Same volume of fibers inserted in the domain to deposit on (a) flat screen, (b) 30°-tilted screen, and (c) 60°-tilted screen

In addition, an experiment was also performed to validate the simulation result. The experimental setup consisted of MJF 3D printed screens, a slurry solution, and a natural drying process. The MJF 3D printed screens were printed by HP Jet Fusion 5200 using polyamide-11 (PA11). The stereolithography (STL) files were designed in the same way as the ones used in the DEM model. The slurry solution consists of 0.5% wood fibers blended with water, and then 350 mL of slurry was poured on top of the screen, and due to the gravity, the fibers were collected on top of the screen. Followingly, these molded fiber samples were dried in the oven at a temperature of 80°C for 1 hour to fully dry the samples. The dried samples were used to validate the simulation results.

Table 1. Material properties using in the DEM model

Material properties	Value	Unit	Reference
Density of wet wood	1500	kg/m <sup>3</sup>	[12]
Youngs Modulus	0.01	GPa	[13]
Poisson Ratio	0.25	-	[14]
Coefficient Restitution	0.15	-	[15]

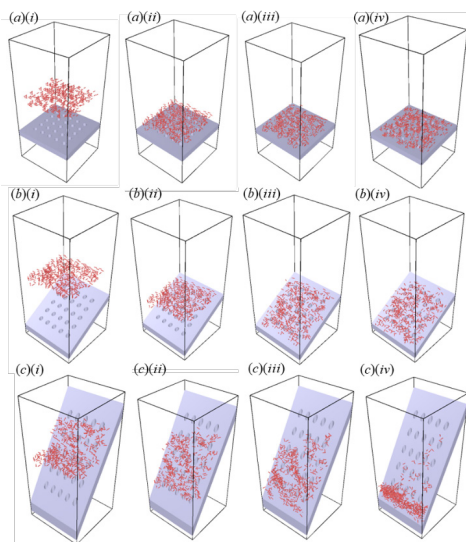
Coefficient Friction	0.11	-	[13]
----------------------	------	---	------

**3. SIMULATION RESULTS**

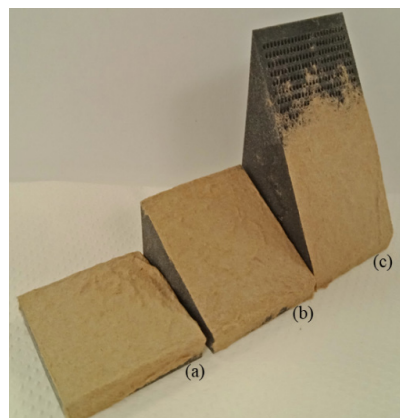
Figure 3 shows the fibers deposited on screens. As observed at 40 ms, the fibers started to touch the 60°-tilted screen and rolled on the screen, while the same fibers had not yet reached the flat screen or the 30°-tilted screen. In addition, when the fibers had settled down, the fiber deposition on the flat screen and on the 30°-tilted screen was more even as compared to that on the 60°-tilted screen. By employing the 60°-tilted screen, the fibers tended to slide to the lower area of the screen. This is due to the gravity force pulling the fibers to slide downward, leading to a large area of the screen in the upper area that was unable to capture the fibers, as shown in Figure 3(c)(iv).

In addition, the experiment was performed to validate the simulation result. It shows that for the flat screen and the 30°-tilted screen, the fibers were distributed and covered the whole screen area. However, for a 60°-tilted screen, a large area of the screen surface was not able to capture the fibers, and most of the fibers sank to the bottom area of the screen. Hence, it formed a large area where fibers were unable to cover. It explains the phenomena obtained in the simulation as shown in Figure 3, where the flat screen and the 30°-tilted screen were well covered by the fibers, while the 60°-tilted screen left a large area with no fibers remaining on the surface of the screen. Although there was a gap between the experiment and the simulation result. Firstly, it can be explained that the number of fibers in the experiment is hundreds of times larger than that in the simulation. Secondly, the DEM model was set up without the effect of surface tension and viscous force. However, the simulation results were obtained that reflect the

ones in the experiment. Therefore, the simulation model can be used to reveal the phenomena in the experiment.



**Figure 3.** Fiber deposition on (a) flat screen, (b) 30°-tilted screen, and (c) 60°-tilted screen at (i)  $t = 40$  ms, (ii)  $t = 50$  ms, (iii)  $t = 60$  ms, and (iv)  $t = 100$  ms.



**Figure 4.** Experimental fiber deposition on (a) flat screen, (b) 30°-tilted screen, and (c) 60°-tilted screen

**4. CONCLUSIONS**

The study investigated the fiber deposition on different molded fiber screens by simulation and experiment. It found that the tilted screen with a high tilted angle, such as 60°, containing a large area of the screen surface, was unable to collect the fibers. It was because

most of the fibers sank to the bottom of the screen. Hence, it explained the defect of uneven thickness of the highly tilted surface of the molded fiber product.

---

**REFERENCES**

[1]. Schneiderman, D. K., and Hillmyer, M. A. (2017). There is a great future in sustainable polymers, *Macromolecules*, vol. 50(10), p. 3733-3749.

[2]. Horejs, C. (2020). Solutions to plastic pollution, *Nature Reviews Materials*, vol. 5(9), p. 641-641.

[3]. Didone, M., and Tosello, G. (2019). Moulded pulp products manufacturing with thermoforming, *Packaging Technology and Science*, vol. 32(1), p. 7-22.

[4]. Keyes, F. E. (1890). Method of molding pulp articles, *Patent No. US424003A*.

[5]. Saxena, P., Bissacco, G., Meinert, K. A., Danielak, A. H., Ribó, M. M., and Pedersen, D. B. (2020). Soft tooling process chain for the manufacturing of micro-functional features on molds used for molding of paper bottles, *Journal of Manufacturing Processes*, vol. 54, p. 129-137.

[6]. Andersson, J, Vogt, U., and Pierce, D. A. (2010). Tool or tool part, system including such a tool or tool part, method of producing such a tool or tool part and method of molding a product from a pulp slurry, *Patent No. US20170370049A1*.

[7]. Lagomarsino, M. C., Pagonabarraga, I., and Lowe, C. P. (2005). Hydrodynamic induced deformation and orientation of a microscopic elastic filament, *Physical review letters*, vol. 94(14), p. 148104.

[8]. Delmotte, B., Climent, E., and Plouraboué, F. (2015). A general formulation of Bead Models applied

to flexible fibers and active filaments at low Reynolds number, *Journal of Computational Physics*, vol. 286, p. 14-37.

[9]. Hosseini, S. A., and Tafreshi, H. V. (2012). Modeling particle-loaded single fiber efficiency and fiber drag using ANSYS-Fluent CFD code, *Computers & Fluids*, vol. 66, p. 157-166.

[10]. Wang, H., Zhao, H., Wang, K., He, Y., and Zheng, C. (2013). Simulation of filtration process for multi-fiber filter using the Lattice-Boltzmann two-phase flow model, *Journal of Aerosol Science*, vol. 66, p. 164-178.

[11]. Shen, Z., Wang, G., Huang, D., and Jin, F. (2022). A resolved CFD-DEM coupling model for modeling two-phase fluids interaction with irregularly shaped particles, *Journal of Computational Physics*, vol. 448, p. 110695.

[12]. Kellogg, R. M., and Wangaard, F. F. (1969). Variation in the cell-wall density of wood, *Wood and Fiber Science*, no. 3, p. 180-204.

[13]. Guo, Y., Chen, Q., Xia, Y., Westover, T., Eksioglu, S., and Roni, M. (2020). Discrete element modeling of switchgrass particles under compression and rotational shear, *Biomass and Bioenergy*, vol. 141, p. 105649.

[14]. Sheng, Y., Wang, M., Zhang, L., and Ren, Q. (2022). Analysis of filtration process of 3-D mesh spacer filter by using CFD-DEM simulation, *Powder Technology*, vol. 396, Part B, p. 785-793.

[15]. Qian, F., Huang, N., Zhu, X., and Lu, J. (2013). Numerical study of the gas-solid flow characteristic of fibrous media based on SEM using CFD-DEM, *Powder technology*, vol. 249, p. 63-70.

# NGHIÊN CỨU TÍNH TOÁN, THIẾT KẾ VÀ CHẾ TẠO MÁY CẮT TIA NƯỚC CÔNG SUẤT NHỎ CHO VẬT LIỆU MỀM

Bùi Vũ Hùng, Bùi Văn Hưng\*

Đỗ Trung Hiếu, Đỗ Đức Long, Dương Công Trung

Khoa Cơ khí, Trường Đại học Giao Thông Vận Tải

\*Email: hungtkm@utc.edu.vn

## TÓM TẮT

Bài báo trình bày phương pháp tính toán, thiết kế và chế tạo mô hình máy cắt tia nước công suất nhỏ, hướng tới ứng dụng trong gia công các vật liệu mềm như mút xốp, cao su non và da nhân tạo. Trên cơ sở phân tích nguyên lý hình thành tia nước áp suất cao và đặc tính cơ học của vật liệu mềm, nghiên cứu đã xác định được dải thông số công nghệ tối ưu, bao gồm áp suất và lưu lượng tối thiểu cần thiết để đảm bảo khả năng cắt hiệu quả. Từ các thông số này, một mô hình máy cắt quy mô nhỏ đã được thiết kế với các thành phần chính gồm: bơm áp suất cao mini, đầu cắt đường kính nhỏ nhằm tập trung năng lượng dòng tia, và cơ cấu dẫn hướng có độ chính xác cao để đảm bảo quỹ đạo cắt ổn định. Kết quả mô phỏng CFD kết hợp với thực nghiệm cho thấy hệ thống có khả năng tạo ra tia nước đủ năng lượng để cắt xuyên vật liệu mềm với bề mặt cắt sạch, ít biến dạng và độ ổn định cao. Mô hình không chỉ đáp ứng yêu cầu nghiên cứu, đào tạo mà còn có tiềm năng ứng dụng trong sản xuất quy mô nhỏ và các hệ thống gia công linh hoạt chi phí thấp.

**Từ khóa:** máy cắt tia nước, cắt vật liệu mềm, gia công CNC, nước áp suất cao, thiết kế máy.

## ABSTRACT

This study presents the calculation, design, and development of a low-power waterjet cutting system intended for processing soft materials such as foam, soft rubber, and synthetic leather. Based on an analysis of high-pressure waterjet formation and the mechanical characteristics of soft materials, the research determines the minimum required pressure and flow rate necessary to achieve effective cutting performance. A compact prototype system is subsequently designed and fabricated, consisting of a mini high-pressure pump, a small-diameter nozzle for jet concentration, and a precision guiding mechanism to ensure stable cutting trajectories. Computational fluid dynamics (CFD) simulations, combined with experimental validation, are conducted to evaluate the performance of the system. The results demonstrate that the proposed model is capable of generating a stable and sufficiently energetic waterjet to produce clean and consistent cutting surfaces with minimal deformation. The system shows strong potential for application in educational environments, laboratory-scale research, and small-scale manufacturing due to its cost-effectiveness and operational simplicity.

**Keywords:** waterjet cutting machine, cutting soft materials, CNC machining, high-pressure water, machine design.

Ngày nhận bài: 14/01/2026; Ngày sửa bài: 22/03/2026; Ngày duyệt đăng: 02/04/2026

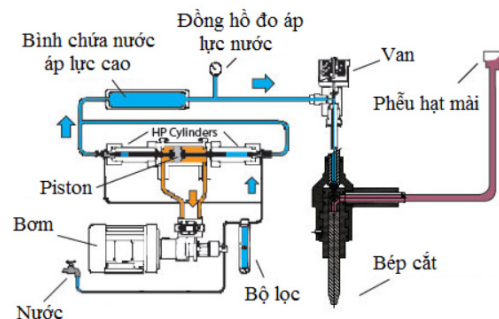
## 1. GIỚI THIỆU CHUNG

Công nghệ gia công bằng tia nước (Water Jet Machining – WJM) hiện được ứng dụng rộng rãi trong lĩnh vực gia công cắt kim loại tấm nhờ khả năng xử lý linh hoạt đa dạng vật liệu. Công nghệ này cho phép cắt hiệu quả từ kim loại và hợp kim (như thép không gỉ, thép carbon, đồng, nhôm...) đến các vật liệu phi kim như nhựa, gỗ, gốm, đá, kính và vật liệu composite, bằng cách sử dụng tia nước áp suất siêu cao hoặc hỗn hợp nước – hạt mài, áp suất cực đại của tia nước gia tốc tập trung lên một vùng diện tích rất nhỏ, với chiều rộng rãnh cắt khoảng 1 mm và đường kính lỗ nhỏ nhất có thể cắt đạt khoảng 1.5 mm [1, 2].

Quá trình gia công bằng tia nước được thực hiện bằng cách ép một lưu lượng nước lớn đi qua một ống có tiết diện rất nhỏ, làm cho vận tốc của dòng nước tăng lên nhanh chóng theo nguyên lý bảo toàn lưu lượng. Khi dòng nước đã được tăng tốc rời khỏi đầu phun với vận tốc điển hình từ 200 đến 500 m/s, nó tạo ra một lực cắt lớn tác động lên bề mặt vật liệu gia công [2]. Một ưu điểm quan trọng của WJM là quá trình cắt không sinh nhiệt đáng kể, do đó không làm thay đổi cấu trúc vi mô hay tính chất cơ học của vật liệu, đặc biệt phù hợp đối với các vật liệu nhạy với vùng ảnh hưởng nhiệt [1,3]. Dưới tác dụng của áp suất cao thường trong khoảng  $4 \times 10^8$  đến  $6 \times 10^8$  Pa các vi nứt được hình thành tại bề mặt vật liệu và tiếp tục phát triển dưới tác động kết hợp của áp suất thủy lực và động năng của các phân tử nước. Dòng nước đồng thời cuốn trôi vật liệu bị bóc tách ra khỏi vùng gia công, góp phần mở rộng và kéo dài vết nứt cho đến khi vật liệu bị cắt rời hoàn toàn [2, 3].

Nhờ những đặc tính này, công nghệ cắt tia nước được ứng dụng trong nhiều ngành công nghiệp như cơ khí chế tạo, xây dựng và hàng không vũ trụ để thực hiện các công đoạn cắt biên dạng, tạo hình, tách phôi và khoan lỗ. Hệ thống

cắt tia nước có thể gia công kim loại với chiều dày lớn (lên tới khoảng 150 mm tùy thuộc vào điều kiện công nghệ), đồng thời tạo ra vết cắt có độ chính xác và chất lượng bề mặt cao, vượt trội so với các phương pháp cắt plasma hoặc cắt khí oxy–gas [4].



Hình 1. Sơ đồ nguyên lý máy cắt tia nước [5]

Có nhiều nghiên cứu quốc tế nổi bật chủ yếu tập trung vào hệ thống cắt tia nước áp suất cao và ứng dụng trong gia công vật liệu kim loại hoặc vật liệu cứng ở quy mô công nghiệp [1, 2]. Các nghiên cứu này ít đề cập đến việc phát triển các hệ thống công suất nhỏ, chi phí thấp hoặc tối ưu hóa cho vật liệu mềm. Tại Việt Nam, công nghệ gia công bằng tia nước đã được ứng dụng chủ yếu trong các ngành cơ khí chế tạo, vật liệu xây dựng và gia công đá – kính thông qua các hệ thống máy cắt CNC nhập khẩu từ châu Âu và Trung Quốc. Tuy nhiên, có thể nhận thấy rằng phần lớn các nghiên cứu đều tập trung vào vật liệu kim loại và hợp kim, đồng thời sử dụng thiết bị cắt tia nước công suất lớn vốn được thiết kế cho mục đích công nghiệp. Việc làm chủ thiết bị trong nước còn hạn chế, nên các nghiên cứu chuyên sâu về tối ưu thông số công nghệ, mô phỏng và dự đoán đặc trưng vết cắt bằng tia nước chỉ mới được triển khai trong khoảng hơn một thập kỷ gần đây. Một số nhóm nghiên cứu đã tiến hành khảo sát ảnh hưởng của áp suất, tốc độ cắt, lưu lượng và kích thước hạt mài

đến chiều sâu cắt, độ nhám bề mặt và hình dạng rãnh khi gia công bằng tia nước có hạt mài, chủ yếu trên vật liệu hợp kim nhôm và thép [6]. Các công trình này góp phần bước đầu xây dựng mô hình thực nghiệm và đánh giá khả năng ứng dụng công nghệ cắt tia nước trong điều kiện thiết bị sẵn có tại Việt Nam.

Các công trình liên quan đến máy cắt tia nước công suất nhỏ, thiết bị giá rẻ hoặc ứng dụng cho vật liệu mềm gần như chưa được triển khai rộng rãi. Điều này dẫn tới khoảng trống nghiên cứu trong lĩnh vực thiết kế – phát triển mô hình máy cắt tia nước phục vụ đào tạo, nghiên cứu ứng dụng và sản xuất quy mô nhỏ, đặc biệt đối với các vật liệu mềm như mút xốp, cao su non hay da nhân tạo. Vì vậy, hướng nghiên cứu về tính toán, thiết kế và chế tạo máy cắt tia nước công suất nhỏ cho vật liệu mềm vẫn còn nhiều tiềm năng và có ý nghĩa thực tiễn đối với môi trường đào tạo kỹ thuật và nhu cầu sản xuất trong nước.

**2. PHƯƠNG PHÁP TÍNH TOÁN – THIẾT KẾ**

**2.1. Xác định yêu cầu kỹ thuật**

Trong tính toán và thiết kế mô hình máy cắt tia nước, việc đầu tiên là xác định các thông số đầu vào làm cơ sở cho mô hình tính toán và lựa chọn chế độ làm việc. Các thông số gồm đặc tính cơ lý của vật liệu Fomex, đặc trưng của chất lỏng cắt (nước) và yêu cầu về bề rộng, độ sâu vết cắt của sản phẩm. Đây là cơ sở để xác định áp suất, lưu lượng và các thông số chính của hệ thống cắt tia nước công suất nhỏ. Bảng 1 trình bày các thông số được sử dụng làm cơ sở cho tính toán thiết kế.

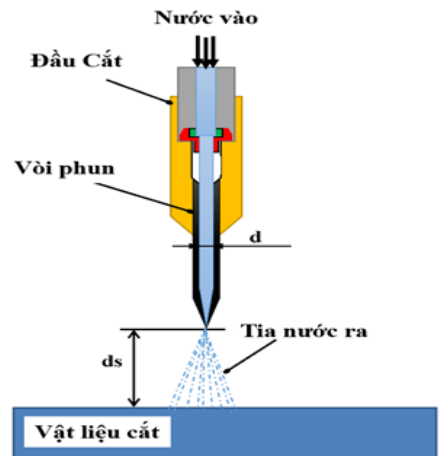
*Bảng 1. Các thông số thiết kế*

STT	Tên	Thông số xác định	Kí hiệu	Giá trị
1	Loại vật liệu: Fomex	Ứng suất bền	$\sigma_{\text{bền}}$	15 MPa

2	Chất lưu để cắt vật liệu: Nước	Tỉ khối của nước	$\rho$	1000 kg/m <sup>3</sup>
3	Yêu cầu sản phẩm	Bề rộng vết cắt	a	0.3 mm
		Độ sâu vết cắt	h	2 mm

**2.2. Tính toán thông số làm việc**

**a. Xác định áp suất cắt cần thiết ( $P_{\text{ct}}$ )**



*Hình 2. Đầu cắt tia nước hạt mài*

Trong trường hợp chất lỏng lí tưởng, kích thước lỗ đường kính đầu phun có thể lấy gần đúng  $d = 0.3 \text{ mm}$  [7]. Khoảng cách từ đầu phun đến bề mặt gia công là được xác định từ công thức:

$$d_s = Xc_{\text{Max}} = 135 \times d = 135 \times 0.27 = 36.45 \text{ mm.}$$

Điều kiện để máy cắt tia nước có thể cắt được vật liệu là: Công suất của dòng tia nước trộn hạt mài cần phải đủ lớn để thắng sức kháng cắt của vật liệu. Nếu gọi là ứng suất cần thiết để bào mòn và cắt vật liệu thì điều kiện cắt là:

$$P_{\text{cut}} \geq \sigma_{\text{cut}} \cdot A \cdot v_{\text{feed}} \tag{1}$$

Trong đó:  $P_{\text{cut}}$ : Công suất cắt do dòng nước cung cấp (W)

$\sigma_{\text{cut}}$ : Ứng suất kháng cắt của vật liệu (Pa)

A: Diện tích mặt cắt trên vật liệu (m<sup>2</sup>)

$v_{\text{feed}}$ : Tốc độ di chuyển của đầu cắt (m/s)

Công suất cắt của dòng nước biểu diễn qua tổng động năng của nước theo công thức sau:

$$P_{cut} = \frac{E_{water}}{t} \quad (2)$$

Trong đó:  $E_{water}$  là động năng của dòng nước và  $t$  là thời gian cắt (s)

Áp dụng phương trình Bernoulli ta xác định được vận tốc của dòng nước được xác định theo công thức sau (3):

$$v_{water} = \sqrt{\frac{2 \cdot p_{ct}}{\rho}} = \sqrt{\frac{2 \cdot p_{ct}}{1000}} = \sqrt{\frac{p_{ct}}{500}} \quad (3)$$

Trong đó:  $p$ : áp suất (Pa)

Động năng của dòng nước thuần khiết được xác định theo công thức (4):

$$E_{water} = \frac{1}{2} \cdot m_{water} \cdot v_{water}^2 \quad (4)$$

Thay các giá trị vào phương trình (4), ta có:

$$E_{water} = \frac{1}{2} \cdot \rho \cdot Q \cdot t \cdot v_{water}^2 = \frac{1}{2} \cdot \rho \cdot A \cdot t \cdot v_{water}^3$$

$$\frac{1}{2} \cdot 1000 \times 5.7 \times 10^{-8} \cdot \left( \sqrt{\frac{p}{500}} \right)^3 \cdot t$$

$$E_{water} = 2.55 \times 10^{-9} \times p^{\frac{3}{2}} \cdot x \cdot t \quad (5)$$

Diện tích mặt cắt trên vật liệu cần cắt được xác định theo công thức sau:

$$A = w \cdot T = 0.3 \times 10^{-3} \times 2 \times 10^{-3} = 6 \times 10^{-7} (m^2) \quad (6)$$

Trong đó:  $w$ : bề rộng vết cắt (mm),  $T$ : độ dày vật liệu cần cắt (mm).

Điều kiện cắt được vật liệu là:

$$P_{cut} \geq \sigma_{cut} \cdot A \cdot v_{feed} \quad (7)$$

Chọn tốc độ di chuyển của đầu cắt là  $v_{feed} = 2.5 (m/s)$  Từ các phương trình (3), (4), (5), (6), (7) xác định được áp suất cần thiết để cắt được vật liệu Fomex có giá trị  $\sigma_{bền} = 15 \text{ MPa}$  là:

$$p_{ct} \geq 4270003 (Pa)$$

Lựa chọn áp suất cần thiết:

$$p_{ct} = 4.5 (MPa).$$

Trong đó:  $p_{ct}$ : áp suất cần thiết để cắt được vật liệu (Mpa)

### b. Tính toán công suất máy bơm

Lưu lượng nước chảy qua đầu cắt trong một đơn vị thời gian được xác định như sau:

$$Q = A \cdot v \quad (8)$$

Trong đó:  $A$  là diện tích tiết diện lỗ phun tia nước ( $d = 0.3 \text{ mm}$ );  $v$  là tốc độ của tia nước (m/s).

Tốc độ của nước được xác định như sau:

$$v = \sqrt{\frac{2 \cdot p_{ct}}{\rho}} = \sqrt{\frac{2 \times 4.5 \times 10^6}{1000}}$$

$$= 94.86 (m/s)$$

Lưu lượng nước được xác định:

$$Q = A \cdot v = 7.07 \times 10^{-8} \times 94.86$$

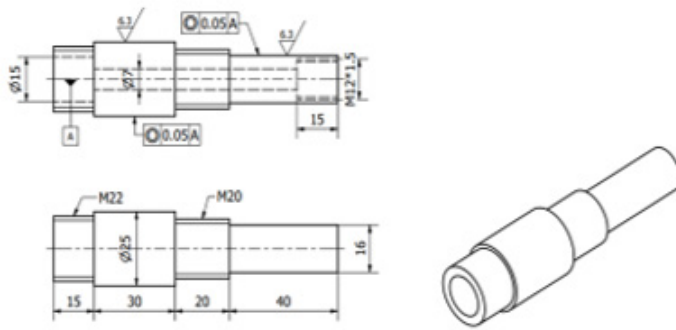
$$= 6.7 \times 10^{-6} (m^3/s)$$

Với áp suất cần thiết  $p_{ct} = 4.5 \times 10^6 (Pa)$ , công suất của bơm được tính theo công thức (9):

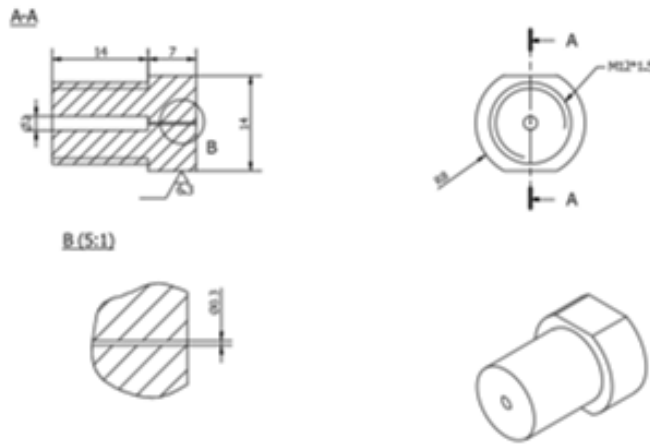
$$P = \frac{p_{ct} \cdot Q}{\eta} = \frac{4.5 \times 10^6 \times 6.7 \times 10^{-6}}{0.9} = 33.55 (W) \quad (9)$$

### 2.3. Thiết kế và chế tạo mô hình máy

Trong hệ thống máy cắt tia nước, đầu phun là chi tiết làm việc trực tiếp với dòng nước áp suất cao, quyết định đến hình dạng, tốc độ và năng lượng của tia cắt, do đó ảnh hưởng lớn tới chất lượng bề mặt và độ chính xác đường cắt. Vì vậy, việc thiết kế đầu phun cần được xem xét cẩn thận về kích thước lỗ phun, chiều dài dẫn hướng, kiểu ren lắp ghép, vật liệu chế tạo và khả năng thay thế, bảo dưỡng. Trên cơ sở yêu cầu kỹ thuật của mô hình máy và điều kiện gia công thực tế, phần này tiến hành xây dựng phương án thiết kế đầu phun, lựa chọn các thông số kết cấu chủ yếu và thể hiện trên bản vẽ chi tiết phục vụ chế tạo (Hình 3a).



a) Thân đầu phun



b) Đầu phun

**Hình 3.** Bản vẽ của cụm đầu phun: a) Thân đầu phun

Kết quả chế tạo của cụm đầu phun được thể hiện trong Hình 4.



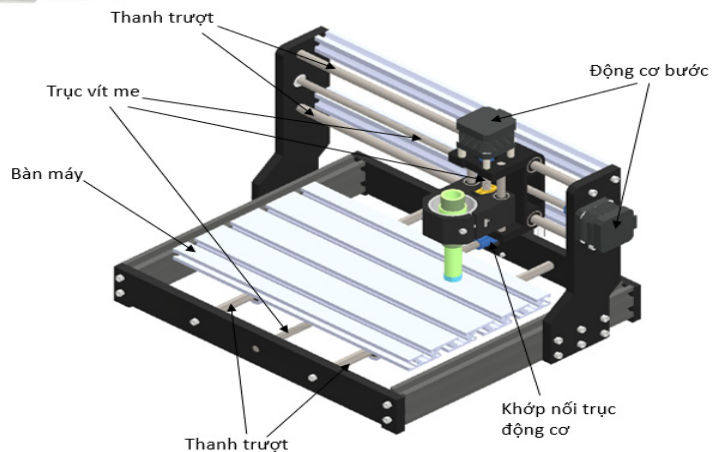
**Hình 4.** Chi tiết đầu phun sau khi gia công

**2.4. Thiết kế mô hình máy cắt tia nước công suất nhỏ**

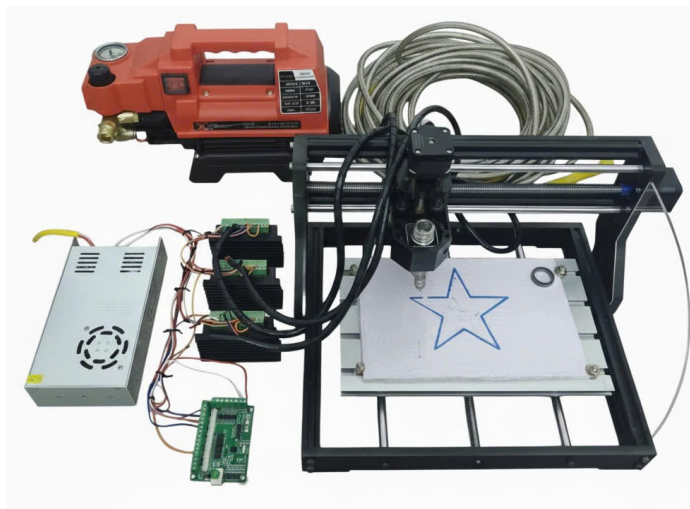
Để hiện thực hóa mô hình máy cắt tia nước công suất nhỏ đã đề xuất, trước hết cần lựa chọn phương án thiết kế và chế tạo hợp lý cho kết cấu cơ khí và hệ thống truyền

động của máy. Phương án này phải đảm bảo các yêu cầu cơ bản về độ cứng vững, độ chính xác chuyển động và hành trình làm việc phù hợp với kích thước chi tiết gia công, đồng thời đảm bảo tính đơn giản trong gia công – lắp ráp và chi phí chế tạo thấp, phù hợp với điều kiện trang thiết bị hiện có tại cơ sở đào tạo. Các nguyên tắc này được tham khảo từ các tài liệu về công nghệ chế tạo máy và thiết kế hệ thống CNC [8-10]. Trên cơ sở đó, nghiên cứu tiến hành phân tích và lựa chọn phương án thiết kế tối ưu cho mô hình máy cắt tia nước, đồng thời xem xét các yếu tố liên quan đến dòng chảy và đặc tính của tia nước dựa trên cơ sở lý thuyết cơ học lưu chất [11]. Kết quả thiết kế mô hình máy cắt tia nước công suất nhỏ sử dụng phần mềm

Autodesk Inventor được thể hiện trong **Hình 5a**. Kết quả chế tạo máy cắt tia nước công suất nhỏ được thể hiện trong **Hình 5b**



a)

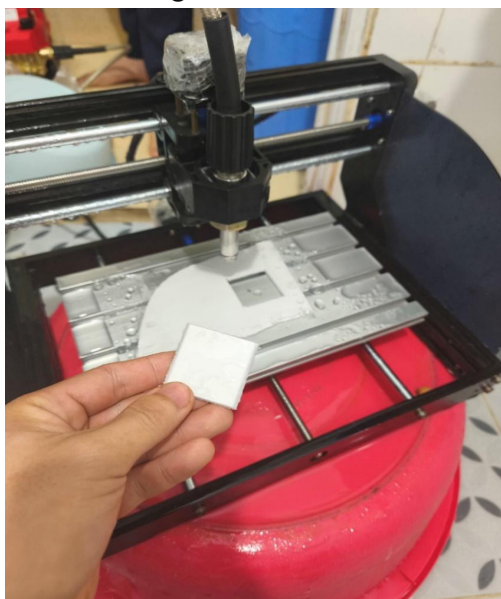


b)

**Hình 5.** Mô hình 3D máy cắt tia nước: a) Mô hình 3D thiết kế trên phần mềm CAD;

#### b) Mô hình máy sau khi chế tạo THỬ NGHIỆM VÀ ĐÁNH GIÁ

Để thực nghiệm đánh giá khả năng công nghệ của máy cắt tia nước công suất nhỏ, nhóm nghiên cứu sử dụng vật liệu Fomex có chiều dày 2 mm với tiết diện cắt là hình vuông có kích thước 20mm x 20 mm. Kết quả thực nghiệm được thể hiện như trong **Hình 6**.



**Hình 6.** Sản phẩm sau khi cắt

Sau khi tiến hành cắt thử nghiệm trên vật liệu Fomex, có thể nhận thấy máy hoạt động ổn định, đường chạy dao bám đúng quỹ đạo đã lập trình. Sản phẩm sau cắt cho biên dạng khá chính xác, các cạnh vuông góc rõ ràng, không xuất hiện hiện tượng cháy xém, biến dạng nhiệt hay cong vênh tấm – đây là ưu điểm nổi bật so với các phương pháp cắt sinh nhiệt. Bề mặt vết cắt tương đối phẳng, độ sứt mẻ mép nhỏ, lớp

bọt bên trong tấm Fomex không bị bóc tách nhiều. Kích thước chi tiết đo được sai lệch không lớn so với kích thước thiết kế, nằm trong phạm vi chấp nhận được đối với mô hình máy công suất nhỏ. Tuy nhiên, tại một số vị trí góc chuyển hướng và đoạn bắt đầu/kết thúc đường cắt vẫn còn dấu hiệu rung nhẹ, bề mặt chưa thật “sắc”, cho thấy cần tiếp tục tối ưu các thông số như áp suất nước, tốc độ tiến dao và khoảng cách từ đầu phun tới bề mặt vật liệu để nâng cao hơn nữa chất lượng vết cắt.

#### KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN

Nghiên cứu đã thực hiện việc tính toán, thiết kế, chế tạo và thử nghiệm mô hình máy cắt tia nước công suất nhỏ nhằm gia công các vật liệu mềm. Kết quả cho thấy mô hình máy CNC sử dụng cơ cấu truyền động vít me – động cơ bước theo ba trục X, Y, Z có khả năng làm việc ổn định, đáp ứng được yêu cầu cơ bản về độ chính xác hình học và chất lượng đường cắt. Thử nghiệm trên vật liệu Fomex cho thấy máy có thể tạo ra biên dạng cắt bám sát quỹ đạo lập trình,

mép cắt tương đối phẳng, bề mặt vật liệu không xuất hiện ảnh hưởng nhiệt hay hư hại đáng kể, qua đó khẳng định tính khả thi của phương pháp cắt tia nước đối với nhóm vật liệu mềm.

Bên cạnh đó, kết quả nghiên cứu cũng chỉ ra rằng các thông số công nghệ như áp suất nước, tốc độ tiến dao và khoảng cách từ đầu phun đến bề mặt vật liệu có ảnh hưởng rõ rệt đến chất lượng bề mặt cắt và độ chính xác kích thước. Điều này cho thấy việc lựa chọn và tối ưu hóa chế độ cắt là yếu tố quan trọng nhằm nâng cao hiệu quả gia công và mở rộng khả năng ứng dụng của hệ thống.

Trong thời gian tới, nghiên cứu có thể được tiếp tục theo hướng hoàn thiện hệ thống điều khiển và kết cấu cơ khí của máy, đặc biệt là tăng độ cứng vững của khung và bàn máy, cải tiến cơ cấu dẫn hướng và truyền động nhằm hạn chế hiện tượng trượt bước, rung động và sai số vị trí trong quá trình gia công. Đồng thời, cần mở rộng thực nghiệm trên nhiều loại vật liệu mềm khác như cao su, da, vải và xốp, cũng như xây dựng mô hình tối ưu hóa thông số công nghệ để xác lập chế độ cắt phù hợp cho từng loại vật liệu. Đây sẽ là cơ sở quan trọng để nâng cao tính ứng dụng thực tiễn của mô hình trong đào tạo, nghiên cứu và sản xuất quy mô nhỏ.

---

### TÀI LIỆU THAM KHẢO

[1]. Hashish, M. (1989). *Machining with Abrasive Waterjets*. Journal of Engineering Materials and Technology.

[2]. Momber, A. W., & Kovacevic, R. (1998). *Principles of Abrasive Waterjet Machining*. Springer.

[3]. Kulekci, M. (2009). *Processes and apparatus developments in industrial waterjet applications*. International Journal of Machine Tools & Manufacture.

[4]. Maurel, B., et al. (2014). *High-pressure waterjet cutting of metals*. Journal of Manufacturing Processes.

[5]. Website : <https://smartsheetmetal.com.vn/tin-tuc/tim-hieu-ve-nguyen-ly-hoat-dong-cua-may-cat-kim-loai-bang-nuoc.html>

[6]. Bùi Văn hưng, Ngô Anh Vũ & Khuong Quang Sơn (2022). *Xây dựng mô hình thực nghiệm dự đoán chiều sâu phay bằng tia nước chứa hạt mài trên hợp kim nhôm*. Tạp chí KHCN, ĐH Công nghiệp Hà Nội, website: <https://doi.org/10.57001/huinh5804.90>.

[7]. David A. Summer (2003). *Waterjetting Technology*, E&FN Spon, An Imprint of Chapman & Hall.

[8]. Trần Văn Địch (2004). *Giáo trình Công nghệ CNC*, NXB khoa học và kỹ thuật Hà nội.

[9]. Nguyễn Đắc Lộc (2007). *Sổ tay công nghệ chế tạo máy tập 1*, NXB Khoa học kỹ thuật.

[10]. Nguyễn Đắc Lộc (2005). *Sổ tay công nghệ chế tạo máy tập 2*, NXB Khoa học kỹ thuật.

[11]. Phạm Văn Vĩnh (1994). *Cơ học lưu chất ứng dụng*, Trường ĐH Giao thông vận tải Hà Nội.

# ĐỀ XUẤT ỨNG DỤNG CÔNG NGHỆ TRÍ TUỆ NHÂN TẠO (AI) TRONG HƯỚNG DẪN THỰC HÀNH HỌC PHẦN AUTOCAD CHO SINH VIÊN TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á

**Bùi Quang Toàn**

ThS. Khoa Cơ khí, Trường Đại học Công nghệ Đông Á

Email: toanbq@eaut.edu.vn

## TÓM TẮT

Bài báo tập trung nghiên cứu và đề xuất giải pháp ứng dụng công nghệ Trí tuệ nhân tạo (AI) vào quá trình giảng dạy thực hành học phần AutoCAD cho sinh viên Trường Đại học Công nghệ Đông Á. Tác giả đề xuất triển khai trợ giảng AI dưới các hình thức như: trợ lý ảo học tập sử dụng chatbot AI, thư viện video mô phỏng thao tác AutoCAD, phần mềm theo dõi và đánh giá thao tác người học, hệ thống giám sát lớp học bằng AI thị giác máy tính, và môi trường học tập trực tuyến tích hợp AI chấm bài tự động. Các giải pháp này nhằm nâng cao hiệu quả giảng dạy, tăng tính cá nhân hóa trong hỗ trợ sinh viên, đồng thời giúp giảm tải cho giảng viên khi lớp học đông sinh viên. Nghiên cứu cũng đưa ra lộ trình triển khai theo giai đoạn và nhấn mạnh tiềm năng của AI trong cải thiện chất lượng đào tạo thực hành kỹ năng phần mềm kỹ thuật trong giáo dục đại học.

**Từ khóa:** Trí tuệ nhân tạo, AutoCAD, trợ giảng AI, tương tác, thao tác.

## ABSTRACT

The paper focuses on investigating and proposing solutions for integrating Artificial Intelligence (AI) technologies into the practical teaching of the AutoCAD course for students at East Asia University of Technology. The author proposes implementing AI-powered teaching assistants in various forms, including: AI-based learning chatbots, a video library simulating AutoCAD operations, software for tracking and evaluating students' actions, AI-driven classroom monitoring systems using computer vision, and an online learning environment integrated with automated AI-based grading tools. These solutions aim to enhance teaching effectiveness, increase the level of personalized support for students, and reduce the workload for instructors in large classes. The study also outlines a phased implementation roadmap and highlights the potential of AI in improving the quality of practical training in technical software within higher education.

**Keywords:** Artificial Intelligence, AutoCAD, AI teaching assistant, interaction, operation.

Ngày nhận bài: 20/06/2025; Ngày sửa bài: 14/11/2025; Ngày duyệt đăng: 02/04/2026

**1. Đặt vấn đề**

Tại Trường Đại học Công nghệ Đông Á, học phần AutoCAD là môn thực hành quan trọng trong chương trình đào tạo kỹ thuật. Tuy nhiên, việc giảng dạy đang gặp nhiều khó khăn như sĩ số lớp đông (trung bình 40-60 sinh viên/lớp), chất lượng đầu vào chưa đồng đều, sinh viên tiếp thu chậm, tính tự học thấp. Ngoài ra, hệ thống máy tính chưa đồng bộ, lớp học rộng khiến giảng viên khó quan sát và hỗ trợ sinh viên kịp thời.

Khảo sát nội bộ tại Khoa Cơ khí cho thấy hơn 60% sinh viên gặp khó khăn khi thao tác phần mềm, thường xuyên chậm tiến độ bài thực hành. Trong khi đó, giảng viên không thể theo sát từng cá nhân, ảnh hưởng đến chất lượng dạy - học.

Trước thực trạng này, việc ứng dụng trí tuệ nhân tạo (AI) như một công cụ trợ giảng là giải pháp khả thi, giúp hỗ trợ sinh viên theo hướng cá nhân hóa, đồng thời giảm tải cho giảng viên. Bài báo đề xuất các giải pháp ứng dụng AI trong giảng dạy AutoCAD nhằm nâng cao hiệu quả thực hành và thúc đẩy quá trình chuyển đổi số trong đào tạo kỹ thuật.

**2. Giải quyết vấn đề**

**2.1. Cơ sở lý thuyết**

**2.1.1. Trí tuệ nhân tạo trong giáo dục (AI in Education)**

Theo nghiên cứu của Woolf (2010) và các báo cáo của UNESCO [7], AI trong giáo dục có thể hỗ trợ cá nhân hóa học tập, cung cấp phản hồi tức thời, theo dõi tiến trình người học và hỗ trợ giảng viên ra quyết định. Trong giảng dạy thực hành kỹ thuật, AI có thể đóng vai trò như “trợ lý ảo”, giúp sinh viên truy xuất thông tin, hướng dẫn thao tác và giải đáp thắc mắc mà không cần chờ giảng viên.

**2.1.2. Mô hình trợ giảng ảo (Intelligent Tutoring Systems – ITS)**

Hệ thống trợ giảng

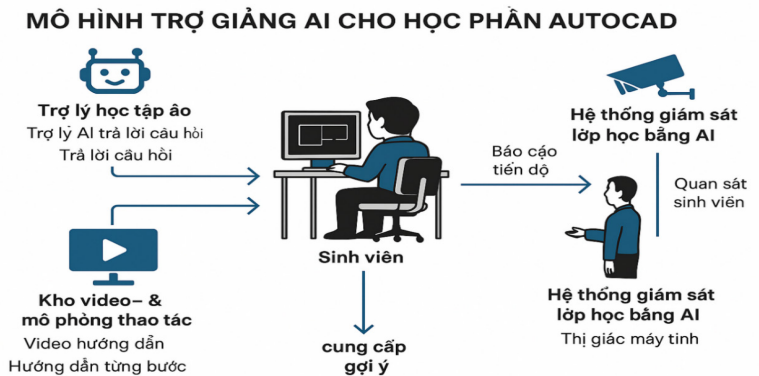
thông minh là các phần mềm sử dụng AI để mô phỏng vai trò người hướng dẫn. Các hệ thống này có thể phân tích thao tác người học, nhận diện lỗi sai và đưa ra gợi ý sửa lỗi theo thời gian thực [6]. Trong môi trường AutoCAD, ITS có thể theo dõi thao tác vẽ, cảnh báo khi sinh viên sử dụng sai lệnh hoặc sai lớp (layer), giúp cải thiện hiệu quả thực hành [8].

**2.1.3. Giao diện tương tác tự nhiên và học tập tích cực**

Sự phát triển của các mô hình ngôn ngữ lớn (LLMs) như ChatGPT, Claude, Gemini... cho phép xây dựng các trợ lý học tập tương tác qua ngôn ngữ tự nhiên. Điều này đặc biệt phù hợp với sinh viên có khả năng tự học còn hạn chế, giúp họ dễ tiếp cận kiến thức hơn qua việc đặt câu hỏi và nhận hướng dẫn ngắn gọn, trực tiếp từ AI [4,5].

Việc ứng dụng trí tuệ nhân tạo (AI) trong giáo dục đã trở thành xu hướng tất yếu trong quá trình chuyển đổi số, đặc biệt ở các môn học thực hành kỹ thuật đòi hỏi thao tác phần mềm như AutoCAD. Dựa theo các nền tảng lý thuyết về trí tuệ nhân tạo, hệ thống trợ giảng thông minh và giao diện học tập tương tác đang tạo điều kiện thuận lợi để ứng dụng AI vào hỗ trợ dạy học AutoCAD [1]. Việc tận dụng các nền tảng này không chỉ giúp khắc phục những bất cập trong giảng dạy truyền thống mà còn tạo bước đột phá trong đào tạo thực hành kỹ thuật.

**2.2. Mô hình đề xuất**



**Hình 1.** Mô hình trợ giảng AI cho học phần Autocad

2.2.1. Trợ lý học tập ảo – AI Learning Assistant

\* **Dạng:** Chatbot AI (ứng dụng LLM như ChatGPT hoặc Claude). [4]

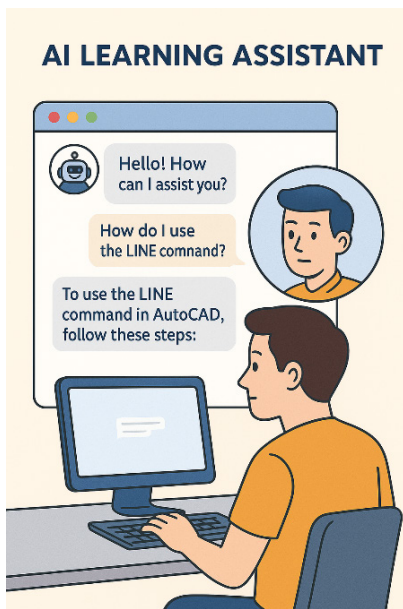
\* **Chức năng:**

- Trả lời các câu hỏi về lệnh AutoCAD, lỗi thao tác, quy trình thực hành.

- Gợi ý theo ngữ cảnh, giúp sinh viên tự học và tra cứu trong quá trình thực hành.

Ví dụ: Sinh viên có thể hỏi các câu như: “Lệnh vẽ hình chữ nhật?”, “Cách dùng lệnh OFFSET?”, “Lỗi không hiện lưới phải làm sao?”, v.v.

\* **Triển khai:** Nhúng trong giao diện LMS, Zalo OA, Google Chat, hoặc nền tảng web nội bộ.



Hình 2. Mô hình trợ lý học tập ảo

2.2.2. Kho học liệu mô phỏng – Video & Module thao tác

\* **Nội dung:** Các video ngắn hướng dẫn từng lệnh, ví dụ thực hành theo bước.

\* **Tích hợp:**

- Video 2D/3D mô phỏng trực quan thao tác AutoCAD.

- Liên kết từ chatbot hoặc LMS khi

sinh viên cần minh họa thao tác.

- Hỗ trợ dưới dạng pop-up hướng dẫn ngay trong phần mềm AutoCAD, có thể chạy song song.

\* Giúp giảm phụ thuộc vào giảng viên khi cần hỗ trợ kỹ thuật cơ bản.



Hình 3. Mô hình kho học liệu mô phỏng

2.2.3. Công cụ theo dõi & phân tích thao tác AutoCAD – AI AutoCAD Tracker

\* **Cơ chế:** Plugin chạy nền trong AutoCAD ghi nhận thao tác của sinh viên.

\* **Tính năng:**

- Phân tích thao tác đúng/sai theo bài mẫu.

- Phát hiện lỗi phổ biến (vẽ sai layer, không ghi kích thước, bố cục sai tỷ lệ...).

\* **Xuất báo cáo** cho giảng viên theo từng sinh viên, từng buổi học: Giảng viên có thể nhận được báo cáo tổng hợp cuối buổi (ai làm tốt, ai mắc lỗi gì, ai không thao tác gì cả).



Hình 4. Mô hình công cụ theo dõi và phân tích thao tác

## Cơ khí - Động lực

### 2.2.4. Hệ thống phân tích lớp học – AI Classroom Monitor (tùy chọn mở rộng)

\* Ứng dụng thị giác máy tính (Computer Vision) qua camera lớp.

\* Chức năng: Phát hiện sinh viên không hoạt động, mất tập trung, hoặc giờ tay yêu cầu hỗ trợ.

\* Giúp giảng viên phản ứng nhanh hơn với các tình huống trong lớp đông.

### 2.2.5. Môi trường học tập trực tuyến tích hợp AI

Khi sinh viên học tại nhà, hệ thống có thể:

- Giao bài tập AutoCAD theo cấp độ.
- AI tự động chấm bài bằng cách so sánh bản vẽ với đáp án mẫu (dựa trên nhận dạng hình học, layer, tỷ lệ...).
- Hướng dẫn lỗi nếu bản vẽ sai (VD: “bạn đặt khung sai tỷ lệ 1:100, nên sửa lại...”).

### 2.3. Quy trình vận hành tổng thể của mô hình:



Hình 5. Quy trình vận hành tổng thể của mô hình

#### 2.3.1. Ưu điểm

- **Tăng khả năng cá nhân hóa học tập:** Hệ thống AI cho phép sinh viên nhận hỗ trợ theo nhu cầu cá nhân, phù hợp với tốc độ tiếp thu và năng lực mỗi người.

- **Giảm tải cho giảng viên:** AI hỗ trợ xử lý câu hỏi thường gặp, phát hiện lỗi thao tác cơ bản, giúp giảng viên tập trung vào phần chuyên sâu và điều phối lớp học hiệu quả hơn.

- **Hỗ trợ sinh viên học mọi lúc, mọi nơi:** Chatbot, kho học liệu và video mô phỏng có thể truy cập mọi thời điểm, kể cả ngoài giờ học chính khóa.

- **Theo dõi và đánh giá tiến độ học tập:** Hệ thống phân tích thao tác AutoCAD giúp phát hiện lỗi sớm, ghi nhận tiến độ và sinh viên yếu kém để giảng viên can thiệp kịp thời.

- **Khả năng mở rộng cao:** Mô hình có thể áp dụng cho các môn học kỹ thuật khác, không chỉ riêng AutoCAD, tạo nền tảng cho chuyển đổi số toàn diện trong đào tạo.

#### 2.3.2. Nhược điểm

- **Phụ thuộc vào hạ tầng công nghệ:** Máy tính cấu hình yếu, mạng không ổn định sẽ làm giảm hiệu quả hệ thống AI, đặc biệt với plugin giám sát thao tác AutoCAD.

- **Đòi hỏi kỹ năng sử dụng công nghệ từ sinh viên và giảng viên:** Cần thời gian làm quen, tập huấn để khai thác đúng cách; sinh viên yếu công nghệ có thể gặp khó khăn khi sử dụng AI trợ giảng.

- **Khó kiểm soát chất lượng phản hồi của AI:** Nếu AI chưa được huấn luyện đủ tốt từ dữ liệu thực tế, có thể đưa ra gợi ý chưa phù hợp hoặc thiếu chính xác.

- **Chi phí đầu tư ban đầu và nhân lực triển khai:** Xây dựng hệ thống, đào tạo AI, thiết kế học liệu số và plugin cần đầu tư thời gian, công sức và chi phí đáng kể.

**Bảng 1. So sánh mô hình trợ giảng AI đề xuất với phương pháp truyền thống**

Tiêu chí	Phương pháp truyền thống	Mô hình trợ giảng AI đề xuất
Khả năng tương tác	Hạn chế khi lớp đông, giảng viên khó hỗ trợ từng sinh viên kịp thời.	AI hỗ trợ sinh viên 24/7, trả lời câu hỏi và gợi ý thao tác theo ngữ cảnh.
Theo dõi tiến độ	Phụ thuộc vào cảm nhận chủ quan của giảng viên, khó bao quát toàn bộ lớp.	Plugin theo dõi thao tác giúp ghi nhận, phân tích sai sót và báo cáo chi tiết.
Tính cá nhân hóa	Tất cả sinh viên học theo một tốc độ và cách tiếp cận giống nhau.	AI cá nhân hóa gợi ý, cung cấp tài liệu phù hợp năng lực từng sinh viên.
Khả năng tự học	Sinh viên thường bị động, chờ giảng viên hướng dẫn.	Kho học liệu và chatbot khuyến khích sinh viên tự học, tự tra cứu.
Hiệu quả với lớp đông	Hiệu quả giảm rõ rệt khi sĩ số vượt quá khả năng tương tác của giảng viên.	Hệ thống AI có thể hỗ trợ đồng thời nhiều sinh viên mà không bị quá tải.
Chi phí triển khai ban đầu	Thấp, chỉ cần giảng viên và phòng máy cơ bản.	Cao hơn (đào tạo AI, phát triển hệ thống, nâng cấp thiết bị).
Phụ thuộc công nghệ	Không bị ảnh hưởng bởi lỗi phần mềm hay hệ thống.	Phụ thuộc vào máy tính, mạng, và độ ổn định của hệ thống AI.
Mức độ linh hoạt mở rộng	Giới hạn theo năng lực giảng viên và thời gian học.	Có thể mở rộng quy mô lớp và tích hợp thêm học phần khác dễ dàng.

**2.4. Lộ trình triển khai**

Lộ trình triển khai trợ giảng AI trong dạy học AutoCAD được chia thành 4 giai đoạn rõ ràng, đi từ thử nghiệm nhỏ đến mở rộng toàn diện, thể hiện tính khả thi, từng bước và phù hợp với điều kiện của nhà trường. [3]

**Bảng 2. Lộ trình triển khai thực tế**

Giai đoạn	Công việc chính	Thời gian	Mục tiêu	Ghi chú
1. Khởi tạo	Đào tạo AI bằng dữ liệu học phần	1-3 tháng	Tạo nền tảng dữ liệu cho AI	Cần phối hợp giảng viên - kỹ thuật viên - chuyên gia
2. Phát triển	Xây dựng chatbot, plugin, kho video	3-5 tháng	Tạo sản phẩm dùng thử (MVP)	Ưu tiên tính khả thi, đơn giản hóa giao diện
3. Thử nghiệm	Triển khai tại 1 lớp học cụ thể	1 học kỳ	Lấy phản hồi, kiểm nghiệm thực tế	Cần đo lường mức độ tương tác, lỗi
4. Mở rộng	Ứng dụng cho toàn bộ học phần	các học kỳ kế tiếp	Áp dụng toàn diện	Đòi hỏi hệ thống ổn định và hỗ trợ kỹ thuật

**2.4.1. Giai đoạn 1: Khởi tạo**

\* **Công việc chính:** Đào tạo AI với tài liệu học phần AutoCAD sẵn có của trường/khoa. [2]

\* **Thời gian:** 1-3 tháng.

\* **Ý nghĩa:**

- Đây là bước nền tảng quan trọng để hệ thống hiểu nội dung cần hướng dẫn.

- Giai đoạn này cần sự phối hợp giữa

chuyên gia AI và giảng viên chuyên môn để chuẩn hóa lệnh, lỗi thường gặp, câu hỏi mẫu.

**Lưu ý:** Tài liệu học liệu càng rõ ràng thì AI càng hiệu quả.

**2.4.2. Giai đoạn 2: Phát triển**

\* **Công việc chính:** Xây dựng các thành phần chính của hệ thống như:

- Chatbot học tập.
- Plugin theo dõi thao tác AutoCAD.
- Hệ thống chấm lỗi và gợi ý.
- Kho video mô phỏng.

\* **Thời gian:** 3-5 tháng.

- **Ý nghĩa:** Tạo sản phẩm mẫu đủ chức năng cốt lõi.

- **Tính thực tiễn:** Có thể làm theo mô hình MVP (Minimum Viable Product) – tức là chỉ cần phiên bản đầu đủ dùng thử, không cần hoàn thiện ngay từ đầu.

**2.4.3. Giai đoạn 3: Thử nghiệm**

\* **Công việc chính:** Triển khai tại 1 lớp học cụ thể (ví dụ lớp AutoCAD năm nhất Khoa Cơ khí).

\* **Thời gian:** 1 học kỳ.

\* **Ý nghĩa:**

- Thu thập phản hồi từ sinh viên & giảng viên.

- Kiểm tra tính ổn định của AI, mức độ sinh viên tương tác, hiệu quả gợi ý lỗi.

\* **Lợi ích:** Phát hiện vấn đề thực tế như tốc độ phản hồi, độ chính xác của AI, tính thân thiện giao diện.

#### 2.4.4. Giai đoạn 4: Mở rộng

\* **Công việc chính:** Triển khai diện rộng cho tất cả các lớp học phần AutoCAD.

\* **Thời gian:** Học kỳ tiếp theo.

\* **Điều kiện triển khai:**

- Có cơ sở đánh giá từ thử nghiệm thành công.

- Có bổ sung cải tiến từ phản hồi thực tế.

- Có kế hoạch đầu tư về máy chủ hoặc tích hợp hệ thống LMS chính thức.

### 2.5. Mức độ phù hợp

#### 2.5.1. Rõ ràng, tuần tự và dễ triển khai

- Bảng so sánh cho thấy một lộ trình hợp lý, bắt đầu từ việc chuẩn bị nền tảng (dữ liệu, tài nguyên), sau đó phát triển công cụ, thử nghiệm ở quy mô nhỏ và cuối cùng là nhân rộng.

- Thể hiện rõ nguyên tắc phát triển từng bước: “Bắt đầu nhỏ – thử nghiệm – tối ưu – mở rộng”.

#### 2.5.2. Mục tiêu cụ thể theo từng giai đoạn

- Mỗi giai đoạn đều có mục tiêu rõ ràng: tạo nền, phát triển tính năng, kiểm chứng thực tế và ứng dụng đại trà.

- Điều này giúp nhà trường dễ dàng đánh giá hiệu quả ở từng bước và giảm thiểu rủi ro đầu tư sai hướng.

#### 2.5.3. Có tính khả thi cao trong điều kiện thực tế

- Thời gian triển khai từng giai đoạn

hợp lý (1–5 tháng), phù hợp với lịch học và khung đào tạo tại trường đại học.

- Cách tiếp cận này phù hợp với điều kiện cơ sở vật chất, nhân lực và trình độ công nghệ hiện có tại các khoa kỹ thuật, đặc biệt với lớp sĩ số đông, đầu vào chưa đồng đều.

#### 2.5.4. Có thể mở rộng cho các học phần kỹ thuật khác

Mặc dù tập trung vào AutoCAD, nhưng mô hình lộ trình này hoàn toàn có thể được tái sử dụng cho các môn như SolidWorks, Revit, Inventor, hoặc các học phần lập trình, kỹ thuật điện tử v.v.

## 3. Thảo luận và hướng phát triển

### 3.1. Thảo luận

Việc ứng dụng trợ giảng AI trong hướng dẫn thực hành AutoCAD là một bước đi phù hợp với xu hướng chuyển đổi số trong giáo dục đại học, đặc biệt trong các học phần yêu cầu thao tác kỹ thuật phần mềm như tại Trường Đại học Công nghệ Đông Á. Qua phân tích mô hình đề xuất, một số vấn đề cần được trao đổi và thảo luận sâu hơn nhằm đảm bảo khả năng triển khai hiệu quả trên thực tế, cần lưu ý một số vấn đề:

- **Năng lực sinh viên:** Nhiều sinh viên còn thụ động, kỹ năng số hạn chế. Do đó, cần có buổi hướng dẫn sử dụng hệ thống AI ngay từ đầu khóa.

- **Hạ tầng công nghệ:** Phòng máy còn thiếu đồng bộ, cấu hình yếu, mạng không ổn định. Cần nâng cấp hoặc tối ưu hệ thống sao cho chạy nhẹ, có thể dùng offline.

- **Vai trò giảng viên:** AI không thay thế giảng viên, mà hỗ trợ. Giảng viên cần được tập huấn để khai thác tốt công cụ AI trong giảng dạy.

- **Tính mở rộng:** Mô hình có thể áp dụng cho nhiều học phần kỹ thuật khác. Tuy nhiên, cần đầu tư thêm vào việc xây dựng học liệu số và ngân hàng câu hỏi chuyên sâu.

Mô hình trợ giảng AI khả thi và hữu ích, nhưng cần triển khai từng bước, có sự phối hợp giữa nhà trường – giảng viên – sinh viên – kỹ thuật viên để đạt hiệu quả tối ưu.

### 3.2. Hướng phát triển

Đề tài ứng dụng trợ giảng AI trong giảng dạy thực hành AutoCAD có tiềm năng mở rộng và phát triển theo nhiều hướng trong tương lai:

- **Mở rộng sang môn học khác:** Áp dụng cho các học phần như SolidWorks, Revit, CNC, lập trình kỹ thuật...

- **Tích hợp với LMS:** Kết nối AI vào hệ thống quản lý học tập để đồng bộ dữ liệu và báo cáo.

- **Cá nhân hóa học tập:** Gợi ý bài học và bài tập phù hợp năng lực từng sinh viên.

- **Phân tích dữ liệu học tập:** Sử dụng AI để theo dõi, đánh giá và cải tiến phương pháp giảng dạy.

- **Ứng dụng công nghệ VR/AR:** Tạo môi trường học tập trực quan, tăng tương tác trong thực hành kỹ thuật.

### 4. Kết luận

Đề tài đã tập trung nghiên cứu và đề xuất một mô hình ứng dụng trợ giảng AI nhằm hỗ trợ hiệu quả cho giảng viên và sinh viên trong quá trình thực hành học phần AutoCAD. Trong bối cảnh lớp học đông, năng lực sinh viên không đồng đều và hạn chế về hạ tầng thiết bị, việc tích hợp AI không chỉ góp phần cải thiện khả năng tiếp cận kiến thức mà còn nâng cao tính chủ động, cá nhân hóa trong học tập.

Mô hình được xây dựng với các thành phần rõ ràng như: trợ lý học tập ảo, kho học liệu mô phỏng, hệ thống theo dõi thao tác và phản hồi lỗi tự động. Bên cạnh đó, lộ trình triển khai được đề xuất theo hướng từng bước, phù hợp với điều kiện thực tiễn và đảm bảo khả năng mở rộng.

Kết quả nghiên cứu khẳng định rằng, việc ứng dụng AI vào giảng dạy thực

hành kỹ thuật là xu thế tất yếu và khả thi. Tuy nhiên, để thành công, cần có sự đầu tư đồng bộ về hạ tầng, đào tạo nhân lực và cơ chế phối hợp chặt chẽ giữa các bên liên quan. Đề tài không chỉ có ý nghĩa thực tiễn đối với học phần AutoCAD, mà còn mở ra triển vọng ứng dụng rộng rãi trong các học phần kỹ thuật khác tại nhà trường trong thời gian tới.

### TÀI LIỆU THAM KHẢO

[1]. Nguyễn Văn Hiếu (2020). *Ứng dụng trí tuệ nhân tạo trong giáo dục đại học tại Việt Nam: Thực trạng và định hướng*. Tạp chí Khoa học và Công nghệ Đại học Quốc gia Hà Nội.

[2]. Trường Đại học Công nghệ Đông Á (2023). *Giáo trình AutoCAD cơ bản*. Tài liệu lưu hành nội bộ.

[3]. Trịnh Hữu Anh (2022). *Hướng dẫn xây dựng hệ thống học tập thông minh tích hợp AI*. Tạp chí Công nghệ Thông tin và Truyền thông, số 4, tr. 45–50.

[4]. Woolf, B. P. (2010). *Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing E-learning*. Morgan Kaufmann.

[5]. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education.

[6]. Chien, C. F., Lin, Y. H., & Hsu, C. C. (2017). Development of a CAD E-learning Platform Based on Intelligent Tutoring and Student Behavior Analysis. *Computer Applications in Engineering Education*, 25(2), 203–215.

[7]. UNESCO (2021). *AI and Education: Guidance for Policy-makers*. Paris: UNESCO.

[8]. VanLehn, K. (2006). The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.

# NÂNG CAO SẢN XUẤT TINH GỌN TẠI CÔNG TY TNHH SẢN XUẤT HÀNG MAY MẶC ESQUEL VIỆT NAM HÒA BÌNH

**Đỗ Đức Bình**

GS.TS. Viện Thương Mại và Kinh tế quốc tế, Đại học Kinh tế quốc dân

Email: dd302@gmail.com

## TÓM TẮT

Trong bối cảnh cạnh tranh toàn cầu gia tăng và chuyển đổi công nghệ diễn ra nhanh chóng, các doanh nghiệp sản xuất chịu áp lực nâng cao năng suất, giảm chi phí và cải thiện chất lượng sản phẩm, đồng thời bảo đảm phát triển bền vững. Quản trị tinh gọn (QTTG) và sản xuất tinh gọn được xem là phương pháp hiệu quả giúp doanh nghiệp nâng cao hiệu suất thông qua cải tiến liên tục và loại bỏ lãng phí. Nghiên cứu này đánh giá việc triển khai quản trị tinh gọn tại Công ty TNHH Sản xuất Hàng May mặc Esquel Việt Nam Hòa Bình và đề xuất các giải pháp nhằm nâng cao hiệu quả sản xuất.

Nghiên cứu sử dụng kết hợp phương pháp định tính và định lượng, bao gồm phân tích dữ liệu thứ cấp và khảo sát 218 cán bộ công nhân viên, nhằm xem xét bốn yếu tố ảnh hưởng đến việc áp dụng quản trị tinh gọn: lãnh đạo, nhân viên, chính sách và cơ sở hạ tầng gắn với văn hóa doanh nghiệp. Kết quả cho thấy doanh nghiệp đạt được những cải thiện đáng kể sau khi áp dụng các công cụ tinh gọn. Sản lượng tăng từ 3,3 triệu sản phẩm năm 2013 lên 9,1 triệu sản phẩm năm 2025; hiệu suất sản xuất tăng từ 40% lên khoảng 77%; và thời gian sản xuất giảm từ 130 ngày xuống còn 60–80 ngày. Đồng thời, số lượng ý tưởng cải tiến tăng, phản ánh sự hình thành văn hóa cải tiến liên tục.

Tuy nhiên, một số hạn chế vẫn tồn tại như mức độ tham gia cải tiến của người lao động chưa cao, đào tạo chưa phù hợp và truyền thông nội bộ chưa hiệu quả. Vì vậy, nghiên cứu đề xuất tăng cường cam kết lãnh đạo, hoàn thiện đào tạo, cải thiện truyền thông và cơ chế khuyến khích nhằm củng cố việc triển khai quản trị tinh gọn trong dài hạn.

Từ khóa: Quản trị tinh gọn, Sản xuất tinh gọn, Ngành dệt may, cải tiến liên tục, hiệu quả sản xuất

## ABSTRACT

In the context of increasing global competition and rapid technological transformation, manufacturing enterprises are facing growing pressure to enhance productivity, reduce operational costs, and improve product quality while ensuring sustainable development. Lean management and lean production are considered effective approaches that help organizations improve operational performance through continuous improvement and waste elimination. This study evaluates the implementation of lean management at Esquel Vietnam Hoa Binh Garment Manufacturing Co., Ltd. and proposes solutions to enhance production efficiency.

The research employs a combination of qualitative and quantitative methods, including secondary data analysis and a survey of 218 employees, to examine four

*key factors influencing lean management adoption: leadership, employees, policies, and infrastructure associated with organizational culture. The findings indicate that the company has achieved significant improvements after applying lean tools. Specifically, production output increased from 3.3 million units in 2013 to 9.1 million units in 2025; production efficiency rose from 40% to approximately 77%; and total production time decreased from 130 days to 60–80 days. In addition, the number of improvement initiatives increased, reflecting the gradual development of a continuous improvement culture within the organization.*

*However, several limitations remain, including low employee participation in improvement activities, inadequate training programs, and ineffective internal communication. Therefore, the study recommends strengthening leadership commitment, developing systematic training programs, improving internal communication, and enhancing incentive mechanisms to sustain lean management implementation in the long term.*

**Keywords:** *Lean management, Lean production, Garment industry, Kaizen, Operational performance*

**Ngày nhận bài:** 28/02/2026; **Ngày sửa bài:** 10/03/2026; **Ngày duyệt đăng:** 02/04/2026

## **1. MỞ ĐẦU**

Trong bối cảnh môi trường kinh doanh ngày càng cạnh tranh và quá trình chuyển đổi số diễn ra mạnh mẽ dưới tác động của cuộc Cách mạng công nghiệp 4.0 và xu hướng phát triển công nghệ 5.0, các doanh nghiệp buộc phải tìm kiếm những giải pháp quản trị và sản xuất hiệu quả nhằm duy trì khả năng cạnh tranh và phát triển bền vững. Mục tiêu của doanh nghiệp không chỉ là gia tăng sản lượng mà còn phải tối ưu hóa chi phí, rút ngắn thời gian sản xuất, nâng cao chất lượng sản phẩm và đáp ứng ngày càng tốt hơn nhu cầu của khách hàng. Do đó, cải tiến liên tục, đổi mới phương thức quản trị và nâng cao hiệu quả hoạt động đã trở thành yêu cầu tất yếu đối với các doanh nghiệp trong bối cảnh hiện nay.

Một trong những phương thức quản trị hiện đại được nhiều doanh nghiệp trên thế giới áp dụng là quản trị tinh gọn (Lean Management). Phương thức này được hình thành và phát triển tiêu biểu tại Nhật Bản thông qua Hệ thống sản

xuất Toyota (Toyota Production System – TPS). TPS được xây dựng trên nền tảng các nguyên tắc như tiêu chuẩn hóa công việc, tổ chức dòng chảy sản xuất hợp lý và loại bỏ các yếu tố gây lãng phí trong quá trình sản xuất. Trọng tâm của TPS là nhận diện và loại bỏ ba dạng tổn thất chính, bao gồm lãng phí (Muda), mất cân đối (Mura) và quá tải (Muri). Đồng thời, hệ thống này áp dụng nhiều công cụ và phương pháp quản trị hiệu quả như 5S, Kaizen, Kanban, quản lý trực quan, dòng chảy liên tục và hệ thống kéo. Nhờ triển khai thành công quản trị tinh gọn, Toyota đã trở thành hình mẫu điển hình về quản trị sản xuất hiện đại và được nhiều doanh nghiệp trên thế giới học tập và vận dụng.

Các nghiên cứu quốc tế tiếp theo đã góp phần hoàn thiện cơ sở lý luận về quản trị tinh gọn. Tiêu biểu, Liker (2004) trong tác phẩm *The Toyota Way – 14 Management Principles from the World’s Greatest Manufacturer* đã hệ thống hóa 14 nguyên tắc quản trị của Toyota, nhấn mạnh vai trò của chất lượng, con người

và cải tiến liên tục trong hoạt động sản xuất. Trước đó, Womack và Jones (1996) trong cuốn *Lean Thinking* đã đề xuất năm nguyên tắc cốt lõi của quản trị tinh gọn, bao gồm xác định giá trị, nhận diện chuỗi giá trị, thiết lập dòng chảy liên tục, vận hành hệ thống kéo và hướng tới sự hoàn thiện không ngừng. Những nguyên tắc này đã trở thành nền tảng lý thuyết quan trọng cho việc nghiên cứu và triển khai quản trị tinh gọn trong nhiều lĩnh vực khác nhau.

Tại Việt Nam, quản trị tinh gọn đã được nghiên cứu và áp dụng trong nhiều ngành nghề, đặc biệt trong lĩnh vực sản xuất và dịch vụ. Một số nghiên cứu tập trung vào việc xác định các điều kiện và yếu tố ảnh hưởng đến hiệu quả triển khai quản trị tinh gọn trong doanh nghiệp. Các kết quả nghiên cứu cho thấy hiệu quả áp dụng quản trị tinh gọn phụ thuộc vào nhiều yếu tố, trong đó vai trò của lãnh đạo, nhận thức và kỹ năng của người lao động, hệ thống đào tạo, cơ chế khuyến khích và văn hóa doanh nghiệp được xem là những yếu tố có ảnh hưởng quan trọng. Tuy nhiên, phần lớn các nghiên cứu tại Việt Nam vẫn mang tính mô tả hoặc tập trung vào từng trường hợp cụ thể, chưa xây dựng được mô hình quản trị tinh gọn mang tính hệ thống và bền vững cho các doanh nghiệp sản xuất quy mô lớn.

Trong ngành dệt may và sản xuất hàng tiêu dùng, việc áp dụng quản trị tinh gọn đã bước đầu mang lại những kết quả tích cực, góp phần nâng cao năng suất lao động, giảm chi phí sản xuất và cải thiện chất lượng sản phẩm. Tuy nhiên, quá trình triển khai vẫn gặp phải nhiều khó khăn như hạn chế về tư duy quản trị, thiếu cam kết dài hạn từ lãnh đạo, nguồn lực tài chính và nhân lực chưa đáp ứng yêu cầu, cũng như sự thiếu đồng bộ trong việc áp dụng các công cụ và phương pháp quản trị tinh gọn. Bên cạnh đó, việc mở rộng

áp dụng quản trị tinh gọn sang các lĩnh vực dịch vụ như tài chính – ngân hàng và y tế cũng cho thấy hiệu quả rõ rệt trong việc cải tiến quy trình và nâng cao chất lượng dịch vụ. Tuy nhiên, các nghiên cứu trong lĩnh vực này chủ yếu tập trung vào từng quy trình chức năng riêng lẻ, chưa xem xét toàn diện hệ thống quản trị tinh gọn ở cấp độ tổ chức.

Tổng hợp các công trình nghiên cứu cho thấy quản trị tinh gọn đã được tiếp cận dưới nhiều góc độ và lĩnh vực khác nhau, song vẫn còn thiếu các nghiên cứu chuyên sâu về việc xây dựng và hoàn thiện hệ thống quản trị tinh gọn một cách tổng thể tại các doanh nghiệp sản xuất quy mô lớn, đặc biệt trong ngành dệt may – lĩnh vực có đặc điểm sản xuất phức tạp, lực lượng lao động lớn và yêu cầu cao về chất lượng, chi phí và thời gian giao hàng. Bên cạnh đó, mối quan hệ giữa yếu tố con người, hệ thống quản trị và việc duy trì hiệu quả quản trị tinh gọn trong dài hạn vẫn chưa được làm rõ, đặt ra nhu cầu tiếp tục nghiên cứu trong bối cảnh doanh nghiệp Việt Nam hiện nay.

Công ty TNHH Sản xuất Hàng May mặc Esquel Việt Nam Hòa Bình được thành lập năm 2013, trực thuộc Tập đoàn Esquel (Hồng Kông), là một trong những doanh nghiệp sản xuất quy mô lớn tại tỉnh Hòa Bình với khoảng 4.500 lao động và sản lượng hơn 10 triệu sản phẩm mỗi năm. Mặc dù đạt được nhiều kết quả tích cực trong hoạt động sản xuất, doanh nghiệp vẫn đang đối mặt với một số hạn chế trong công tác quản trị và tổ chức sản xuất. Do đó, việc nghiên cứu và hoàn thiện hệ thống quản trị tinh gọn tại doanh nghiệp được xem là giải pháp quan trọng nhằm nâng cao hiệu quả hoạt động, tăng cường năng lực cạnh tranh và hướng tới phát triển bền vững trong bối cảnh hội nhập kinh tế quốc tế.

## 2. GIẢI QUYẾT VẤN ĐỀ

### 2.1. Mô hình lý thuyết

Ngôi nhà Lean được xây dựng trên nền móng của Triết lý dài hạn gồm: văn hóa 5E (Engage (Gắn kết), Explore (Khám phá), Explain (Giải thích), Elaborate (Áp dụng/Mở rộng), và Evaluate (Đánh giá), môi trường xanh, 3T (tiết giảm, tái sử dụng, tái chế), văn hóa không tăng ca và cân bằng giữa công việc và cuộc sống (No OT and work life balance).

liên tục cải tiến - Kaizen, gồm các cải tiến để loại bỏ lãng phí và diễn ra liên tục.

+ Mái của ngôi nhà Lean là SQDCM (Best Safety - Best Quality - Shorter lead time (Delivery) - Lowest Cost - High Morale). Đây chính là kết quả mà công ty hướng tới: An toàn hàng đầu - Chất lượng tốt nhất - Giao hàng đúng hạn - Chi phí thấp nhất - Tinh thần làm việc cao.

- 8 Nguyên tắc Lean – 8 Lean principles

### Ngôi Nhà Lean



Hình 1: Mô hình ngôi nhà Lean  
(Nguồn: Eduardo L. Garcia, 2016)

Nền móng phía trên là quản lý trực quan, trao quyền cho người lao động và ổn định-chuẩn hóa quy trình. Đặc biệt, trao quyền cho người lao động nhấn mạnh đến sơ đồ tổ chức dạng phẳng, tạo ra môi trường làm việc minh bạch và tinh gọn.

+ Tiếp theo là hai cột trụ giúp cho ngôi nhà Lean vươn cao và phát triển, lần lượt là kiểm soát hàng tồn và tự động ngăn ngừa lỗi. Trái tim của ngôi nhà Lean là

So với Lean house thì 8 nguyên tắc Lean là một danh sách gồm 8 nguyên tắc cụ thể hơn, chi tiết hơn, dễ dàng thực hiện hơn và dễ dàng tiếp cận tới người lao động.

+ Triết lý dài hạn: phát triển văn hóa 5E, tạo môi trường xanh, 3T (tiết giảm, tái sử dụng, tái chế), no OT, cân bằng giữa công việc - cuộc sống.

+ Quản lý trực quan: áp dụng quản lý trực quan bằng xác định quy trình, gắn nhãn, bảng hiển thị, xây dựng bảng KPI và số liệu hợp lý.

+ Trao quyền cho nhân viên: xây dựng sơ đồ tổ chức dạng phẳng. Một chuyên trưởng quản lý một chuỗi công đoạn: may-PI-hoàn thiện. Giá trị cốt lõi là dòng chảy sản phẩm được quản lý như một thực thể thống nhất và duy nhất.

+ Ổn định và chuẩn hóa quy trình: trên nền tảng 5S xây dựng 6S (5S + An toàn). Xây dựng quy trình công đoạn, sơ đồ chuyên và hướng dẫn công việc chuẩn.

+ Kiểm soát hàng tồn: áp dụng phương pháp quản lý của hệ thống kéo: chỉ sản xuất ra những gì cần và khi cần, tạo dòng chảy liên tục và xây dựng PST.

+ Tự động hóa thông minh: ngăn chặn lỗi, máy tự động dừng khi phát hiện ra hàng lỗi, xây dựng chất lượng tại gốc, nguyên tắc 3D, mỗi cá nhân tự kiểm hàng, sử dụng gá cũ hỗ trợ, giảm thời gian cài đặt, tạo ra nhiều ý tưởng cải tiến.

+ Phát triển con người: thiết lập ma trận kỹ năng phát triển tài năng cho mỗi công nhân viên và xây dựng đội ngũ chuyên gia đào tạo nội bộ.

+ Giảm lãng phí: tìm và loại bỏ 8 loại lãng phí.

- Hệ thống chuyên tốc hành – eXML (Express Manufacturing Line)

Phương thức cơ bản của chuyên eXML là chuyên may sử dụng hệ thống kéo và sản xuất đơn chiếc, với sự tham gia hỗ trợ của tất cả bộ phận support trong đó chuyên trưởng là người đứng đầu, với mục tiêu sản xuất ra sản phẩm nhanh nhất, chất lượng tốt nhất, giải quyết vấn đề phát sinh kịp thời nhất. Đối với sản xuất đơn chiếc, tại mỗi trạm của công đoạn chính, lượng bán thành phẩm tồn tối thiểu là 01 chiếc và tối đa là 3 chiếc. Nếu lượng bán thành phẩm tồn này ít hơn một chiếc tức là năng lực của công đoạn trước chậm hơn công đoạn sau hoặc năng lực của công đoạn sau đang nhanh hơn công đoạn trước. Nếu lượng bán thành phẩm tồn trên giá nhiều hơn ba

chiếc tức là năng lực của công đoạn trước nhanh hơn công đoạn sau hoặc năng lực công đoạn sau đang chậm hơn công đoạn trước. Chuyên trưởng cần nhanh đề khắc phục: bố trí công nhân, máy móc, sử dụng máy tự động, sử dụng gá cũ cải tiến, đào tạo lại kỹ năng cho công nhân, hoặc điều chỉnh lại mục tiêu cũng như nhịp sản xuất (takt time) để hỗ trợ công đoạn. Thiết lập quy trình: hệ thống kéo tức là chỉ sản xuất loại hàng cần và khi cần, không sản xuất ồ ạt dẫn đến lãng phí. Pull từ khu vực đóng hàng kéo đến DC.

### **2.2. Mô hình quản trị tinh gọn tại công ty TNHH sản xuất hàng may mặc Việt Nam Hòa Bình**

Tại Esquel Hòa Bình, có 4 phương pháp để có thể cân bằng chuyên hiệu quả:

+ Loại bỏ thao tác hoặc công đoạn thừa

+ Kết hợp gộp 2 hoặc 3 công đoạn

+ Sắp xếp lại trình tự thao tác hoặc trình tự công đoạn

+ Sử dụng công cụ, vật dụng để hỗ trợ sản xuất (rập, cũ, khuôn...).

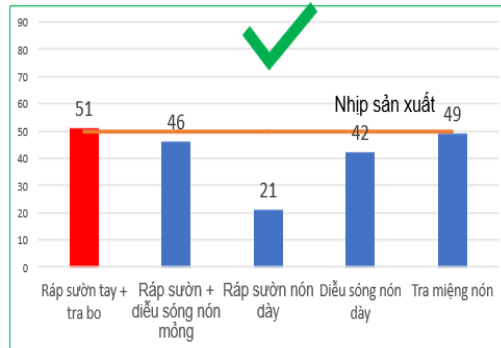
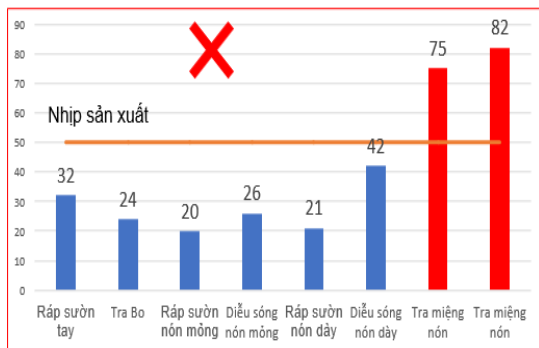
Ở cấp độ 1 gồm các kiến thức tổng quan và quan trọng nhất về QTTG tại Esquel Việt Nam Hòa Bình. Nếu không được đào tạo kỹ lưỡng sẽ áp dụng sai cách và dẫn đến những hệ lụy tiêu cực. Cấp độ này là bắt buộc với cấp giám đốc và quản lý, giúp họ có được tư duy định hướng và các biện pháp giải quyết vấn đề một cách hệ thống.

- PST

PST là viết tắt của Pit Stop Team hay còn gọi là đội phản ứng nhanh cho hoạt động rải chuyên sang đơn hàng mới. Đội PST được hình thành theo khái niệm từ đường đua công thức 1 và nguyên lý hoạt động của hệ thống. Nhiệm vụ là hỗ trợ chuyên may để thời gian rải chuyên là nhanh nhất (2 giờ), hiệu suất bị giảm thấp nhất (0%) (Hình 2).

Công đoạn	Công nhân	Thời gian (giây)
Ráp sườn tay	Dương	32
Tra bo	Hà	19
Ráp sườn nón mỏng	Hiển	20
Điều sườn nón mỏng	Hoa	26
Ráp sườn nón dày	Hoàn	21
Điều sườn nón dày	Trúc	42
Tra miệng nón	Vũ Linh	75
Tra miệng nón	Yến	82

Công đoạn	Công nhân	Thời gian (giây)
Ráp sườn tay + tra bo	Dương	51
Ráp sườn + điều song	Hiển	46
Ráp sườn nón dày	Hoa	21
Điều sườn nón dày	Trúc	42
Tra miệng nón	Linh+Yến	49



**Hình 2. Cân bằng chuyền sản xuất năm 2025**

(Nguồn: Phòng đào tạo PST, công ty Esquel Hòa Bình)

- Sơ đồ chuỗi giá trị - VSM

VSM đã được áp dụng từ năm 2016 như một công cụ hữu hiệu nhằm trực quan hóa dòng chảy của sản phẩm. Phương pháp này giúp xác định các hoạt động tạo ra giá trị và các hoạt động không tạo ra giá trị. Với các hoạt động không tạo ra giá trị, người thực hiện sẽ đưa ra phương pháp cải tiến nhằm loại bỏ các lãng phí này

Cấp độ 2 là bắt buộc với khối sản xuất. Công cụ eXLM và cân bằng chuyền được triển khai từ những ngày đầu khi thành lập nhà máy năm 2013. Riêng VSM được triển khai từ năm 2016, tuy nhiên VSM đã chứng tỏ được tầm quan trọng và hiệu quả của mình khi giúp người quản lý trực quan hơn về dòng chảy của sản phẩm và cơ hội cải tiến giúp tăng năng suất làm việc. Cũng như cấp độ 1, bốn phương pháp của cấp độ 2 được đưa vào chương trình đào tạo nội bộ. Đã biết, đây còn là

các khía cạnh liên quan đến chỉ số KPIs để đánh giá năng lực của cấp quản lý và chuyên trưởng khối sản xuất.

- 5S

Esquel Hòa Bình xác định 5S là yêu cầu cơ bản nhất khi áp dụng QTTG, hay nói cách khác nếu không có 5S sẽ không có tinh gọn (no 5S no Lean). Đây là một phương pháp đơn giản và dễ thực hiện, gồm các hướng dẫn về cách sắp xếp lại nơi làm việc nhằm tạo ra sự tối ưu, nâng cao an toàn hiệu quả công việc và tạo kỷ luật trong môi trường làm việc. Do vậy, phương pháp 5S là kiến thức và công việc yêu cầu 100% cán bộ công nhân viên của công ty phải được đào tạo và thực hành. Tại công ty, thời gian thực hiện 5S diễn ra trong 2 phút và có bốn mốc giờ thực hiện như sau: 8:55, 1:25, 4:55 và 7:55. Sau khi có nhạc chuông báo hiệu, tất cả công nhân viên sẽ dừng công việc đang làm và thực hiện 5S trong khu vực làm

## Kinh tế

việc của mình.

- Quản lý trực quan – Visual management

Esquel Hòa Bình sử dụng quản lý trực quan cho phép mọi người nhìn vào đó để nhận ra tiêu chuẩn và sự sai lệch ngoài tiêu chuẩn. Mục đích là muốn công nhân viên nhận ra được sự sai lệch. Những sự sai lệch này có thể là theo hướng tiêu cực hoặc tích cực. Đối với sự sai lệch tiêu cực, đây là cơ hội để sử dụng công cụ cải tiến nhằm cải thiện khu vực làm việc. Với những sai lệch tích cực là đáng ghi nhận và là cơ hội căn cứ để cập nhật và ban hành tiêu chuẩn, quy trình mới. Tại Esquel Hòa Bình, quản lý trực quan được chia thành ba mức độ thực hiện (Hình 3):

+ Mức độ 1: bắt đầu với nền tảng 5S, tổ chức lại nơi làm việc.

+ Mức độ 2: hiển thị trực quan

+ Mức độ 3: kiểm soát trực quan (kiểm soát các tiêu chuẩn)

quá trình di chuyển chất lượng sản phẩm sẽ bị ảnh hưởng.

Nguyên nhân: sắp xếp vị trí chưa hợp lý, quá trình vận chuyển phụ thuộc nhiều vào sức lao động của công nhân, chưa áp dụng các công cụ giảm thời gian.

Biện pháp: sắp xếp lại vị trí làm việc hợp lý bằng 5S và tự động hóa.

+ Hàng tồn (inventory):

Ảnh hưởng tới: kéo dài thời gian sản xuất, chiếm diện tích và tăng chi phí lưu kho bãi, cản trở giao tiếp, có xu hướng che đậy vấn đề, ảnh hưởng tới chất lượng sản phẩm.

Nguyên nhân: chưa áp dụng quy tắc FIFO, cân bằng dòng chảy chưa tốt.

Biện pháp: lập kế hoạch cân đối giữa sản xuất và nhập nguyên phụ liệu, áp dụng hệ thống sản xuất đúng lúc (Just in time) hoặc áp dụng FIFO.

+ Thao tác (motion):

Ảnh hưởng: lãng phí thời gian, ảnh



Mức độ 1: Thực hiện 5S



Mức độ 2: Hiển thị trực quan



Mức độ 3: Kiểm soát trực quan

**Hình 3.** Ba mức độ quản lý trực quan tại công ty

- 8 lãng phí – 8 wastes

Công ty đã nhận diện ra 8 loại lãng phí đó là: vận chuyển, hàng tồn, thao tác, chờ đợi, sản xuất thừa, công đoạn thừa, hàng lỗi và tài năng con người. Cụ thể:

+ Vận chuyển (transportation):

Ảnh hưởng: lãng phí thời gian, tác động xấu tới hiệu suất làm việc và trong

hưởng hiệu suất và sức khỏe lao động.

Nguyên nhân: chưa có tiêu chuẩn về thao tác làm việc, vị trí làm việc chưa sắp xếp hợp lý, dụng cụ làm việc chưa phù hợp với người lao động.

Biện pháp: sắp xếp lại vị trí và công cụ làm việc hợp lý, đào tạo công nhân theo thao tác chuẩn và cải tiến tìm ra thao tác tốt nhất.

+ Chờ đợi (waiting):

Ảnh hưởng: Gây lãng phí lớn về thời gian sản xuất, công sức, chi phí.

Nguyên nhân: kế hoạch sản xuất chưa hợp lý, sản xuất hàng loạt, con người.

Biện pháp: kế hoạch sản xuất hợp lý, sắp xếp lại dòng chảy, sử dụng tự động hóa.

+ Sản xuất thừa (over-production):

Ảnh hưởng: tạo ra hàng tồn lớn, ảnh hưởng đến chất lượng sản phẩm, tăng lượng hàng tồn và tăng chi phí thuê kho bãi.

Nguyên nhân: dự báo sai về nhu cầu sản xuất và tư duy của người quản lý.

Biện pháp: áp dụng hệ thống kéo: chỉ sản xuất hàng có thể bán hoặc xuất đi ngay. Cải thiện khả năng chuyển đổi/lắp đặt máy móc.

+ Công đoạn thừa (over-processing):

Biện pháp: phương pháp 3D, tìm nguyên nhân gốc rễ, chất lượng tại gốc.

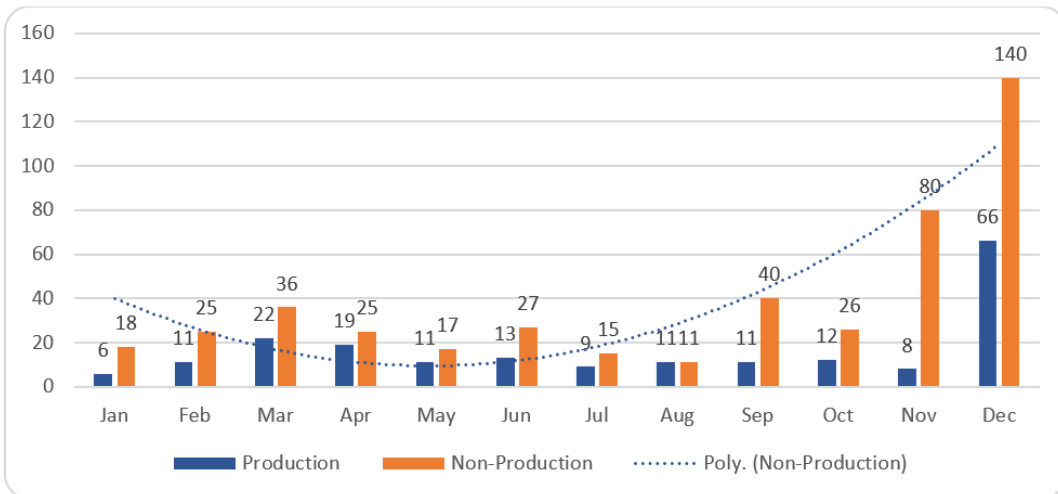
+ Tài năng (talent):

Ảnh hưởng: không trọng dụng được tài nguyên có sẵn là khả năng đặc biệt, tài năng, kinh nghiệm, sáng tạo, kỹ năng, kiến thức tốt nhất của công nhân viên.

Biện pháp: xây dựng chương trình đào tạo nội bộ và ma trận kỹ năng.

- Ý tưởng cải tiến – Kaizen (Idea system)

Hệ thống ý tưởng của Esquel Hòa Bình lấy cảm hứng từ Kaizen: “Kai” là thay đổi, “zen” là tốt hơn. Kaizen là thay đổi để tốt hơn, mỗi người mỗi ngày bắt cứ nơi đâu luôn liên tục cải tiến, sự thay đổi giúp cải thiện công việc trở nên dễ dàng hơn nhằm nâng cao hiệu suất và năng lực làm việc, cũng như môi trường làm việc an toàn.



Ảnh hưởng: dẫn tới lãng phí về chờ đợi, thao tác, nhân lực.

Nguyên nhân: chưa có tiêu chuẩn về quy trình, layout, thông số kỹ thuật.

Biện pháp khắc phục: kiểm tra quy trình và cần luôn đổi mới phương pháp.

+ Hàng lỗi (defect):

Ảnh hưởng: đây được coi là mẹ của mọi lãng phí, dẫn tới ảnh hưởng và thiệt hại lớn về chi phí cho nhà máy.

Hình 4. Số lượng cải tiến 2025

Nguồn: Phòng QC của công ty

Một ý tưởng SMART phải rõ ràng về mục tiêu, tính khả thi, rõ ràng các bên liên quan, thời gian thực hiện và số liệu (cải tiến được bao nhiêu công việc, giảm được bao nhiêu thời gian, cách áp dụng như thế nào). Nếu công nhân viên chỉ đưa ra được ý tưởng nhưng không trực tiếp thực hiện thì không được coi là một ý

tưởng cải tiến, mà chỉ được coi là một gợi ý (suggestion). Một ý tưởng cải tiến là một ý tưởng đã hoàn thành trên 1 khu vực hoặc hoàn thành 1 mẫu thử, là ý tưởng mới là không trùng với ý tưởng đã có, không đi ngược với tiêu chuẩn đã có (trừ khi đề xuất 1 tiêu chuẩn mới thay thế), không phải là ý tưởng mang tính tạm bợ. Tuy nhiên, Esquel Hòa Bình cũng khuyến khích công nhân viên thực hiện ý tưởng đã có với quy mô lớn hơn, và việc thực hiện này được công ty công nhận là “spreading idea” (Hình 4).

Nhìn chung về cấp độ 3, Esquel Hòa Bình đã thực hiện một cách bài bản và nghiêm túc khi đưa 4 phương pháp này vào công việc hàng ngày của công nhân viên, không chỉ trong khối sản xuất mà cả trong khối văn phòng. Phòng Quản lý Tinh gọn đã kết hợp với bộ phận đào tạo và các trưởng bộ phận, đưa bốn loại phương pháp này vào trong chương trình đào tạo nội bộ cũng như giám sát thực hiện trong thực tế công việc tại bộ phận. Cụ thể: lớp đào tạo Lean có lồng ghép kiến thức 4 phương pháp trên được đào tạo trong ngày nhận việc đầu tiên của công nhân viên. Hàng năm, lớp học này sẽ mở lại để tái đào tạo giúp công nhân viên ôn lại và ghi nhớ về các kiến thức cơ bản của quản trị tinh gọn. Thực tế, công nhân viên đã có nhận thức căn bản và thực hiện 4 phương pháp này trong công việc hàng ngày.

Một doanh nghiệp được gọi là tinh gọn là khi họ cho thấy sự thay đổi tích cực trong hoạt động sản xuất kinh doanh, tức là doanh nghiệp đó gặt hái được thành công và hoàn thành những mục tiêu tinh gọn đã đề ra.

Dựa trên số liệu tổng hợp, có thể nhận thấy sự chuyển biến rõ rệt trong quá trình triển khai quản trị tinh gọn tại doanh nghiệp qua các giai đoạn từ năm 2013 đến giai đoạn 2023–2025. Giai đoạn

2013–2015 được xem là thời kỳ khởi động khi doanh nghiệp từng bước tiếp cận quản trị tinh gọn thông qua đào tạo và truyền thông nội bộ. Đến năm 2016, cùng với việc mở rộng quy mô sản xuất, các vấn đề vận hành phát sinh đã thúc đẩy doanh nghiệp áp dụng mạnh mẽ hơn các công cụ quản trị tinh gọn. Giai đoạn 2022–2025 ghi nhận những cải thiện rõ rệt về hiệu quả sản xuất, chất lượng và năng suất lao động, phản ánh mức độ trưởng thành của hệ thống quản trị tinh gọn.

*Bảng 1. Kết quả thực hiện Lean tại Esquel Hòa Bình*

Chỉ số đo lường	2013	2016	2023	2024	2025
	Thành lập	Mở 2 ca			
Sản lượng (triệu chiếc)	3.3	5.4	9.8	14.3	9.1
Hiệu suất EFF (%)	40	80	82.5	75.8	77.46
Tổng số công nhân may	750	1300	2630	2405	2623
e X M L s (Physical line)	25	43	42	39	44
Số lượng chuyển đổi	0	8	42	39	44
Final LAR (%)	N/A	99.2	91.7	99.8	99.8
Chỉ số an toàn lao động (%)	0.3	0.23	0.19	0.09	0.19
Ý tưởng cải tiến	0	299	605	834	679
Total lead-time (days)	130	120	90-115		60-80
dCT Cut to Pack (days) *	32.3	7.1	7.2	9.3	5.5
OTD (%) **	99	99.3	99.8	100	100
Số lượng nhà cung cấp	N/A	201	177	177	158
Số lượng khách hàng	5	17	16	17	17
* dCT – dynamic cycle time: thời gian sản xuất					
** OTD – On time delivery: tỷ lệ giao hàng đúng hạn cho khách hàng					

(Nguồn: Phòng Quản lý Tinh gọn)

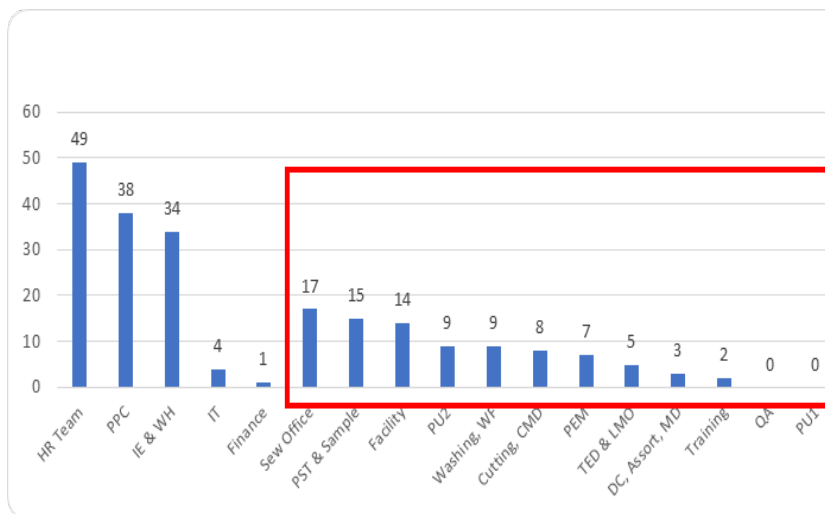
Các chỉ số hoạt động cho thấy sự cải thiện đáng kể về hiệu quả và tính bền vững. Chỉ số tai nạn lao động giảm từ 0,3 năm 2013 xuống còn 0,19 năm 2025, cho thấy môi trường làm việc và mức độ tuân thủ quy trình an toàn được cải thiện. Đồng thời, thời gian chu kỳ sản xuất của các công đoạn chính giảm mạnh từ 32,3

ngày xuống còn 5,5 ngày, phản ánh việc tối ưu hóa dòng chảy giá trị và loại bỏ các hoạt động không tạo giá trị. Số lượng ý tưởng cải tiến cũng tăng đáng kể, từ 0 ý tưởng năm 2013 lên 299

ý tưởng năm 2016 và đạt 679 ý tưởng vào năm 2025, cho thấy văn hóa cải tiến liên tục đã từng bước được hình thành trong doanh nghiệp. Bên cạnh đó, sản lượng sản xuất tăng từ 3,3 triệu sản phẩm lên 9,1 triệu sản phẩm trong cùng giai đoạn, thể hiện hiệu quả của việc nâng cao năng suất và cải tiến quy trình. Tổng thời gian sản xuất cho một đơn hàng giảm từ khoảng 130 ngày xuống còn 60–80 ngày, đồng thời số lượng nhà cung cấp được tinh gọn từ 201 xuống còn khoảng 158 đơn vị, hỗ trợ triển khai hiệu quả nguyên tắc sản xuất đúng thời điểm (Just-in-Time).

Tuy nhiên, quá trình triển khai QTTG vẫn còn một số hạn chế, như việc thực hiện 5S tại một số bộ phận chưa thực sự bền vững, mức độ tham gia cải tiến của người lao động còn thấp và chương trình đào tạo nội bộ chưa phù hợp với từng nhóm đối tượng. Nhìn chung, kết

quả nghiên cứu cho thấy QTTG đã mang lại những cải thiện đáng kể về hiệu quả hoạt động và năng suất, song doanh nghiệp cần tiếp tục hoàn thiện cơ chế đào tạo, tăng cường sự tham gia của người lao động và nâng cao cam kết của lãnh đạo để duy trì hiệu quả cải tiến trong dài hạn.



Hình 5. Tổng hợp ý tưởng cải tiến theo bộ phận tháng 12/2025

(Nguồn: Phòng Quản lý Tinh gọn)

Kết quả kiểm tra và đánh giá quy trình cho thấy một số bộ phận, đặc biệt thuộc khâu sản xuất, chưa tuân thủ nghiêm ngặt các quy định vận hành. Việc giám sát và thực thi quy trình chưa thường xuyên đã dẫn đến các sai sót như ghi nhầm cỡ số, sai thông tin sản phẩm khi giao hàng, phải làm lại sản phẩm hoặc xuất hiện côn trùng trong thùng hàng. Những sai sót này không chỉ làm gia tăng chi phí và thời gian xử lý mà còn ảnh hưởng trực tiếp đến uy tín doanh nghiệp, nhất là khi công ty cung cấp sản phẩm áo sơ mi cao cấp cho các khách hàng có tiêu chuẩn chất lượng khắt khe như Muji (Nhật Bản), Hugo Boss (Đức) và Charles Tyrwhitt (Anh).

Một vấn đề đáng chú ý khác là tình trạng tồn kho nguyên phụ liệu vượt mức quy định. Theo tiêu chuẩn nội bộ, thời gian lưu kho (aging) đối với vải là 21 ngày và phụ liệu là 14 ngày (Hình 6). Tuy nhiên, trên thực tế, một số loại vải có thời gian lưu kho từ 3 tháng đến 1 năm, trong khi một số phụ liệu như cúc có thể tồn kho đến 6 tháng. Điều này không chỉ làm tăng chi phí lưu kho mà còn ảnh hưởng đến chất lượng nguyên vật liệu và hiệu quả sử dụng diện tích kho bãi.

Bên cạnh đó, chương trình đào tạo nội bộ về quản trị tinh gọn chưa được thiết kế phù hợp với từng cấp bậc nhân sự. Các khóa đào tạo thường mang tính lý thuyết, thiếu thực hành và chưa có lộ trình đào tạo rõ ràng, dẫn đến nhiều học viên chưa nắm vững kiến thức và khó áp dụng vào thực tế công việc. Hệ thống truyền thông và khuyến khích cải tiến cũng chưa thực sự hiệu quả, khiến mức độ tham gia của người lao động vào các hoạt động cải tiến còn hạn chế.

**2.2. Các yếu tố ảnh hưởng đến mô hình quản trị tinh gọn tại công ty**

Kết quả khảo sát nội bộ đối với 218 cán bộ công nhân viên vào tháng 9/2025 cho thấy yếu tố lãnh đạo đóng vai trò quan trọng trong việc triển khai QTTG. Trong số này, 166 người (76%) đánh giá ban lãnh đạo có sự quan tâm, cam kết và định hướng chiến lược rõ ràng đối với quản trị tinh gọn, trong khi 52 người (24%) cho rằng mức độ cam kết và định hướng vẫn chưa được thể hiện rõ ràng. Điều này phản ánh sự cần thiết phải tăng cường vai trò của quản lý cấp trung trong việc kết nối giữa chiến lược của lãnh đạo và quá trình thực thi tại hiện trường sản xuất, nhằm bảo đảm việc triển khai quản trị tinh gọn đạt hiệu quả bền vững. Các yếu tố được xem xét chính là: Công nhân

viên, chính sách, yếu tố cơ sở hạ tầng và văn hóa doanh nghiệp

*Bảng 2. Kết quả khảo sát đánh giá về yếu tố Lãnh đạo*

TT	Nội dung	Trả lời "Có"	Trả lời "Không"
1.1	Lãnh đạo có cam kết dài hạn khi áp dụng	166	52
1.2	Lãnh đạo đưa ra định hướng trong việc áp dụng QTTG và khuyến khích CBNV cùng tham gia	166	52
1.3	Lãnh đạo có cam kết cung cấp tài chính cho các dự án cải tiến	166	52
1.4	Lãnh đạo đưa ra tầm nhìn và chiến lược rõ ràng về việc áp dụng QTTG	166	52

*(Nguồn: Khảo sát tháng 9 năm 2025)*

- Về yếu tố công nhân viên:

Nếu như lãnh đạo nắm vai trò định hướng dẫn dắt thì nhân viên chính là nhóm trực tiếp thực thi. Tại Esquel, nhân viên là những người trực tiếp thực thi, tham gia thực hiện cải tiến. Số lượng nhân viên, đặc biệt là công nhân cao gấp nhiều lần so với cấp lãnh đạo, điều này ảnh hưởng rất nhiều tới sự thành công của QTTG.

Theo kết quả khảo sát, trong 218 phiếu trả lời có 211 người (tương đương 97%) khẳng định CBNV làm việc tại công ty có trình độ và tiếp thu nhanh. Điều này khá sát với thực tế khi độ tuổi trung bình của nhà máy là 30.2 tuổi. Với 159 người (tương đương 73%) tham gia khảo sát nói rằng CBNV có khả năng làm việc nhóm tốt. 204 người (94%) khẳng định họ không được đào tạo tốt về QTTG, 163 người (75%) người khảo sát cho rằng CBNV chưa chủ động áp

dụng QTTG (Bảng 3).

CBNV chưa nhận thức được lợi ích của QTTG nên việc thực hiện QTTG diễn ra hời hợt, qua loa. Từ đây, với lợi thế nhân lực và sự tiếp thu, công ty nên cải thiện việc đào tạo nội bộ và thúc đẩy CBNV tham gia áp dụng QTTG một cách nhiệt tình và chủ động hơn. Bên cạnh đó, công ty cần mạnh mẽ triển khai truyền thông về các kiến thức, ý tưởng và dự án cải tiến để QTTG được tiếp cận tốt hơn tới CBNV, tăng độ nhận diện của QTTG trong môi trường làm việc.

Bảng 3. Kết quả khảo sát đánh giá về yếu tố Công nhân viên tại công ty

TT	Nội dung	Trả lời “Có”	Trả lời “Không”
2.1	Cán bộ công nhân viên luôn chủ động tham gia vào việc áp dụng QTTG	55	163
2.2	CBNV có trình độ, khả năng tiếp thu nhanh	211	8
2.3	Công nhân viên có khả năng làm việc nhóm tốt	159	59
2.4	Công nhân viên được đào tạo tốt về QTTG	14	204

(Nguồn: Khảo sát tháng 9 năm 2025)

- Về yếu tố chính sách

Các doanh nghiệp truyền thống thường sẽ tập trung vào việc CBNV có hoàn thành kế hoạch, gia tăng sản xuất. Nhưng đối với doanh nghiệp tinh gọn như Esquel Hòa Bình thì công ty cần có một chính sách hoạt động, nhân sự và tài chính chặt chẽ.

Qua kết quả khảo sát có 196 người (90%) trả lời công ty không xây dựng một chính sách đào tạo nội bộ về QTTG cho tất cả CBNV (Bảng 4). 90% khẳng định rằng chiến lược đào tạo về QTTG cho công nhân viên chưa hiệu quả, nguyên nhân có thể đến từ phương thức

đào tạo chưa phù hợp. Phương pháp đào tạo lại đang gặp vấn đề trên ba khía cạnh sau: nội dung đào tạo, cách thức đào tạo và thời gian đào tạo.

Nội dung đào tạo chưa phù hợp, cách thức và thời gian đào tạo chưa thực sự linh hoạt và hấp dẫn, chưa giúp học viên tiếp thu bài nhanh và hiệu quả. Phần lớn hình thức đào tạo mà công ty áp dụng là đào tạo ngoài nơi làm việc, truyền đạt kiến thức. Trong khi đó, việc đào tạo cần đến đào tạo trực tiếp tại chỗ, cầm tay chỉ việc. Do đó, học viên khó có thể triển khai áp dụng một cách bài bản và chia sẻ chính xác kiến thức tới những nhân viên cấp dưới.

Bảng 4. Kết quả khảo sát đánh giá về yếu tố Chính sách

TT	Nội dung	Trả lời “Có”	Trả lời “Không”
3.1	Công ty quản lý hoạt động sản xuất kinh doanh theo quy trình	218	0
3.2	Công ty thiết lập một hệ thống đánh giá kết quả khi tiến hành áp dụng QTTG	195	23
3.3	Công ty có chính sách nhân sự khuyến khích CBNV tham gia áp dụng QTTG	185	33
3.4	Công ty có xây dựng một chiến lược đào tạo nội bộ về QTTG cho tất cả CBNV	22	196
3.5	Có hệ thống quản lý đánh giá hiệu quả làm việc	196	25
3.6	Công ty có triển khai các cải tiến liên tục	210	8

(Nguồn: Khảo sát tháng 9 năm 2025)

- Về yếu tố cơ sở hạ tầng và văn hóa doanh nghiệp

Công ty luôn nỗ lực xây dựng văn

hóa riêng nhằm phát huy sức mạnh của QTTG. Về cơ sở hạ tầng, Esquel Hòa Bình may mắn khi là một nhà máy mới được trang bị đầy đủ máy móc và cập nhật thường xuyên các công nghệ hiện đại.

Thông qua kết quả khảo sát thực tế, điểm tích cực, 204 người (tương đương 94%) tin rằng công ty có cơ cấu tổ chức và văn hóa phù hợp với việc áp dụng QTTG. Tuy nhiên, 178 người (tương đương 82%) cho rằng công ty còn chưa tốt về tiềm lực tài chính cho các dự án cải tiến; 166 người (tương đương 76%) nói rằng công ty còn hạn chế về cơ chế khuyến khích khen thưởng khi áp dụng QTTG. Vấn đề tồn tại lớn nhất của yếu tố chính sách đến từ hệ thống thông tin giao tiếp về QTTG và các dự án cải tiến với 196 người đồng tình (chiếm 90%).

*Bảng 5. Kết quả khảo sát đánh giá về yếu tố cơ sở hạ tầng và văn hóa doanh nghiệp*

TT	Nội dung	Trả lời “Có”	Trả lời “Không”
4.1	Công ty xây dựng một hệ thống thông tin giao tiếp hiệu quả về QTTG và các dự án cải tiến	22	196
4.2	Công ty có cơ chế khuyến khích động viên khen thưởng khi áp dụng thành công QTTG	52	166
4.3	Công ty có tiềm lực tài chính để đầu tư cho các dự án cải tiến	40	178
4.4	Văn hóa và cơ cấu tổ chức công ty phù hợp với việc triển khai QTTG	204	14

(Nguồn: tác giả thực hiện tháng 9 năm 2025)

Ngoài ra tại ở công ty Esquel Hòa Bình còn một số yếu tố khác gồm:

i) Lãnh đạo chưa có cam kết mạnh mẽ và dài hạn, định hướng và tầm nhìn chiến lược khi áp dụng QTTG.

ii) Công nhân viên thiếu sự chủ động tham gia áp dụng QTTG.

iii) Chương trình và phương thức đào tạo QTTG nội bộ chưa phù hợp.

iv) Hệ thống thông tin giao tiếp về QTTG và các dự án cải tiến chưa hiệu quả.

v) Công ty chưa có cơ chế khen thưởng tốt khi áp dụng thành công QTTG.

### **3. Giải pháp hoàn thiện hệ thống quản trị tinh gọn của công ty**

#### **3.1. Về con người**

Giải pháp trọng tâm nhằm tăng cường sự cam kết và tính sở hữu của mỗi cá nhân đối với việc áp dụng quản trị tinh gọn là xây dựng đội ngũ nhân sự có định hướng dài hạn, chủ động và am hiểu về tinh gọn. Lãnh đạo cấp cao giữ vai trò then chốt trong việc định hướng chiến lược, truyền đạt tư duy và hình thành văn hóa tinh gọn trong toàn tổ chức. Quản lý cấp trung và chuyên trưởng đóng vai trò cầu nối, trực tiếp thúc đẩy thay đổi hành vi của người lao động thông qua đào tạo chuyên sâu về kỹ năng lãnh đạo, quản lý nhóm và giải quyết vấn đề.

Bên cạnh đó, doanh nghiệp cần xây dựng niềm tin của công nhân viên đối với lợi ích của QTTG, khuyến khích sự tham gia cải tiến thông qua hệ thống đề xuất ý tưởng đa kênh và cơ chế khen thưởng minh bạch. Việc thành lập Hội đồng Tinh gọn với cơ chế lãnh đạo luân phiên cũng góp phần nâng cao tính chủ động, trách nhiệm và hiệu quả triển khai các dự án cải tiến. Về chương trình đào tạo quản trị tinh gọn

#### **3.2. Về chương trình đào tạo quản trị tinh gọn**

Doanh nghiệp cần xây dựng khung chương trình đào tạo chuẩn về quản trị tinh gọn phù hợp với định hướng phát triển và đặc thù từng cấp bậc nhân sự, gắn với các chỉ tiêu đánh giá kết quả thực hiện (KPIs). Nội dung đào tạo cho công nhân và nhân viên bao gồm các kiến thức nền tảng như 5S, 8 loại lãng phí, quản lý trực quan và các mô hình mô phỏng dòng chảy sản xuất, được tổ chức định kỳ nhằm củng cố nhận thức và kỹ năng thực hành.

Ngoài ra, phương thức đào tạo kết hợp giữa đào tạo tại chỗ (On-the-job training) và đào tạo ngoài công việc (Off-the-job training) giúp tăng cường tính tương tác và gắn kết lý thuyết với thực tiễn sản xuất. Đồng thời, việc phát triển đội ngũ giảng viên nội bộ thông qua chương trình “Be a trainer” góp phần thúc đẩy văn hóa học tập chủ động, khuyến khích chia sẻ tri thức, trao quyền cho người lao động và nâng cao năng lực cải tiến liên tục trong tổ chức.

### 3.3. Về phương pháp áp dụng QTTG

Thứ nhất, công ty cần cải tiến hệ thống thông tin giao tiếp về QTTG. Bên cạnh kết hợp chặt chẽ và nhuần nhuyễn giữa yếu tố con người và đào tạo nội bộ; công ty cần quan tâm nhiều hơn nữa đến các kênh truyền thông nội bộ và cách thức tương tác giữa công ty với CBNV bà ngược lại. Mục đích là để QTTG được tiếp cận sâu rộng và bao quát đến người lao động, để họ được tiếp cận nhiều hơn với QTTG và cùng xây dựng môi trường làm việc và văn hóa tinh gọn.

Thứ hai, đẩy mạnh khuyến khích khen thưởng: hiệu quả hơn cả là có thể xã hội hóa sâu rộng các hoạt động thi đua, đề cao việc khen thưởng. Một trong những yếu tố quan trọng của việc khen thưởng là đề tạo động lực phấn đấu. Thông qua chính

sách khen thưởng rõ ràng, minh bạch và công khai, ban lãnh đạo cho CBNV thấy quyết tâm và cam kết của mình đối với việc áp dụng QTTG, từ đó giúp người lao động tin tưởng hơn vào sự điều hành quản lý của công ty.

Kết hợp với các giải pháp nêu trên: sự cam kết của con người, tính nhất quán và liên kết trong thực hiện các dự án cải tiến sẽ được thay đổi và cải thiện; một hệ thống đào tạo nội bộ chặt chẽ, linh hoạt và hiệu quả sẽ phát huy tính ứng dụng cao trong áp dụng QTTG vào công việc; một hệ thống thông tin mạnh mẽ, truyền tải thông điệp về QTTG; đây sẽ là các giải pháp hiệu quả giúp Esquel Hòa Bình giải quyết được những vấn đề còn đang tồn tại và dần hoàn thiện hệ thống tinh gọn.

### 3.4. Kiểm soát quản trị tinh gọn hiệu quả

Quản trị tinh gọn là một quá trình cải tiến liên tục nhằm phát hiện sai lệch và áp dụng các công cụ quản trị để nâng cao hiệu quả công việc. Trên cơ sở đó, các tiêu chuẩn được thiết lập như hướng dẫn công việc mới, chuẩn hóa quy trình, thông số và bảng theo dõi nhằm tăng cường kiểm soát hoạt động. Việc áp dụng công cụ phòng chống sai lỗi (Poka-Yoke) giúp bảo đảm các giải pháp được thực hiện chính xác. Để triển khai QTTG hiệu quả, doanh nghiệp cần xây dựng kế hoạch chiến lược dài hạn, có sự cam kết của lãnh đạo cấp cao, triển khai từng bước với quy mô phù hợp và nhận được sự tư vấn chuyên môn từ các chuyên gia.

## 4. Kết luận

“Sản xuất tinh gọn” được phát triển mạnh mẽ tại Công ty Toyota từ những năm 1950 với tên gọi “Hệ thống sản xuất Toyota” (Toyota Production System – TPS). Đây là phương thức quản lý nhằm

tạo ra giá trị cao nhất cho khách hàng thông qua việc nhận diện và loại bỏ các lãng phí trong quá trình sản xuất. Việc áp dụng sản xuất tinh gọn góp phần xây dựng môi trường làm việc an toàn, nâng cao hiệu suất, đảm bảo chất lượng sản phẩm, rút ngắn thời gian giao hàng, giảm chi phí và gia tăng sự hài lòng của khách hàng, từ đó giúp doanh nghiệp nâng cao năng lực cạnh tranh trên thị trường.

Trải qua 10 năm xây dựng và phát triển, cùng với chiến lược áp dụng QTTG từ những ngày đầu, công ty Esquel Hòa Bình đã đạt được những thành tựu đáng kể. Tuy nhiên, với sự lớn mạnh và mở rộng quy mô công ty, sự thay đổi của thị trường, công ty Esquel Hòa Bình gặp phải một số vấn đề trong vận hành sản xuất kinh doanh. Các phương pháp nghiên cứu định tính, thu thập dữ liệu sơ cấp thứ cấp, thực hiện khảo sát, nghiên cứu tài liệu, thực hiện dự án cải tiến thực tế,... đã được sử dụng để tìm ra những yếu tố thành công và nguyên nhân thất bại khi áp dụng QTTG tại công ty. Thông qua bài viết này, tác giả đã giải quyết được các mục tiêu sau: hệ thống cơ sở lý luận và thực tiễn về áp dụng QTTG tại doanh nghiệp, đánh giá thực trạng và phân tích các yếu tố ảnh hưởng đến QTTG tại công ty giai đoạn 2019-2025. Từ đây, tác giả đề xuất giải pháp để hoàn thiện và tăng cường áp dụng QTTG hiệu quả tại công ty Esquel Hòa Bình; nhằm khắc phục những hạn chế, tiến tới hoàn thiện hệ thống và áp dụng hiệu quả QTTG trong những năm tới.

Qua nghiên cứu, công ty khi muốn áp dụng QTTG đều cần phải lưu ý một số điểm sau: i) nhận thức rõ ràng về lợi ích của tinh gọn mang lại chính là yêu cầu không chỉ đối với ban lãnh đạo mà còn là yêu cầu đối với toàn cấp quản lý

và CBNV, ii) công ty cần xây dựng một chiến lược QTTG dài hạn với dự trù các nguồn lực như nguồn vốn, nguồn nhân lực, cơ sở vật chất phù hợp để đảm bảo quá trình thực hiện dự án, iii) ban lãnh đạo của công ty cần có sự quan tâm và chú trọng vào công tác kiểm tra, giám sát để việc áp dụng và xây dựng văn hóa “tinh gọn” trong môi trường làm việc.

Với những vấn đề được nêu lên trong nghiên cứu, cùng với các giải pháp, đặc biệt là các bài học kinh nghiệm khi tác giả triển áp QTTG tại công ty Esquel Hòa Bình.

Thứ nhất, nghiên cứu được thực hiện tại công ty Esquel Hòa Bình, có tham khảo tài liệu đào tạo và kinh nghiệm quản lý thực tế tại hai nhà máy Esquel Đồng Nai và Esquel Bình Dương. Tuy nhiên, các công ty này đều thuộc quản lý của tập đoàn Esquel, do đó khả năng tổng quát hóa kết quả nghiên cứu chưa cao vì mỗi ngành khác nhau sẽ có thể có những đặc điểm riêng biệt khi nghiên cứu.

Hệ thống QTTG được nghiên cứu và xây dựng là áp dụng riêng và mang đậm nét văn hóa Esquel Hòa Bình. Nếu như muốn áp dụng nghiên cứu này tại công ty khác, thậm chí là công ty cùng ngành may mặc khác thì sẽ phải mất nhiều thời gian để điều chỉnh lại.

Thứ hai, bài nghiên cứu chỉ đề cập đến ba nguyên nhân gốc rễ tồn tại và bốn yếu tố chính ảnh hưởng đến hiệu quả áp dụng QTTG. Điều này có thể cải thiện ở các nghiên cứu tiếp theo trong tương lai để có thể có nhiều nghiên cứu khác đề cập tới nhiều yếu tố hơn để đạt được mức độ tổng quát hóa cao hơn.

---

## TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt:

1. Lê Thị Vân Dung (2022), Các yếu tố ảnh hưởng đến hiệu quả áp dụng quản trị tinh gọn tại doanh nghiệp sản xuất hàng xuất khẩu tại Việt Nam, Đại học Ngoại thương đăng trên Tạp chí Kế toán và Kiểm toán số tháng 7/2022.

2. Bùi Thị Thu Hằng (2016), Áp dụng quản trị tinh gọn trong hoạt động mua sắm tập trung tại ngân hàng TMCP Quân Đội, ĐH QG Hà Nội, luận văn thạc sĩ.

3. Nguyễn Đăng Minh (2015), Quản trị tinh gọn tại Việt Nam – Đường tới thành công. Hà Nội: Nhà xuất bản Đại học Quốc Gia Hà Nội.

4. Nguyễn Đạt Minh và cộng sự (2017), Các nhân tố tác động đến việc áp dụng QTTG trong các doanh nghiệp tại Việt Nam. Tạp chí Những vấn đề kinh tế và chính trị thế giới, (Số 12), trang 2-3.

5. Khánh Lan (2022), Lean Manufacturing: Xây dựng ba thương hiệu lớn nhất thế giới, Speedmaint, truy cập 8/3/2022, <https://speedmaint.com/lean-manufacturing-xay-dung-thuong-hieu/>

6. Lê Thị Kim Oanh (2015), Áp dụng sản xuất tinh gọn – Lean vào công ty CP May Sài Gòn năm 2015. TC Khoa học Công nghệ và Thực phẩm, p 147-148.

7. Phòng CPSI (2020). May 10: Đầu tư quản lý sản xuất và chất lượng sản phẩm là chìa khóa để thành công, SCP, truy cập 21/2/2023, <https://scp.gov.vn/tin-tuc/t2142/may-10-dau-tu-quan-ly-san-xuat-va-chat-luong-san-pham-la-chia-khoa-de-thanh-cong.html>

8. Phan Thị Thu Sương (2020), Nghiên cứu hoàn thiện việc áp dụng quản trị tinh gọn tại công ty TNHH Jia Hsin, Trường ĐH Kinh tế TP HCM, luận văn thạc sĩ.

9. VP NSCL (2020). Ngành dệt may

Việt Nam đẩy mạnh áp dụng sản xuất tinh gọn Lean nhằm nâng cao năng suất và sức cạnh tranh, VP Năng Suất Chất Lượng, Vụ Khoa học và Công nghệ, truy cập ngày 21/2/2023, <https://khcncongthuong.vn/tin-tuc/t7614/nganh-det-may-viet-nam-day-manh-ap-dung-san-xuat-tinh-gon-lean-nham-nang-cao-nang-suat-va-suc-can-phanh.html>

10. VietQ.vn (2020). Hiệu quả nâng cao năng suất nhờ áp dụng Lean tại may Nhà Bè, TBT AN GIANG, truy cập ngày 21/2/2023, <http://tbtagi.angiang.gov.vn/hieu-qua-nang-cao-nang-suat-nho-ap-dung-lean-tai-may-nha-be-9381.html#:~:text=Nh%E1%BB%9D%20%C3%A1p%20d%E1%BB%A5ng%20Lean%20m%C3%A0,t%C4%83ng%20t%E1%BB%AB%2010%2D15%25.>

#### Tài liệu tiếng Anh

1. Eduardo L. Garcia (2016). Summary of Lean Manufacturing/ Management V3.6 (2016), retrieved 21 February 2023 from <https://www.slideshare.net/enateduardo/the-house-of-lean-v35-kki>

2. Mark Onetto (2011). Lean and Sigsixma at Amazon.com, retrieved 8 March 2023 from <https://opexlearning.com/resources/marc-onetto-lean-six-sigma-at-amazon/8042/>

3. Ohno, T. (1988). *Toyota production system: beyond large-scale production*. Portland, Oregon: Productivity Press.

4. Womack, J.P., Jones, D.T. & Roos, D. (1990). *The machine that changed the world*. New York: Rawson Associates.

5. Womack, James P.; Jones, Daniel T. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. Taylor & Francis Group.

# DU LỊCH VĂN HÓA VÀ NGHỆ THUẬT BIỂU DIỄN: MỐI QUAN HỆ NHÌN TỪ GÓC ĐỘ NGÀNH CÔNG NGHIỆP VĂN HÓA

Lê Thị Hoài Phương

GS.TS. Viện Văn hóa, Nghệ thuật, Thể thao và Du lịch Việt Nam

Email: bachduong.786@gmail.com

## TÓM TẮT

Đối tượng nghiên cứu của công trình này là mối quan hệ giữa Du lịch Văn hóa (DLVH) và Nghệ thuật Biểu diễn (NTBD) nhìn từ góc độ các ngành công nghiệp văn hóa (CNVH) Việt Nam. Mục đích của nghiên cứu là để khẳng định luận điểm: DLVH và NTBD có mối quan hệ tương hỗ, “ràng buộc” lẫn nhau, cần hợp tác với nhau để hoàn thành vai trò của các ngành CNVH mũi nhọn trong Chiến lược phát triển CNVH đến năm 2030, tầm nhìn 2045. Tác giả nêu ra một số câu hỏi mang tính giả định, rồi thông qua phương pháp tiếp cận liên ngành (sử học, văn hóa học, du lịch học, nghệ thuật học, khoa học quản lý) để lần lượt trình bày các vấn đề cốt lõi xoay quanh mối quan hệ của DLVH và NTBD như: những đặc điểm riêng của mỗi lĩnh vực; những điểm tương đồng và khác biệt giữa DLVH và NTBD; thực trạng hợp tác giữa DLVH và NTBD với sự đánh giá các mặt tích cực cũng như hạn chế; những cơ hội và thách thức mà DLVH và NTBD đang đối diện... Từ đó tác giả đề xuất một số giải pháp để DLVH và NTBD cần làm trong những năm tới. Nội dung nghiên cứu và kết quả nghiên cứu là thống nhất với luận điểm đã nêu.

**Từ khóa:** Công nghiệp văn hóa, Du lịch văn hóa, Nghệ thuật biểu diễn

## ABSTRACT

The subject of this study is the relationship between Cultural Tourism and Performing Arts, examined from the perspective of Vietnam's cultural industries. The purpose of the research is to affirm the argument that Cultural Tourism and Performing Arts have a mutually supportive and interdependent relationship; they are “bound” to one another and need to collaborate closely in order to fulfill their roles as key sectors in the Cultural Industries Development Strategy to 2030, with a vision to 2045.

The author raises several hypothetical research questions and, through an interdisciplinary approach—drawing on history, cultural studies, tourism studies, arts studies, and management science—systematically presents the core issues surrounding the relationship between Cultural Tourism and Performing Arts. These include: the distinctive characteristics of each field; the similarities and differences between Cultural Tourism and Performing Arts; the current state of cooperation between the two sectors, along with an assessment of both strengths and limitations; as well as the opportunities and challenges that Cultural Tourism and Performing Arts are currently facing.

Based on this analysis, the author proposes a number of solutions for both Cultural Tourism and Performing Arts to implement in the coming years. The research content and findings are consistent with the initial argument presented.

**Keywords:** Cultural industries; Cultural tourism; Performing arts

Ngày nhận bài: 02/03/2026; Ngày sửa bài: 30/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. Đặt vấn đề

Du lịch vốn là một hoạt động tự thân của con người có từ xa xưa để thỏa mãn nhu cầu rất tự nhiên là muốn tìm hiểu, khám phá thiên nhiên và xã hội. Trải qua bao biến thiên của xã hội loài người, du lịch từ chỗ khởi đầu là hoạt động mang tính cá nhân, hoặc của một nhóm, hay một tập thể người có chung sở thích trong hoạt động du lịch, càng về sau du lịch càng có tính tổ chức cao hơn, phát triển hơn về quy mô và phong phú hơn về loại, kiểu, hình thức du lịch...

Trong xã hội hiện đại, từ lâu du lịch được gọi là “ngành công nghiệp không khói”, được nhiều quốc gia coi trọng là ngành kinh tế mũi nhọn, vì có đóng góp rất lớn trong tỷ trọng GDP và việc làm.

Ở Việt Nam, do đặc điểm lịch sử của đất nước phải trải qua nhiều cuộc chiến tranh nối tiếp liên miên, hoạt động du lịch trước đây rất bị hạn chế. Chỉ từ sau khi đất nước hoàn toàn thống nhất (1975), trong cuộc sống hòa bình thì du lịch mới bắt đầu có cơ hội hoạt động. Từ khi Đảng và Nhà nước chú trọng phát triển các ngành công nghiệp văn hóa (CNVH), ban hành Quyết định số 1755/QĐ-TTg, *phê duyệt Chiến lược phát triển các ngành công nghiệp văn hóa Việt Nam đến năm 2020, tầm nhìn đến năm 2030* (ngày 08/09/2016) [6], ngành Du lịch có tuổi đời non trẻ được đưa vào danh sách 12 ngành CNVH Việt Nam, với tư cách là một ngành kinh tế mũi nhọn.

Gần đây nhất, Thủ tướng Phạm Minh Chính đã ký Quyết định số 2486/QĐ-TTg, *Phê duyệt Chiến lược phát triển công nghiệp văn hóa đến năm 2030, tầm nhìn 2045*, ngày 14/11/2025 [7] (sau đây gọi là *Chiến lược phát triển công nghiệp văn hóa*). Chiến lược mới đã xác định 10 ngành CNVH chủ chốt, bao gồm: Điện ảnh; Mỹ thuật – nhiếp ảnh – triển lãm; Nghệ thuật biểu diễn; Phần mềm & trò chơi giải trí; Quảng cáo; Thủ công mỹ nghệ; Du lịch văn hóa; Thiết kế sáng tạo; Truyền hình &

phát thanh; Xuất bản.

Trong khuôn khổ của bài viết này, chúng tôi tập trung nghiên cứu về lĩnh vực Du lịch văn hóa (DLVH) và Nghệ thuật biểu diễn (NTBD), với tư cách là hai ngành CNVH được Nhà nước chú trọng phát triển, vì chúng tôi nhận thấy hoạt động của hai lĩnh vực này có mối quan hệ tương hỗ lẫn nhau, cùng có những cơ hội và thách thức khá là tương đồng trong bối cảnh xã hội hiện nay. Hoạt động biểu diễn nghệ thuật cũng như du lịch văn hóa vừa có giá trị đóng góp tăng trưởng kinh tế đất nước, vừa góp phần bảo tồn các giá trị văn hóa, nghệ thuật dân tộc, vừa quảng bá bản sắc văn hóa và hình ảnh Việt Nam trên thế giới.

## 2. Phương pháp nghiên cứu và câu hỏi nghiên cứu

### 2.1. Phương pháp nghiên cứu

Phương pháp nghiên cứu được áp dụng là phương pháp nghiên cứu liên ngành, trong đó chủ yếu là sử học, văn hóa học, nghệ thuật học, du lịch học, kinh tế học và quản lý văn hóa nghệ thuật.

### 2.2. Câu hỏi nghiên cứu

Công trình nghiên cứu này nhằm lý giải một số vấn đề mang tính cốt lõi:

- Vì sao là DLVH và NTBD, mà không phải là với ngành nào khác trong CNVH Việt Nam?

- Cơ sở nào để hợp tác giữa DLVH và NTBD là khả thi?

- Trong thực tế, DLVH và NTBD Việt Nam đã kết nối, hợp tác đến mức độ nào?

- Những cơ hội và thách thức mà DLVH và NTBD đang đối diện để thực hiện mục tiêu của *Chiến lược phát triển công nghiệp văn hóa đến năm 2030, tầm nhìn 2045*?

- Giải pháp nào cho những năm sắp tới?

## 3. Cơ sở lý luận và thực tiễn – Lịch sử vấn đề

### 3.1. Các phạm trù liên quan

#### 3.1.1. Du lịch – Du lịch văn hóa

“Du lịch”, hay hoạt động du lịch, hiểu theo nghĩa rộng đã có nghĩa là “du lịch văn hóa” rồi. Từ thời bắt đầu có những hoạt

động du lịch “sơ khai” nhất của con người - kiểu du lịch mang tính chất tìm hiểu, tham quan, khám phá một vùng đất mới, một đất nước xa lạ, một nền văn hóa khác biệt,... dù có chủ đích hay không thì cũng đều hàm chứa yếu tố văn hóa chủ yếu trong đó.

Càng về sau, khi du lịch là nhu cầu không thể thiếu trong đời sống của con người thì hoạt động du lịch càng mang tính tổ chức cao, có kế hoạch, mục tiêu, mục đích cụ thể. Từ đó, hoạt động du lịch lại được phân loại ra nhiều hình thức, kiểu, loại, đáp ứng nhu cầu đa dạng của người du lịch. Cho đến nay, có thể kể tên hàng loạt kiểu du lịch của con người như: Du lịch khám phá - mạo hiểm; Du lịch sinh thái; Du lịch nghỉ dưỡng; Du lịch chăm sóc sức khỏe; Du lịch ẩm thực (Food Tourism); Du lịch văn hóa – lịch sử; Du lịch học tập, nghiên cứu; v.v... Trong các kiểu, loại du lịch thì DLVH là hoạt động phong phú hơn cả, liên quan tới hầu hết các lĩnh vực đời sống xã hội của một đất nước, tộc người hay cộng đồng cư dân.

Xét ở một phạm vi hẹp hơn, cụ thể trong danh mục 10 ngành CNVH mũi nhọn của Việt Nam thì DLVH là một ngành độc lập, có vị thế ngang hàng với các ngành còn lại trong danh sách.

### 3.1.2. Nghệ thuật Biểu diễn

Nghệ thuật biểu diễn Việt Nam bao gồm các lĩnh vực nghệ thuật cơ bản như: sân khấu (kịch); nghệ thuật múa rối; âm nhạc; múa; xiếc... Trong đó, nhiều thể loại vốn khởi nguồn là nghệ thuật dân gian, sau quá trình phát triển lâu dài đã trở thành chuyên nghiệp như nghệ thuật Tuồng (hát Bội), Chèo, Cải lương, dân ca kịch, múa rối nước...

Về phương diện quản lý nhà nước, trong thời kỳ bao cấp (từ 1986 trở về trước) các đơn vị NTBD chuyên nghiệp trực thuộc quyền quản lý của Nhà nước, từ cấp trung ương đến địa phương. Từ khi thực hiện chính sách đổi mới, Nhà nước khuyến khích phát triển thành phần kinh tế tư nhân,

ngày càng xuất hiện nhiều đơn vị NTBD tư nhân (nhiều nhất là ở miền Nam). Từ đây xuất hiện các cụm từ “đơn vị nghệ thuật công lập” và “đơn vị nghệ thuật ngoài công lập”. Hình thức tổ chức và quy mô của các đơn vị nghệ thuật ngoài công lập (dân lập) khá phong phú: một số đơn vị hoạt động biểu diễn chuyên nghiệp, số khác là đơn vị không chuyên, hoạt động dưới hình thức câu lạc bộ, do người dân tự lập ra, tự nuôi dưỡng, tự tổ chức biểu diễn, để phục vụ người dân địa phương trong những dịp lễ tết, hội hè, đồng thời cũng là để tiếp nối, gìn giữ truyền thống nghệ thuật của cha ông truyền lại... [4]

Sở dĩ chúng tôi đề cập tới hoạt động biểu diễn nghệ thuật không chuyên tại các địa phương vì câu chuyện của chúng ta đang nói tới DLVH. Trên giải đất hình chữ S Việt Nam hầu như ở vùng miền nào cũng có một hoặc vài loại hình NTBD dân gian đặc sắc, đậm đà bản sắc văn hóa riêng của từng địa phương. Bàn về mối quan hệ giữa DLVH và NTBD thì không thể không nói tới sự phong phú, đa dạng và đặc sắc của các hình thức nghệ thuật dân gian đang hiện hữu trong cộng đồng các dân tộc Việt Nam trên khắp ba miền Bắc, Trung, Nam, vì đây là nguồn sản phẩm đầy tiềm năng cho hoạt động DLVH ở Việt Nam.

### 3.2. Sự tương quan Du lịch văn hóa và Nghệ thuật biểu diễn

Nếu xem xét DLVH và NTBD từ góc độ ngành CNVH, có thể thấy hai lĩnh vực này có một số đặc điểm tương đồng, đồng thời chúng có các đặc tính khác nhau, nhất là về con đường phát triển để định hình là một ngành công nghiệp sáng tạo của Việt Nam.

#### 3.2.1. Những điểm tương đồng

“Công nghiệp văn hóa” theo định nghĩa của UNESCO là: “các ngành công nghiệp kết hợp sự sáng tạo, sản xuất và khai thác các nội dung có bản chất phi vật thể và văn hóa. Các nội dung này thường được bảo vệ bởi luật bản quyền và thể hiện dưới dạng các sản phẩm hay dịch vụ” [8]. Nói rõ hơn,

CNVH được hình thành từ sự kết hợp của sự sáng tạo, quá trình sản xuất và phân phối hàng hóa, dịch vụ thuộc về văn hóa và thường được các quyền sở hữu trí tuệ bảo vệ.

Điểm chung lớn nhất của DLVH và NTBD là đều hướng tới khai thác và sáng tạo các nội dung đậm tính văn hóa, tạo nên các sản phẩm có bản chất phi vật thể, truyền bá tri thức và các giá trị văn hóa, nghệ thuật (truyền thống và hiện đại), nhằm làm phong phú thêm đời sống tinh thần của con người.

Nhìn chung, ngành nghề nào cũng là phục vụ con người, chỉ khác nhau ở *cách thức phục vụ*, hay nói một cách khác là ở *phương thức tiêu thụ sản phẩm*. Có những ngành phục vụ nhu cầu của con người một cách gián tiếp, và có những ngành phục vụ một cách trực tiếp. DLVH và NTBD giống nhau ở chỗ đều *phục vụ trực tiếp con người*, tiếp xúc trực tiếp với đối tượng phục vụ của mình, đưa sản phẩm đến trực tiếp cho khách hàng của mình (đối với ngành du lịch thì đó là du khách, đối với ngành NTBD thì đó là khán giả).

Từ đặc điểm phục vụ trực tiếp khách hàng của mình, DLVH và NTBD dẫn tới hai ngành này có nhiều điểm tương đồng, ở một số khâu trong quy trình khai thác, sản xuất, sáng tạo và phân phối sản phẩm đến khách hàng, đặc biệt là trong việc xây dựng chiến lược marketing và thực hiện kế hoạch của đơn vị mình.

### 3.2.2. Quan hệ “đối tác” và “đồng minh”

Trong 10 ngành CNVH mũi nhọn của Việt Nam có một số ngành thuộc lĩnh vực nghệ thuật, đó là: *Nghệ thuật Biểu diễn, Điện ảnh, Mỹ thuật – nhiếp ảnh – triển lãm*. Trong đó, ngành có liên quan một cách trực tiếp và chặt chẽ với DLVH chính là NTBD. Bởi sản phẩm của NTBD cũng là sản phẩm của DLVH, gắn liền với hoạt động DLVH, trong DLVH có NTBD.

Từ những yếu tố tương đồng và gắn kết

của DLVH và BDNT, dưới góc độ kinh tế, có thể nói giữa DLVH và NTBD có mối quan hệ vừa là “đối tác” vừa là “đồng minh”.

### 3.2.3. Những điểm khác biệt

Một trong những điểm khác nhau giữa Du lịch và NTBD là: nếu như trên con đường phát triển của mình ngành Du lịch sớm định hình là một ngành kinh tế thực thụ, thì NTBD đã trải qua nhiều thế kỷ chỉ được xem là để giải trí, chứ không có giá trị kinh tế. Đây là vấn đề mang tính lịch sử. Cái quan niệm cho rằng các hình thức nghệ thuật diễn xướng chỉ là để “mua vui” đã ăn sâu vào tiềm thức của nhiều lớp người trong chế độ cũ. Đến khi có sự lãnh đạo của Đảng Cộng sản Việt Nam thì có sự nhận thức mới về vai trò của văn học, nghệ thuật. Năm 1943 Đảng đã ban hành *Đề cương về văn hóa Việt Nam*, trong đó văn hóa – tư tưởng được coi là một trong ba mặt trận quan trọng, ngang hàng với kinh tế và chính trị, cùng nhau hợp lực tiến hành cuộc Cách mạng giải phóng dân tộc dưới sự lãnh đạo của Đảng Cộng sản Việt Nam [1]. Xuyên suốt thời kỳ chiến tranh cũng như sau khi đất nước đã hoàn toàn thống nhất (1975), cho đến giữa thập niên 1980, Đảng ta vẫn luôn coi văn hóa nghệ thuật (VHNT) là “vũ khí tư tưởng”, có giá trị tinh thần, chứ không phải có giá trị kinh tế, và sản phẩm của VHNT không phải là hàng hóa (với ý nghĩa vật chất)... Phải từ sau Đại hội Đảng lần thứ VI (1986), khi Đảng và Nhà nước tiến hành đổi mới toàn diện đất nước, thì sự nhìn nhận về vai trò của VHNT mới có sự thay đổi về chất. Đặc biệt, Nghị quyết Trung ương 5, khóa VIII (ban hành tháng 7 năm 1998) của Đảng đã khẳng định: “Văn hóa vừa là mục tiêu, vừa là động lực thúc đẩy phát triển xã hội” [2]. Sự đổi mới cách nhìn nhận về vai trò của VHNT như một lĩnh vực kinh tế không phải diễn ra suôn sẻ và được chấp nhận nhanh chóng. Ngay cả đối với những người

trong ngành văn hóa, từ các nhà quản lý đến các nghệ sĩ - những người đã quen sống và làm việc trong cơ chế bao cấp kéo dài hơn 40 năm (1945 – 1986), nay phải làm quen với “cung cách làm ăn” của cơ chế kinh tế thị trường, khiến họ vô cùng hoang mang, có lúc mất phương hướng... Bởi vì các nhà quản lý các đơn vị NTBD chưa hề được đào tạo quản lý nghệ thuật (QLNT) trong cơ chế thị trường, phải tự giải bài toán thu – chi cho một nhà hát, đoàn nghệ thuật có nơi đến hàng trăm con người với bộ máy công kênh... Họ càng không biết thế nào là marketing nghệ thuật, vì trong cơ chế bao cấp hàng năm họ được cấp ngân sách nhà nước, nhà hát chỉ cần hoàn thành nhiệm vụ (biểu diễn phục vụ chính trị và nhân dân) theo chỉ tiêu hàng năm trên giao...[4]. Bước sang thế kỷ XXI, sau vài chục năm “tập dượt”, học hỏi kinh nghiệm của các nước phát triển, cho đến nay các nhà quản lý của các đơn vị nghệ thuật, chủ yếu là các nhà hát cấp trung ương, đã quen dần với cách quản lý mới, chủ động hơn trong cơ chế tự chủ về tài chính mà Nhà nước đã “cởi trói”, một số nhà hát đã chủ động hơn trong việc tổ chức biểu diễn, chủ động tìm đến khán giả thông qua nhiều hình thức marketing như quảng cáo, bán vé online, chủ động tìm kiếm nguồn tài trợ, v.v...[5]

Sở dĩ chúng tôi thấy cần phải nhắc lại một thời khó khăn của ngành NTBD ở vài thập niên đầu tiên của công cuộc đổi mới của đất nước là để người đọc hình dung được ngành NTBD Việt Nam đã, đang và sẽ còn khó khăn như thế nào để “nhập cuộc” với các ngành CNVH khác của nước nhà.

Xét riêng trong lĩnh vực nghệ thuật, ngành Điện ảnh có ưu thế hơn cả với tư cách là một ngành CNVH có thể đạt được những lợi ích kinh tế lớn. Trong khi đó, do đặc thù loại hình, NTBD cho tới thời điểm hiện tại vẫn đang phải xoay xở để nuôi sống nghệ sĩ và cố gắng duy trì hoạt động nghề một cách thường xuyên, vì vậy,

cái chỉ tiêu mà *Chiến lược phát triển công nghiệp văn hóa* (ban hành năm 2025) đặt ra cho ngành NTBD là phải đạt được con số 10,834 nghìn tỷ đồng (tương đương 430 triệu USD) doanh thu để đóng góp vào mục tiêu chung của toàn bộ ngành CNVH là đến năm 2030 đóng góp 7% GDP quốc gia thực sự là một thử thách lớn của ngành NTBD trong 5 năm tới.[7]

### **3.3. Thực trạng hợp tác Du lịch văn hóa và Nghệ thuật biểu diễn**

#### **3.3.1. Thực trạng hợp tác**

Nhiều năm về trước vấn đề hợp tác giữa DLVH và NTBD không phải chưa được các nhà quản lý ngành văn hóa nghĩ tới, cũng không phải đến nay chưa có sự kết hợp nào giữa các công ty du lịch và đơn vị NTBD. Tuy nhiên, trên thực tế sự hợp tác này mới chỉ diễn ra chủ yếu thông qua các tour du lịch đơn lẻ, mang tính chất thời vụ, còn đối với các đơn vị NTBD, nhất là biểu diễn chuyên nghiệp thì còn nhiều vấn đề nan giải.

Hàng chục năm trước, Nhà hát Tuồng Việt Nam là một trong những đơn vị NTBD hiếm hoi đã rất cố gắng kết nối với các công ty du lịch để mong muốn hợp tác. Lãnh đạo nhà hát đã mời các nhà quản lý một số đơn vị du lịch đến nhà hát để xem các nghệ sĩ biểu diễn, giới thiệu các chương trình nghệ thuật đặc sắc, để tăng thêm sự hiểu biết lẫn nhau... Tuy nhiên phía du lịch tỏ ra không mặn mà tiến tới. Dẫu vậy, Nhà hát Tuồng Việt Nam vẫn xây dựng chương trình biểu diễn phục vụ khách du lịch, với lịch cố định là mỗi tuần dành hai đêm thứ Năm và thứ Bảy, diễn tại Rạp Hồng Hà, gồm các trích đoạn Tuồng truyền thống đặc sắc, có phụ đề bằng bốn thứ tiếng Anh, Pháp, Đức và Tây Ban Nha. Dẫu rằng Ban lãnh đạo và các nghệ sĩ tài năng của Nhà hát đã rất cố gắng, có những suất diễn chỉ có vài ba khách nhưng các nghệ sĩ vẫn biểu diễn phục vụ...Hầu như các suất diễn phục vụ du lịch nhà hát đều phải bù lỗ. Cuối cùng nhà hát đã phải dừng chương trình này.

Ở Hà Nội, chỉ có Nhà hát Múa rối Thăng Long là được hưởng lợi hơn cả từ khách du lịch do có vị trí đắc địa nằm ngay bên Hồ Hoàn Kiếm. Vị trí biểu diễn thuận lợi và loại hình nghệ thuật độc đáo là hai yếu tố cơ bản giúp nhà hát này không phải bỏ nhiều công sức và tài chính để quảng cáo, làm marketing, mà vẫn thu hút rất đông khách du lịch quốc tế cũng như quốc nội, với tần suất trung bình mỗi ngày có 5 – 8 suất diễn, và diễn liên tục quanh năm suốt tháng, suất diễn nào cũng đông khán giả.

Để khai thác những nét văn hóa độc đáo của đời sống Hà Nội về đêm, gần đây ở Hà Nội còn có tour đêm với tên gọi là *Tinh hoa Bắc Bộ* - vừa là tour văn hóa vừa là show diễn – một chương trình nghệ thuật thực cảnh, kết hợp âm nhạc, múa, diễn xuất ngoài trời cho du khách khám phá Hà Nội vào buổi tối.

Ở Thành phố Hồ Chí Minh, sự gắn kết các tour du lịch với nghệ thuật biểu diễn được tổ chức đa dạng và phong phú hơn, thường là các tour combo kết hợp ẩm thực – tham quan – xem biểu diễn nghệ thuật (hoặc là sự kiện văn hóa, lễ hội truyền thống)... Một số tour du lịch khá nổi tiếng như: *Chào Show Saigon, A O Show*, tour xem Múa rối nước Golden Dragon; Tour tổng hợp (tham quan thành phố + show biểu diễn nghệ thuật + ăn tối), v.v...

Ở các thành phố lớn như Huế, Đà Nẵng, Hội An... các tour du lịch cũng thường kết hợp nhiều hoạt động liên quan tới nhiều hình thức giải trí khác nhau và các sản phẩm VHNT khác nhau, chủ yếu là kết hợp tham quan đi tích + biểu diễn nghệ thuật truyền thống (diễn Tuồng ở Nhà hát Nguyễn Hiền Đình, Nhã nhạc cung đình Huế ở Thành Nội và Ca Huế trên sông Hương ở Huế...) + ẩm thực.

Ngoài ra, còn có một số tour du lịch đi về các địa phương, nơi có những di tích lịch sử hay thắng cảnh thiên nhiên nổi tiếng. Tiêu biểu như: tour du lịch ở Quảng Ninh kết hợp tham quan hang động Ngọc Rồng với

biểu diễn nghệ thuật thực cảnh - kể chuyện văn hóa và truyền thuyết địa phương, tạo trải nghiệm du lịch độc; ở Bình Thuận có Lễ hội Nghệ thuật biểu diễn quốc tế, là sự kiện thường niên, thu hút nhiều đoàn nghệ thuật trong và ngoài nước, là sự kiện văn hóa quy mô lớn kết hợp với tour du lịch biển Phan Thiết – Mũi Né.

3.3.2. *Đánh giá thực trạng hợp tác giữa Du lịch văn hóa và Nghệ thuật Biểu diễn trong bối cảnh đất nước đổi mới*

#### **Mặt tích cực**

Từ khi đất nước đổi mới, nền kinh tế chuyển đổi sang cơ chế thị trường thông thoáng hơn trước, Nhà nước chủ trương mở cửa đón khách quốc tế đến với Việt Nam thì ngành du lịch đã thể hiện có nhiều tìm tòi, sáng tạo trong hoạt động quảng bá, giới thiệu những vẻ đẹp về thiên nhiên cũng như về văn hóa của các vùng miền từ Bắc chí Nam, vì vậy đã thu hút mạnh mẽ du khách quốc tế và trong nước đến với Việt Nam ngày càng nhiều.

Thực trạng cho thấy đã có sự kết nối, phối kết hợp giữa các công ty du lịch với các cấp chính quyền địa phương để xây dựng các sản phẩm du lịch dựa trên thế mạnh của từng vùng, miền, tạo thành chuỗi sự kiện của các tour du lịch. Trong các sản phẩm ấy, hầu như đều có hoạt động biểu diễn nghệ thuật, giới thiệu các hình thức nghệ thuật dân gian, truyền thống.

Theo hướng làm này, năm năm trở lại đây, hình thức du lịch “homestay” ngày càng phổ biến ở nhiều địa phương, nhất là ở các tỉnh miền núi phía Bắc và một số địa phương thuộc miền Tây Nam Bộ... Loại hình du lịch này ngày càng được nhiều du khách quốc tế lựa chọn, sẵn sàng kéo dài thời gian lưu trú để thực sự trải nghiệm các nét văn hóa cổ truyền của thôn quê Việt Nam. Hình thức du lịch “homestay” đang góp phần đáng kể vào cải thiện đời sống kinh tế và tinh thần của người dân bản địa.

Cho đến nay đã có một số công ty du

lịch ở Việt Nam được đánh giá là uy tín và mạnh về các tour du lịch văn hóa, như: *Asiatica Travel*, *Vietnam Cultural Tours*, *Công ty TNHH Du lịch Văn hóa Sài Gòn*, *Công ty TNHH Thương mại Dịch vụ Du lịch Sao Việt*, *Công ty Cổ phần Du lịch Hòa Bình Việt Nam*, *Viet Vision Travel* (thuộc *Viet Vision MICE*), v.v... Đặc điểm chung của các công ty – lữ hành này là tổ chức các tour du lịch kết hợp nhiều sản phẩm văn hóa: tham quan danh thắng, di sản lịch sử - văn hóa, ẩm thực, giao lưu cộng đồng bản địa, xem biểu diễn nghệ thuật dân gian truyền thống, tham gia lễ hội... Tùy vào nhu cầu cụ thể của từng nhóm, thậm chí của cá nhân du khách mà các công ty lữ hành tổ chức các tour phù hợp...

Đây là một hướng đi đúng, khai thác và phát huy thế mạnh của đất nước ta, nơi có nhiều danh lam thắng cảnh, di tích lịch sử, hầu như ở địa phương nào cũng có một vài hình thức nghệ thuật dân ca, dân vũ độc đáo, có bản sắc riêng.

### **Những hạn chế**

Có thể thấy những năm qua ngành du lịch đã có nhiều cố gắng trong việc tìm hướng đi mới để đáp ứng được nhu cầu ngày càng lớn của thị trường, nhưng dù sao ở Việt Nam thì du lịch là một ngành non trẻ, nên còn nhiều hạn chế trong hoạt động, chưa thể bao quát được tất cả các công đoạn của một quy trình khép kín, nơi đòi hỏi phải có sự đầu tư lớn về nhân lực, tài chính, công nghệ và đủ mạnh để xây dựng được các chuỗi giá trị du lịch hấp dẫn, có uy tín.

Có nhiều người cho rằng, ở mảng khách du lịch quốc tế thì các công ty du lịch Việt Nam lâu nay làm việc giống như các “đại lý” của các công ty du lịch lớn ở nước ngoài, chứ không phải là bên trực tiếp bán tour cho khách. Có thể ý kiến này chưa chính xác lắm, nhưng thực tế cho thấy rõ ràng vai trò của các công ty du lịch Việt Nam ở mảng này còn có nhiều khiêm

khuyết. Bởi vì, thông thường một quy trình trọn vẹn của công việc kinh doanh du lịch bao gồm các phân khúc và nhiệm vụ cơ bản là: Marketing → bán tour → thiết kế tour → vận hành dịch vụ, thì lâu nay nhiều công ty du lịch Việt Nam hầu như chỉ thực hiện phân khúc cuối: vận hành dịch vụ, làm đối tác thực hiện (DMC – Destination Management Company - đơn vị tổ chức tại điểm đến), cho các công ty nước ngoài. Mà trong quy trình này thì phân khúc bán tour (distribution) mới là khâu thu lợi nhuận nhiều nhất, chứ không phải vận hành. Bên cạnh đó, cách làm này đương nhiên làm giảm tính chủ động của các công ty du lịch Việt nam và kéo theo các hạn chế khác.

Quay lại với vấn đề về mối quan hệ giữa DLVH và NTBD, qua trải nghiệm thực tế của tác giả bài viết, có thể thấy với cách làm hiện nay thì nhờ các show diễn nghệ thuật tại các điểm du lịch đã làm phong phú hơn sản phẩm của tour, nhưng cũng chỉ là ở mức hình thức, gây không khí, chứ chưa có chiều sâu văn hóa. Nhất là với loại hình du lịch “homestay”, các “show diễn” mang tính chất sinh hoạt cộng đồng, chất lượng nghệ thuật không cao. Ở nhiều nơi, người dân chưa quen làm du lịch, tụt tụt, rời rạc.

Cho đến nay, chất lượng hướng dẫn viên du lịch của Việt Nam chưa cao, còn nhiều hạn chế như về khả năng ngoại ngữ, sự hiểu biết về văn hóa, nghệ thuật dân tộc chưa sâu, nhất là thiếu sự am hiểu về nghệ thuật dân gian các vùng miền để có thể giới thiệu với du khách quốc tế và nội địa...

Nhìn chung, từ khi đất nước đổi mới, Nhà nước khuyến khích mở cửa đón khách quốc tế đến với Việt Nam thì ngành du lịch đã có nhiều hoạt động phong phú hơn, tuy nhiên sự hợp tác giữa hoạt động DLVH và NTBD thì chưa thực sự có sự gắn kết gì đáng kể, thiếu sự chủ động của cả hai bên, nên chưa phát huy được tiềm năng dồi dào của các địa phương.

#### 4. Hợp tác Du lịch văn hóa và Nghệ thuật biểu diễn: cơ hội và thách thức

Bước vào kế hoạch 5 năm phía trước (2025 – 2030) của các ngành CNVH Việt Nam, ngành Du lịch cũng như NTBD đang đứng trước nhiều cơ hội thuận lợi, cũng như cần vượt qua một số thách thức để đạt được các mục tiêu mà Chiến lược phát triển công nghiệp văn hóa đến năm 2030, tầm nhìn 2045 (theo Quyết định số 2486/QĐ-TTg đã được Thủ tướng Chính phủ phê duyệt ngày 14/11/2025).

##### 4.1. Cơ hội

Đều nằm trong top 10 ngành CNVH mũi nhọn của kế hoạch 5 năm tới, Du lịch Việt Nam và NTBD sẽ được Nhà nước chú trọng, ưu tiên hơn về chính sách và nguồn tài chính để thúc đẩy phát triển, góp phần tăng năng lực cạnh tranh của nền kinh tế cả nước và quảng bá hình ảnh Việt Nam với thế giới.

Những năm gần đây chính sách mở cửa mời gọi khách quốc tế đến tham quan, du lịch Việt Nam càng trở nên thông thoáng hơn, bằng các quy định về visa, nhập cảnh và lưu trú thuận tiện hơn trước rất nhiều. Xác định được tầm quan trọng của lĩnh vực du lịch trong việc góp phần tăng trưởng kinh tế, hầu như tỉnh thành nào cũng chú trọng đến phát triển du lịch, vận động và hướng dẫn người dân tham gia làm du lịch...phát huy thế mạnh của từng địa phương.

So với ngành Du lịch thì ngành NTBD có phần nhập cuộc khó khăn hơn, chậm trễ hơn. Nhưng năm 2025 (năm kỷ niệm nhiều sự kiện lịch sử lớn của đất nước) đã đánh dấu sự “bật dậy” của NTBD, nhất là hoạt động âm nhạc. Chưa bao giờ ở Việt Nam lại có thể tổ chức nhiều đại nhạc hội (Music Festival, Concert) hoành tráng đến thế, thu hút hàng chục ngàn khán giả mỗi đêm, với sân khấu ngoài trời rộng lớn và trang thiết bị sân khấu vô cùng hiện đại... Các sự kiện âm nhạc này thường được tổ chức ở các quảng trường, khu du lịch lớn ở Hà Nội, Hạ Long, Nha Trang, Thành phố

Hồ Chí Minh... thu hút rất nhiều khách du lịch. Đây là cơ sở để người Việt Nam, nhất là giới trẻ thấp lên niềm hy vọng cho nghệ thuật nước nhà, mở ra hướng đi mới cho NTBD Việt Nam: kết hợp chất liệu nghệ thuật truyền thống Việt Nam với tinh hoa nghệ thuật hiện đại của thế giới; tăng cường hợp tác quốc tế thông qua các dự án nghệ thuật chung của các nghệ sĩ Việt Nam và nghệ sĩ nước ngoài; ứng dụng kỹ thuật - công nghệ số tân tiến nhất hiện nay, để xây dựng nhiều chương trình nghệ thuật chất lượng cao, thu hút khách du lịch trong và ngoài nước đến với sự kiện... Nếu biết tổ chức tốt, làm việc nghiêm túc thì NTBD Việt Nam đang đứng trước nhiều cơ hội để “cắt cánh”.

Chúng ta đang sống trong thời đại khoa học kỹ thuật – công nghệ số phát triển như vũ bão mang tính toàn cầu. Là một trong những đất nước đang phát triển, Việt Nam được hưởng lợi rất nhiều từ cuộc Cách mạng Công nghiệp 4.0 đang hiện diện trong mọi lĩnh vực đời sống trên đất nước ta hôm nay. Trong đó, ngành Du lịch và ngành NTBD Việt Nam được tiếp thu rất nhiều lợi ích của khoa học công nghệ số.

##### 4.2. Thách thức

Thách thức lớn nhất của ngành Du lịch Việt Nam hiện nay và trong những năm tới vẫn là nâng cao vai trò của các công ty du lịch trong toàn bộ quy trình hoạt động. Với cách thức tổ chức, quản lý và vận hành của hoạt động du lịch như hiện nay, các công ty du lịch đang chủ yếu đảm trách phần ngọn (DMC), nhưng để có thể đảm trách cả phần khúc *distribution*, tức là làm cả vai trò người bán khách, thì đây là thách thức rất lớn của các công ty du lịch Việt Nam, vì việc này đòi hỏi phải có nguồn tài chính rất lớn và nguồn nhân lực chuyên nghiệp chất lượng cao, để có thể mở các văn phòng ở nước ngoài, xây dựng chiến lược marketing dài hạn, trực tiếp thu hút khách, tự bán tour và vận hành dịch vụ.

Đối với ngành NTBD, để có thể tạo nên

những bước đi đột phá thì quãng thời gian 5 năm trước mắt (2026 – 2030) là quá ngắn ngủi, gấp gáp. Những năm gần đây lĩnh vực âm nhạc đã phát huy được thế mạnh của chuyên ngành (không bị rào cản ngôn ngữ) nên đã làm được nhiều chương trình nghệ thuật có chất lượng mang tính quốc tế và tính hiện đại cao. Trong khi đó lĩnh vực sân khấu bị hạn chế hơn do đặc thù loại hình nghệ thuật, khiến cho sự hòa nhập với cuộc sống hiện đại có phần khó khăn hơn, sự hội nhập quốc tế cũng chậm hơn. Vì thế đây là bộ phận đang đứng trước nhiều thách thức nhất của ngành.

Khó khăn nhất là các đơn vị nghệ thuật sân khấu chuyên nghiệp, và khó khăn hơn cả là các nhà hát sân khấu truyền thống (trừ nghệ thuật Múa Rối nước). Sau nhiều cố gắng tìm cách kết nối để hợp tác với các công ty du lịch, đến nay việc này vẫn bế tắc.

Nhìn một cách tổng thể, nhiều chuyên gia cho rằng Việt Nam có tài nguyên du lịch rất mạnh, nhưng sản phẩm du lịch chưa khai thác hết.

### 5. Một số gợi ý về giải pháp

Để có thể bứt phá mạnh mẽ trong 5 năm trước mắt, theo chúng tôi ngành NTBD Việt Nam cần hợp tác chặt chẽ với ngành Du lịch, nhất là DLVH, và ngược lại. Dưới góc độ kinh tế cũng như góc độ tinh thần (lịch sử, văn hóa), đây là hai lĩnh vực, hai ngành rất cần cho nhau, cần phải đi cùng nhau, bởi vì trong du lịch có văn hóa, nói chính xác hơn, VHNT phục vụ du lịch, và bản thân tác phẩm VHNT là sản phẩm du lịch), và ngược lại, DLVH là nguồn cung dồi dào, đa dạng khách hàng cho NTBD).

#### 5.1. Đối với ngành Du lịch

Ai cũng biết, vấn đề quan trọng nhất của mọi doanh nghiệp là thương hiệu. ***Du lịch Việt Nam cần phải có thương hiệu chất lượng, uy tín và bền vững.***

Xây dựng thương hiệu của du lịch Việt Nam, nâng cao vị thế của du lịch Việt Nam trên trường quốc tế là nhiệm vụ, là mục tiêu của ngành Du lịch thuộc Chiến lược phát

triển Công nghiệp Văn hóa từ nay đến năm 2030, tầm nhìn 2045. Xu hướng phát triển tất yếu của Du lịch Việt Nam là cần vươn ra thế giới, cố gắng phấn đấu đạt được các văn phòng ở nước ngoài để chủ động trong hoạt động, tăng hiệu quả và chất lượng dịch vụ, phấn đấu ngày càng có nhiều sản phẩm du lịch mới để thu hút du khách đến Việt Nam. Đây quả là việc lớn, là thử thách lớn của ngành Du lịch non trẻ của Việt Nam, nhưng không phải không thể làm được. Chúng ta có thể học hỏi kinh nghiệm của nhiều nước, nhất là các nước ở châu Á (như Nhật Bản, Hàn Quốc, Trung Quốc), và các nước cùng khu vực Đông Nam Á, điển hình là Thailand, một nước láng giềng của chúng ta, có thâm niên về kinh doanh du lịch lâu hơn Việt Nam rất nhiều, và hoạt động rất hiệu quả.

Đối với thị trường trong nước, ngành Du lịch cũng cần có nhiều cải tiến, sáng tạo để thu hút người dân đi du lịch nhiều hơn. Do đời sống xã hội và mức sống kinh tế của người dân đang ngày càng được nâng cao nên nhu cầu du lịch càng lớn, đây là cơ hội của ngành Du lịch. Các công ty du lịch cần thay đổi “thói quen” nhiều năm nay là chỉ muốn phục vụ các sản phẩm có sẵn, ngại xây dựng sản phẩm mới cho các tour, thay vào đó cần phải có sáng tạo, tìm tòi để xây dựng chuỗi giá trị mới mẻ, phù hợp với sở thích du lịch của người dân hôm nay.

Với sự phát triển quá nhanh của thị trường và nhu cầu du lịch của xã hội ngày càng lớn, ngành Du lịch cần tăng cường công tác đào tạo nguồn nhân lực du lịch có chất lượng, thông thạo về nghiệp vụ và am hiểu văn hóa dân tộc, có kiến thức về công nghệ hiện đại, để những năm tới mới có thể đáp ứng được nhu cầu của các công ty du lịch.

#### 5.2. Đối với ngành Nghệ thuật Biểu diễn

Các đơn vị NTBD, các nhà hát sân khấu chuyên nghiệp cũng cần có hướng đi đa dạng để thu hút khán giả. Các nhà hát sân

khẩu truyền thống cần tiếp tục tìm phương cách để hợp tác với ngành Du lịch, trên quan điểm kinh tế là “hai bên cùng có lợi”, cùng phát triển. Các nhà hát chuyên nghiệp không chỉ là địa chỉ để các du khách đến xem vở diễn sân khấu, mà còn có thể là nơi đón các chuyên gia nước ngoài đến học tập, nghiên cứu chuyên sâu về sân khấu Việt Nam... Mặt khác, để tạo cho mình nguồn khán giả lâu dài, bền vững, các nhà hát không thể coi nhẹ thị trường trong nước, đặc biệt để thu hút giới trẻ thì cần chủ động hợp tác với các trường (THPT, đại học...) để tạo nên các lớp khán giả có độ tuổi khác nhau, họ sẽ là khán giả trung thành của sân khấu trong tương lai.

### 5.3. Tạo sức mạnh tổng hợp

Để sự hợp tác đi vào chiều sâu bền vững thì hai ngành DLVH và NTBD cần tìm tòi, sáng tạo, cùng nhau xây dựng các dự án thực sự gắn kết hoạt động của hai ngành, trong đó đặc biệt cần thiết là xây dựng kế hoạch marketing chung, cùng nhau mở rộng khai thác thị trường tiềm năng của du lịch cũng như của nghệ thuật, cùng có chính sách chăm sóc khách hàng, coi khách hàng của du lịch cũng là khán giả của nghệ thuật, và ngược lại.

### 7. Kết luận

Dưới góc độ là các ngành CNVH mũi nhọn của *Chiến lược phát triển công nghiệp văn hóa*, DLVH và NTBD là hai ngành độc lập, có nhiều sự khác biệt về chuyên môn, về cơ chế quản lý và vận hành, về nguồn nhân lực lao động... Nhưng trong hoạt động nghề nghiệp thì giữa DLVH và NTBD cũng có nhiều điểm tương đồng, có sự “ràng buộc” lẫn nhau, tương hỗ lẫn nhau. Đây là cơ sở vững chắc, có tính khả thi cho sự hợp tác của DLVH và NTBD trong thời kỳ mới, có thể “nắm tay nhau” đi tới những mục tiêu kinh tế - văn hóa – nhân văn, với phương châm kinh doanh “hai bên cùng có lợi”, cùng phát triển.

Hơn bao giờ hết, DLVH và NTBD cần

phải chủ động tìm hướng đi sáng tạo, mới mẻ, chủ động đến với nhau và xây dựng các dự án chung để phát triển các thị trường khách hàng tiềm năng trong nước và quốc tế. Có như vậy mới đủ sức để cùng các ngành CNVH khác tiến tới “mục tiêu đưa đóng góp của các ngành công nghiệp văn hóa đạt 7% GDP vào năm 2030, đạt 9% GDP vào năm 2045 như *Chiến lược phát triển công nghiệp văn hóa* đã đề ra.

### TÀI LIỆU THAM KHẢO

- [1]. Đảng Cộng sản Việt Nam (1943), *Đề cương về văn hóa Việt Nam*, trong Văn kiện Đảng toàn tập, tập 7, Nxb Chính trị quốc gia Sự thật, 2000, Hà Nội, tr. 316 - 319.
- [2] Đảng Cộng sản Việt Nam (1998), *Nghị quyết Hội nghị lần thứ năm Ban Chấp hành Trung ương khóa VIII*, trong Văn kiện Đảng toàn tập, tập 57, Nxb Chính trị quốc gia, 2005, Hà Nội, tr.500 – 520.
- [3]. *Giáo trình Marketing du lịch* (2023), TS. Nguyễn Đức Thắng (chủ biên) – TS. Nguyễn Thành Nam – Ths. Lê Đức Hiếu – Ths. Đồng Thùy Dương – Ths. Hoàng Quang Huy, Nxb. Lao động, Hà Nội.
- [4]. Lê Thị Hoài Phương (2016), *Quản lý hoạt động Nghệ thuật Biểu diễn trong cơ chế thị trường*, Nxb. Sân khấu, Hà Nội.
- [5]. Lê Thị Hoài Phương (2023), *Nghệ thuật Biểu diễn Việt Nam thời kỳ Đổi mới: Chính sách và thực tiễn*, Nxb Sân khấu, Hà Nội.
- [6]. Quyết định số 1755/QĐ-TTg, *phê duyệt Chiến lược phát triển các ngành công nghiệp văn hóa Việt Nam đến năm 2020, tầm nhìn đến năm 2030*, ngày 08/09/2016.
- [7]. Quyết định số 2486/QĐ-TTg, *Phê duyệt Chiến lược phát triển công nghiệp văn hóa đến năm 2030, tầm nhìn 2045*, ngày 14/11/2025
- [8]. UNESCO, *Các ngành Công nghiệp văn hóa - Tâm điểm của văn hóa trong tương lai*, [portal.unesco.org](http://portal.unesco.org).

# PHÁT TRIỂN DU LỊCH VĂN HÓA TRONG HỆ SINH THÁI CÔNG NGHIỆP VĂN HÓA VIỆT NAM

Nguyễn Thị Thu Trang<sup>1</sup>  
Lê Mạnh Hùng<sup>2</sup>

<sup>1</sup>Viện Văn hóa, Nghệ thuật, Thể thao và Du lịch Việt Nam,  
Bộ Văn hóa, Thể thao và Du lịch.  
<sup>2</sup>Khoa Kinh tế, Đại học Thủy lợi.  
Email: trangvicas79@gmail.com

## TÓM TẮT

Bài viết phân tích phát triển du lịch văn hóa Việt Nam dưới góc nhìn hệ sinh thái công nghiệp văn hóa, kết hợp khung lý thuyết chuỗi giá trị và tiếp cận hệ sinh thái để nhận diện các hạn chế trong liên kết ngành, năng lực sáng tạo, hạ tầng số và quản trị di sản. Chuyển đổi số được đánh giá vừa là cơ hội nâng cao trải nghiệm và giá trị di sản văn hóa, vừa đặt ra thách thức về bản quyền, tiêu chuẩn dữ liệu và chất lượng nhân lực. Trên cơ sở đó, bài viết đề xuất nhóm giải pháp theo mô hình “kiềng ba chân”: hoàn thiện thể chế theo hướng kiến tạo, ứng dụng công nghệ trong kể chuyện di sản và phát triển nguồn nhân lực liên ngành. Những gợi mở này nhằm góp phần thúc đẩy du lịch văn hóa trở thành câu chuyện quan trọng trong hệ sinh thái công nghiệp văn hóa Việt Nam.

**Từ khóa:** Chuyển đổi số; Chuỗi giá trị; Du lịch văn hóa; Hệ sinh thái công nghiệp văn hóa.

## ABSTRACT

This article examines the development of cultural tourism in Vietnam through the lens of the cultural industries ecosystem. By integrating the theoretical frameworks of the value chain and the ecosystem approach, the study identifies key constraints related to cross-sectoral linkages, creative capacity, digital infrastructure, and heritage management. Digital transformation is assessed as both an opportunity to enhance cultural experiences and a tool for enriching heritage interpretation, while also posing challenges concerning copyright, data standards, and workforce readiness. Based on these insights, the article proposes a set of solutions aligned with the “three-legged stool” model, emphasizing institutional reform toward an enabling state, the application of digital storytelling technologies, and the development of interdisciplinary human resources. These recommendations aim to position cultural tourism as a dynamic and integral component of Vietnam’s cultural industries ecosystem.

**Keywords:** Cultural tourism; Cultural industries ecosystem; Digital transformation; Value chain.

Ngày nhận bài: 01/03/2026; Ngày sửa bài: 16/03/2026; Ngày duyệt đăng: 02/04/2026

## 1. Đặt vấn đề

Trong bối cảnh nền kinh tế toàn cầu đang chuyển dịch mạnh mẽ sang mô hình kinh tế tri thức và sáng tạo, công nghiệp văn hóa (Cultural Industries) đã được UNESCO và nhiều quốc gia xác định là động lực tăng trưởng bền vững mới. Theo Báo cáo Kinh tế Sáng tạo của UNESCO (2013)<sup>1</sup> và UNCTAD (2022)<sup>2</sup>, sự giao thoa giữa văn hóa, thương mại và công nghệ không chỉ tạo ra giá trị kinh tế trực tiếp mà còn gia tăng sức mạnh mềm quốc gia. Trong hệ sinh thái đa dạng đó, du lịch văn hóa (Cultural tourism) đóng vai trò là “kênh phân phối” và “không gian tiêu dùng” quan trọng bậc nhất, nơi các sản phẩm của công nghiệp sáng tạo (như di sản, nghệ thuật biểu diễn, thủ công mỹ nghệ, thiết kế) được thương mại hóa và gia tăng giá trị thông qua trải nghiệm của du khách.

Đối với Việt Nam, sự chuyển dịch nhận thức về vai trò của văn hóa trong phát triển đã được thể chế hóa rõ ràng trong các văn bản chiến lược của Nhà nước. Chiến lược phát triển các ngành công nghiệp văn hóa Việt Nam đến năm 2030, tầm nhìn đến năm 2045 (Quyết định số 2486/QĐ-TTg ngày 14/11/2025) đặt mục tiêu các ngành công nghiệp văn hóa đạt tốc độ tăng trưởng bình quân khoảng 10%/năm và đóng góp 7% GDP của đất nước. Lực lượng lao động trong các ngành công nghiệp văn hóa bình quân tăng 10%/năm, chiếm tỷ trọng 6% trong tổng lực lượng lao động của nền kinh tế. Trong đó, ngành du lịch văn hóa được xác

định là một trong 6 lĩnh vực trọng điểm<sup>3</sup>. Song song với đó là *Chiến lược phát triển du lịch Việt Nam đến năm 2030* (Quyết định số 147/QĐ-TTg, 2020) định hướng phát triển du lịch theo hướng bền vững, chuyên nghiệp, hiện đại và mang đậm bản sắc văn hóa dân tộc. Sự tương thích về tầm nhìn giữa hai chiến lược này tạo ra nền tảng quan trọng để hình thành hệ sinh thái công nghiệp văn hóa – du lịch, nơi các ngành sáng tạo cùng vận hành và hỗ trợ cho nhau.

Việt Nam có lợi thế nổi trội về tài nguyên văn hóa sở hữu kho tàng 9 di sản thế giới<sup>4</sup>, hàng nghìn lễ hội truyền thống, ẩm thực đa dạng vùng miền, hàng nghìn làng nghề thủ công, cùng hệ thống nghệ thuật dân gian phong phú... Ngoài ra, Việt Nam có tới 16 di sản văn hoá phi vật thể được UNESCO ghi danh là di sản đại diện của nhân loại. Những “nguồn nguyên liệu văn hóa” này là cơ sở để phát triển chuỗi giá trị du lịch văn hóa – sáng tạo: từ bảo tồn di sản, sáng tạo sản phẩm, thiết kế trải nghiệm, đến quảng bá và thương mại hóa. Bên cạnh đó, tiến trình chuyển đổi số và hội nhập toàn cầu đang mở ra cơ hội chưa từng có để số hóa di sản, ứng dụng công nghệ thực tế ảo, trí tuệ nhân tạo, dữ liệu lớn... nhằm tạo ra những trải nghiệm du lịch cá nhân hóa, thông minh và hấp dẫn hơn.

3. <https://vanban.chinhphu.vn/?pageid=27160&docid=215888>

4. Theo Cục Di sản Văn hoá, tính đến tháng 7/2025, Việt Nam đã sở hữu: 6 di sản văn hoá thế giới (Quần thể di tích Cố đô Huế, Khu di tích Chăm – Mỹ Sơn, Đô thị cổ Hội An, Hoàng thành Thăng Long, Thành nhà Hồ, Quần thể di tích và danh thắng Yên Tử – Vĩnh Nghiêm – Côn Sơn, Kiếp Bạc), 2 di sản thiên nhiên thế giới (Vịnh Hạ Long – Quần đảo Cát Bà và Phong Nha – Kẻ Bàng) và 1 di sản hỗn hợp là Quần thể danh thắng Tràng An. [<https://nhandan.vn/inforgraphic-9-di-san-the-gioi-o-viet-nam-duoc-unesco-vinh-danh-post893439.html>].

1. UNESCO. (2013). *Creative Economy Report 2013 Special Edition: Widening Local Development Pathways*. Paris: UNESCO/UNDP.

2. UNCTAD. (2022). *Creative Economy Outlook 2022*. Geneva: United Nations Conference on Trade and Development.

Tuy nhiên, quá trình phát triển du lịch văn hóa ở Việt Nam vẫn đang đối diện nhiều thách thức: liên kết ngành chưa chặt chẽ; năng lực sáng tạo và nhân lực quản trị còn hạn chế; sự thương mại hóa thiếu kiểm soát ở một số điểm đến làm suy giảm giá trị nguyên bản của di sản; trong khi công nghệ và chính sách bản quyền chưa đáp ứng yêu cầu của thời đại số. Những vấn đề này đặt ra yêu cầu cấp thiết phải tiếp cận, phát triển du lịch văn hóa trong khuôn khổ hệ sinh thái công nghiệp văn hóa, thay vì chỉ xem đây là một hoạt động dịch vụ đơn lẻ.

Từ góc nhìn đó, bài viết sẽ tập trung phân tích về: (1) thực trạng vận hành chuỗi giá trị du lịch văn hóa ở Việt Nam; (2) Đánh giá tác động của chuyển đổi số đối với việc gia tăng trải nghiệm văn hóa; (3) Đề xuất các giải pháp nhằm thúc đẩy sự phát triển du lịch văn hóa trong hệ sinh thái công nghiệp văn hóa, hướng tới mục tiêu phát triển bền vững.

### 2. Phương pháp nghiên cứu

Phương pháp nghiên cứu của bài viết dựa trên cách tiếp cận liên ngành, kết hợp phân tích lý thuyết và tổng hợp dữ liệu thứ cấp nhằm làm rõ mối quan hệ giữa du lịch văn hóa và hệ sinh thái công nghiệp văn hóa Việt Nam. Bài viết xây dựng khung phân tích dựa trên hai nền tảng lý thuyết: chuỗi giá trị (Value chain)<sup>5</sup> và tiếp cận hệ sinh thái (Ecosystem approach)<sup>6</sup>. Khung chuỗi giá trị được sử dụng để phân tích cấu trúc hoạt động của du lịch văn hóa qua các khâu sáng tạo, sản xuất, phân phối, truyền thông, trải nghiệm và

tái đầu tư giá trị; qua đó nhận diện những điểm nghẽn trong liên kết ngành và mức độ tạo giá trị gia tăng. Đồng thời, tiếp cận hệ sinh thái cho phép xem du lịch văn hóa như một thành tố trong mạng lưới các tác nhân (gồm: cơ quan quản lý, cộng đồng di sản, doanh nghiệp du lịch, doanh nghiệp sáng tạo, nền tảng số và du khách...) với các tương tác phụ thuộc lẫn nhau được chi phối bởi điều kiện công nghệ, thiết chế văn hóa và chính sách. Việc kết hợp hai khung lý thuyết giúp bài viết đánh giá đồng thời cả logic vận hành nội tại của chuỗi giá trị và động lực cộng hưởng của toàn hệ sinh thái, từ đó phân tích tác động của chuyển đổi số và đề xuất giải pháp chính sách hướng tới phát triển bền vững du lịch văn hóa trong bối cảnh công nghiệp văn hóa đang mở rộng tại Việt Nam.

### 3. Thực trạng vận hành chuỗi giá trị du lịch văn hóa ở Việt Nam

Trong những năm gần đây, du lịch văn hóa ở Việt Nam ngày càng được nhìn nhận không chỉ như một hoạt động dịch vụ du lịch đơn thuần mà còn như một cấu phần quan trọng của hệ sinh thái công nghiệp văn hóa. Đặc biệt trong tinh thần của Nghị quyết 80 của Bộ Chính trị, các ngành công nghiệp văn hóa được xác định là một trong những trụ cột quan trọng của chiến lược phát triển văn hóa. Nghị quyết nhấn mạnh yêu cầu “thúc đẩy phát triển các ngành công nghiệp văn hóa, thị trường văn hóa, định vị sản phẩm văn hoá thương hiệu quốc gia gắn với phát triển du lịch văn hóa”<sup>7</sup>. Trong bối cảnh

5. Porter, Michael E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: The Free Press.

6. Moore, James F. (1996). *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. New York: HarperBusiness.

7. Báo điện tử Chính phủ (2026); *Toàn văn Nghị quyết số 80-NQ/TW của Bộ Chính trị về phát triển văn hóa Việt Nam*; <https://xaydungchinh sach.chinhphu.vn/nghi-quyet-so-80-nq-tw-ve-phat-trien-van-hoa-viet-nam-119260113133313565.htm>; cập nhật ngày 13/03/2026.

đó, du lịch văn hóa ngày càng được xem là một kênh quan trọng nhằm chuyển hóa các giá trị văn hóa thành nguồn lực kinh tế và giá trị xã hội.

Trên thực tế, quá trình phát triển du lịch văn hóa ở Việt Nam vẫn đang trong giai đoạn chuyển đổi từ mô hình khai thác tài nguyên sang mô hình phát triển theo logic chuỗi giá trị và hệ sinh thái sáng tạo. Điều này thể hiện ở nhiều khía cạnh của hệ sinh thái công nghiệp văn hóa.

Trước hết, ở khâu sáng tạo và phát triển nội dung văn hóa, nguồn tài nguyên văn hóa của Việt Nam rất phong phú, bao gồm hệ thống di sản vật thể, phi vật thể, lễ hội truyền thống, nghệ thuật biểu diễn, làng nghề và tri thức bản địa. Tuy nhiên, việc chuyển hóa những chất liệu văn hóa này thành các sản phẩm trải nghiệm mang tính sáng tạo cao vẫn còn hạn chế. Trong nhiều trường hợp, sản phẩm du lịch văn hóa vẫn chủ yếu dựa vào mô hình tham quan – thuyết minh truyền thống, chưa hình thành được các format trải nghiệm mới tích hợp nghệ thuật, thiết kế sáng tạo, công nghệ và kể chuyện văn hóa (cultural storytelling). Điều này cho thấy mối liên kết giữa du lịch văn hóa với các ngành công nghiệp sáng tạo khác như thiết kế, nghệ thuật biểu diễn, điện ảnh, truyền thông số... trong hệ sinh thái CNVH vẫn chưa thực sự chặt chẽ.

Ở khâu sản xuất và tổ chức trải nghiệm, năng lực vận hành chuỗi giá trị du lịch văn hóa vẫn còn phân tán. Nhiều điểm đến thiếu cơ chế phối hợp hiệu quả giữa các chủ thể như cơ quan quản lý di sản, doanh nghiệp du lịch, nghệ nhân, cộng đồng địa phương và các tổ chức sáng tạo. Điều này khiến cho trải nghiệm của du khách chưa đạt tới mức độ đồng bộ và liên tục trong toàn bộ hành trình trải nghiệm văn hóa.

Ngoài ra, năng lực quản trị điểm đến, chất lượng nguồn nhân lực hướng dẫn và khả năng kết nối giữa các dịch vụ hỗ trợ—từ lưu trú, vận chuyển đến dịch vụ trải nghiệm văn hóa—vẫn còn chênh lệch giữa các địa phương.

Ở khâu phân phối và thị trường, du lịch văn hóa Việt Nam vẫn phụ thuộc khá lớn vào các nền tảng trung gian quốc tế và các kênh phân phối truyền thống của ngành du lịch. Việc xây dựng thương hiệu văn hóa của điểm đến còn rời rạc, thiếu chiến lược dài hạn và chưa gắn kết chặt chẽ với hệ sinh thái sáng tạo của CNVH. Do đó, giá trị kinh tế thu được từ du lịch văn hóa chưa phản ánh đầy đủ giá trị biểu tượng và giá trị sáng tạo của các sản phẩm văn hóa.

Ở khâu tiêu dùng văn hóa, trải nghiệm của du khách tại nhiều điểm đến vẫn chủ yếu dừng ở mức quan sát và thưởng thức thụ động, trong khi xu hướng của thị trường du lịch hiện nay ngày càng hướng tới các hình thức trải nghiệm mang tính tương tác, tham gia và đồng sáng tạo. Khoảng cách này cho thấy hệ sinh thái du lịch văn hóa ở Việt Nam vẫn chưa tận dụng hết tiềm năng của các nguồn lực sáng tạo trong cộng đồng, đặc biệt là nghệ nhân, nhà thiết kế, nghệ sĩ và các doanh nghiệp sáng tạo.

Bên cạnh đó, các rào cản thể chế và hạ tầng cũng ảnh hưởng đáng kể đến sự vận hành của hệ sinh thái. Chính sách quản lý văn hóa, du lịch và công nghiệp sáng tạo hiện vẫn còn phân tán theo ngành; cơ chế bảo hộ bản quyền đối với các sản phẩm văn hóa số và sản phẩm sáng tạo còn chưa hoàn thiện; trong khi hạ tầng công nghệ thông tin và nền tảng dữ liệu văn hóa chưa đồng đều giữa các địa phương.

Trong hơn một thập niên qua, Việt

Nam cũng đã chứng kiến nhiều thử nghiệm tích cực trong việc kết nối du lịch văn hóa với các ngành công nghiệp sáng tạo. Một số trung tâm văn hóa – du lịch như Hà Nội, Huế, Hội An, Ninh Bình, Đà Nẵng và TP. Hồ Chí Minh đã bước đầu hình thành các mô hình kết hợp giữa văn hóa, sáng tạo và công nghệ thông qua các sự kiện văn hóa lớn, các chương trình biểu diễn nghệ thuật quy mô và các sản phẩm trải nghiệm di sản như: Festival Huế, Ký ức Hội An, Tinh hoa Bắc Bộ hay các tour trải nghiệm làng nghề. Những mô hình này cho thấy tiềm năng của du lịch văn hóa khi được đặt trong một hệ sinh thái công nghiệp văn hóa rộng hơn, nơi giá trị văn hóa có thể được tái tạo và lan tỏa thông qua các hình thức sáng tạo đa dạng. Nhưng so với tiềm năng của nguồn tài nguyên văn hóa cũng như bối cảnh phát triển mạnh mẽ của công nghiệp văn hóa trong khu vực, đóng góp của du lịch văn hóa Việt Nam trong hệ sinh thái CNVH vẫn còn khiêm tốn, cả về giá trị kinh tế lẫn mức độ lan tỏa sáng tạo.

#### **4. Tác động của chuyển đổi số đối với việc gia tăng trải nghiệm văn hóa**

Tác động của chuyển đổi số đối với việc gia tăng trải nghiệm văn hóa trong du lịch văn hóa Việt Nam ngày càng thể hiện rõ nét, song mang tính hai mặt tùy thuộc vào mức độ sẵn sàng của hệ thống và năng lực triển khai ở từng địa phương. Về phương diện tích cực, chuyển đổi số mở ra khả năng mở rộng không gian trải nghiệm thông qua các công nghệ như số hóa di sản, thực tế ảo (VR), thực tế tăng cường (AR), bản đồ tương tác hay không gian triển lãm trực tuyến, cho phép du khách tiếp cận các giá trị văn hóa một cách sinh động, trực quan và vượt qua giới hạn địa lý – thời gian. Công nghệ số

cũng góp phần tăng chiều sâu trải nghiệm bằng việc cung cấp thông tin đa phương tiện, cá nhân hóa nội dung theo thị hiếu du khách, đồng thời kết nối trải nghiệm tại chỗ với trải nghiệm trực tuyến trước và sau chuyến đi. Việc ứng dụng phân tích dữ liệu lớn (big data) và AI giúp điếm đến hiểu rõ hành vi, nhu cầu của du khách, từ đó thiết kế sản phẩm văn hóa phù hợp hơn, nâng cao mức độ hài lòng và gắn kết cảm xúc.

Tuy nhiên, chuyển đổi số cũng đặt ra những thách thức đáng kể. Sự thiếu đồng bộ về hạ tầng công nghệ và mức độ sẵn sàng chuyển đổi số giữa các địa phương khiến trải nghiệm số không nhất quán, thậm chí gây đứt gãy trong hành trình trải nghiệm. Việc số hóa mà không có chiến lược kiểm định chất lượng và khung bản quyền rõ ràng dễ dẫn tới tình trạng “kỹ thuật hóa” di sản theo hướng trình diễn hóa, làm giảm tính nguyên bản và chiều sâu văn hóa. Một số sản phẩm ứng dụng công nghệ cao còn mang tính phô diễn kỹ thuật hơn là phục vụ mục tiêu trải nghiệm, không tạo được giá trị cảm xúc hoặc ý nghĩa văn hóa cho du khách. Ngoài ra, năng lực số của cộng đồng sở hữu di sản và các chủ thể sáng tạo còn hạn chế khiến họ khó tham gia bình đẳng vào chuỗi cung ứng sản phẩm số, dẫn tới nguy cơ gia tăng chênh lệch giữa các khu vực và nhóm tác nhân.

Nhìn chung, chuyển đổi số có tiềm năng mạnh mẽ trong việc nâng cao chất lượng và tính hấp dẫn của trải nghiệm văn hóa, nhưng hiệu quả của nó phụ thuộc lớn vào các điều kiện nền tảng: khung pháp lý về bản quyền và dữ liệu; năng lực số của các tác nhân; tính liên thông của hạ tầng; và khả năng tích hợp công nghệ vào chiến lược phát triển sản phẩm

du lịch văn hóa. Chỉ khi những điều kiện này được đảm bảo, chuyển đổi số mới thực sự trở thành công cụ gia tăng giá trị trải nghiệm và củng cố sự bền vững của du lịch văn hóa trong hệ sinh thái công nghiệp văn hóa.

## **5. Nhóm giải pháp thúc đẩy phát triển du lịch văn hóa trong hệ sinh thái công nghiệp văn hóa**

### **5.1. Hoàn thiện thể chế và cơ chế điều phối hệ sinh thái**

Một trong những điều kiện tiên quyết để tái cấu trúc du lịch văn hóa theo hướng bền vững là chuyển đổi tư duy quản lý từ “kiểm soát” sang “kiến tạo”. Điều này đòi hỏi Nhà nước không chỉ đóng vai trò ban hành quy định mà còn trở thành tác nhân điều phối, hỗ trợ và kích hoạt các tương tác trong hệ sinh thái. Trước hết, cần xây dựng cơ chế hợp tác công – tư (PPP) chuyên biệt dành cho các dự án hạ tầng văn hóa sáng tạo, bao gồm trung tâm sáng tạo, không gian nghệ thuật hoặc khu trưng bày công nghệ trong môi trường di sản. Bên cạnh đó, việc hoàn thiện khung pháp lý về bản quyền số đối với các sản phẩm phái sinh từ di sản là yêu cầu cấp bách nhằm tạo hành lang pháp lý minh bạch cho doanh nghiệp sáng tạo và công nghệ, đồng thời khuyến khích tái đầu tư giá trị vào hoạt động bảo tồn. Sự thay đổi này giúp gia tăng liên kết giữa các tác nhân và khắc phục tình trạng vận hành rời rạc vốn tồn tại trong chuỗi giá trị du lịch văn hóa.

### **5.2. Đổi mới phương thức kể chuyện di sản thông qua công nghệ**

Ứng dụng công nghệ số vào khai thác di sản không nên chỉ dừng lại ở mức tái hiện thông tin mà cần hướng đến tạo lập trải nghiệm tương tác, giúp du khách tham gia sâu hơn vào câu chuyện văn

hóa. Vì vậy, cần hình thành cơ chế hợp tác giữa các nhà sử học, chuyên gia văn hóa, nhà thiết kế trải nghiệm và doanh nghiệp công nghệ để phát triển các mô hình “kể chuyện di sản” (digital storytelling) mang tính sáng tạo. Công nghệ thực tế tăng cường (AR), thực tế ảo (VR), mô phỏng 3D hay bảo tàng số không chỉ hỗ trợ tái hiện không gian lịch sử mà còn có thể mở ra các lớp thông tin mới, giúp “kích hoạt” di sản trong mắt du khách. Một ví dụ điển hình là việc triển khai ứng dụng AR tại Hoàng thành Thăng Long, cho phép du khách tương tác với bối cảnh và nghi lễ cung đình ngay trên nền di tích khảo cổ. Những đổi mới này góp phần khắc phục hạn chế về tính tĩnh của sản phẩm du lịch văn hóa truyền thống và gia tăng giá trị trải nghiệm trong kỷ nguyên số.

### **5.3. Phát triển nguồn nhân lực liên ngành cho du lịch văn hóa**

Nguồn nhân lực chất lượng cao là “trụ cột thứ ba” bảo đảm sự vận hành hữu hiệu của hệ sinh thái du lịch văn hóa. Trong bối cảnh du lịch gắn ngày càng chặt chẽ với công nghệ và công nghiệp sáng tạo, nhu cầu về lực lượng lao động “lai” (hybrid workforce) trở nên đặc biệt cấp thiết. Mặc dù số lượng lao động du lịch gia tăng, chất lượng đào tạo vẫn chưa đáp ứng yêu cầu hội nhập, thể hiện qua hạn chế về kỹ năng chuyên môn, năng lực sáng tạo và khả năng cạnh tranh quốc tế của đội ngũ hướng dẫn viên cũng như nhà quản lý doanh nghiệp<sup>8</sup>.

8. Việt Nam thiếu trầm trọng đội ngũ lao động có kỹ năng và được đào tạo bài bản trong lĩnh vực du lịch và dịch vụ. Nhân sự trong nhiều khách sạn, nhà hàng và khu du lịch, đặc biệt là ở những khu vực xa thành phố lớn vẫn thiếu chuyên nghiệp, chất lượng kém, chưa đáp ứng được nhu cầu và đòi hỏi cho khách du lịch hạng sang. Theo ông Phạm Hà, Chủ tịch Lux Group: Đội ngũ hướng dẫn viên chuyên nghiệp còn thiếu và chuyên môn yếu, đặc biệt là hướng dẫn viên nói được các ngoại ngữ hiếm như Đức, Ý, Tây Ban

Để khắc phục tình trạng này, người lao động cần đồng thời sở hữu kiến thức nền về văn hóa – di sản, kỹ năng quản trị dịch vụ, tư duy thiết kế sáng tạo và khả năng ứng dụng công nghệ số. Điều này đòi hỏi hệ thống giáo dục – đào tạo chuyển từ mô hình đơn ngành sang liên ngành, mở rộng các khóa bồi dưỡng ngắn hạn, workshop và chương trình hợp tác giữa trường đại học văn hóa – nghệ thuật, doanh nghiệp công nghệ và các đơn vị vận hành điểm đến. Việc hình thành đội ngũ nhân lực “lai” không chỉ nâng cao chất lượng lao động mà còn tạo nền tảng cho sự hội tụ bền vững giữa du lịch và công nghiệp văn hóa.

### 6. Kết luận

Phát triển du lịch văn hóa trong bối cảnh hệ sinh thái công nghiệp văn hóa Việt Nam đang chuyển dịch mạnh mẽ đòi hỏi một cách tiếp cận đa chiều, tích hợp và hướng tới tính bền vững. Qua việc phân tích chuỗi giá trị du lịch văn hóa, bài viết cho thấy những “điểm nghẽn” cốt lõi nằm ở sự liên kết ngành còn rời rạc, năng lực sáng tạo hạn chế, thiếu hụt hạ tầng số và khung pháp lý phù hợp, cũng như sự mất cân bằng giữa khai thác kinh tế và bảo tồn giá trị di sản. Đánh giá tác động của chuyển đổi số cũng cho thấy: công nghệ vừa mở ra cơ hội nâng cao trải nghiệm văn hóa, mở rộng không gian sáng tạo, vừa đặt ra những thách thức mới liên quan đến bản quyền, năng lực

---

Nha, Pháp,... phát triển du lịch phải song song với đào tạo nguồn nhân lực có chất lượng. Muốn phục vụ được khách 4 - 5 sao thì phải có nhân lực chất lượng cao, đáp ứng các tiêu chuẩn này. Công tác đào tạo cần được thực hiện một cách bài bản ngay từ trong trường, gắn với thực tiễn công việc để khi ra trường, sinh viên tốt nghiệp có thể đáp ứng được yêu cầu làm việc tại các doanh nghiệp, nhà hàng, khách sạn mà không phải đào tạo lại. [<https://theleader.vn/4-diem-yeu-cua-du-lich-viet-1611570042997.htm>].

quản trị và tính đồng bộ của hạ tầng.

Từ góc nhìn hệ sinh thái, các giải pháp được đề xuất theo cấu trúc “kiềng ba chân” gồm: thể chế kiến tạo, đổi mới công nghệ và phát triển nguồn nhân lực liên ngành. Góc nhìn này cho thấy, cần thiết lập chặt chẽ mối quan hệ giữa các tác nhân trong hệ sinh thái và thúc đẩy sự cộng hưởng giữa du lịch và công nghiệp văn hóa. Việc hoàn thiện khung pháp lý, tăng cường ứng dụng công nghệ trong kể chuyện di sản và xây dựng đội ngũ nguồn nhân lực không chỉ góp phần nâng cao giá trị trải nghiệm của du khách mà còn tạo nền tảng cho chuỗi giá trị du lịch văn hóa vận hành hiệu quả, sáng tạo và bền vững hơn.

Nhìn từ góc độ tổng thể, phát triển du lịch văn hóa không nên xem như một hoạt động đơn ngành mà cần được đặt trong hệ sinh thái công nghiệp văn hóa - nơi giá trị được tạo lập và gia tăng thông qua sự kết nối, tương tác và đổi mới liên tục. Trên cơ sở đó, kết quả phân tích trong bài viết gợi mở một hướng tiếp cận mới cho hoạch định chính sách, phát triển sản phẩm và quản trị điểm đến trong bối cảnh chuyển đổi số, đồng thời khẳng định tiềm năng của du lịch văn hóa như một cấu phần chiến lược trong tăng trưởng của kinh tế sáng tạo ở Việt Nam.

---

## TÀI LIỆU THAM KHẢO

### I. Tài liệu Tiếng Việt:

1. Lại Thị Thanh Bình (2020). *Công nghiệp văn hóa, công nghiệp sáng tạo và sự lựa chọn mô hình phù hợp ở Việt Nam*. Hà Nội, 22/11/2020.
2. Lê Thị Kim Chi (2024). *Phát triển du lịch sáng tạo tại Hà Nội*. *Tạp chí Khoa học và Công nghệ*, số 43.
3. Phan Thế Công (2015). *Phát triển*

các ngành công nghiệp sáng tạo ở Việt Nam trong bối cảnh hội nhập quốc tế. *Tạp chí Khoa học và Xã hội Việt Nam*, số 1 (86).

4. Trần Thị Ngọc Quyên (2025- Chủ biên); *Hành trình từ di sản đến phát triển công nghiệp văn hóa trong kinh tế sáng tạo*; Nxb Giáo dục Việt Nam; HN,

5. Nguyễn Thị Thu Trang (2022); *Phát triển ngành Du lịch văn hóa ở Việt Nam: cơ hội và thách thức* tại Hội thảo: Đánh giá 5 năm thực hiện chiến lược phát triển các ngành công nghiệp văn hóa ở Việt Nam do Viện Văn hóa Nghệ thuật quốc gia Việt Nam tổ chức 9/2022.

6. Nguyễn Thị Thu Trang, Nguyễn Thị Lan Hương (2025); *Phát triển công nghiệp văn hóa trong mối liên kết với du lịch: Động lực mới cho kinh tế sáng tạo ở Việt Nam* in trong cuốn “Công nghiệp văn hóa với phát triển du lịch bền vững tại Việt Nam trong bối cảnh hội nhập quốc tế”; Nxb KHXH. HN 2025. ISBN 9786326071764

7. Thủ tướng Chính phủ (2025). *Quyết định số 2486/QĐ-TTg ngày 14/11/2025 phê duyệt Chiến lược phát triển các ngành công nghiệp văn hóa Việt Nam đến năm 2030, tầm nhìn đến năm 2045*; [https://vanban.chinhphu.vn/?pageid=27160 & docid=215888](https://vanban.chinhphu.vn/?pageid=27160&docid=215888)

8. Báo điện tử Chính phủ (2026); *Toàn văn Nghị quyết số 80-NQ/TW của Bộ Chính trị về phát triển văn hóa Việt Nam*; <https://xaydungchinhhsach.chinhphu.vn/nghi-quyet-so-80-nq-tw-ve-phat-trien-van-hoa-viet-nam>

[chinhphu.vn/nghi-quyet-so-80-nq-tw-ve-phat-trien-van-hoa-viet-nam-119260113133313565.htm](http://chinhphu.vn/nghi-quyet-so-80-nq-tw-ve-phat-trien-van-hoa-viet-nam-119260113133313565.htm); cập nhật ngày 13/03/2026.

## II. Tài liệu tiếng Anh

9. British Council Vietnam. (2022). *Vietnam Creative Industries Mapping Report*. Hanoi: British Council.

10. Moore, J. F. (1996). *The death of competition: Leadership and strategy in the age of business ecosystems*. HarperBusiness.

11. Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. The Free Press.

12. Richards, G. (2011). Creativity and tourism: The state of the art. *Annals of Tourism Research*, 38(4), 1225-1253. <https://doi.org/10.1016/j.annals.2011.07.008>

13. Throsby, D. (2001). *Economics and Culture*. Cambridge: Cambridge University Press.

14. UNCTAD. (2022). *Creative Economy Outlook 2022*. Geneva: United Nations Conference on Trade and Development.

15. UNDP Vietnam. (2020). *Creative Hubs and Cultural Industries in Vietnam*. Hanoi: UNDP.

16. UNESCO. (2013). *Creative Economy Report 2013 Special Edition: Widening Local Development Pathways*. Paris: UNESCO/UNDP.

# PERSONAL FACTORS AND TEACHER STRESS: IMPLICATIONS FOR EFL LECTURERS IN HIGHER EDUCATION

**Dinh Thi Phuong Thao**

*Faculty of English, East Asia University of Technology*

*Email: thaodtp@eaut.edu.vn*

## **ABSTRACT**

*While much of the existing literature on teacher stress has traditionally focused on institutional pressures such as workload, student performance, and administrative demands, there is a growing recognition that personal circumstances significantly influence how stress is experienced and managed by university lecturers. This review synthesizes key findings from international studies to examine how marital status, caregiving responsibilities, personality traits, and gendered cultural expectations, acting as both significant sources of strain and potential buffers, impact EFL teacher stress levels. In addition, it was found that the sustained stress resulted in severe consequences, including increased burnout, declining job satisfaction, negative impacts on students' learning outcomes, and alarmingly high turnover rates. Based on the findings, some strategies to enhance EFL teacher well-being in higher education will be provided for the reference of instructors and institutions.*

**Keywords:** *EFL teacher stress, EFL teacher well-being, job satisfaction, personal factors/circumstances, students' learning outcomes*

*Ngày nhận bài: 12/11/2025; Ngày sửa bài: 30/12/2025; Ngày duyệt đăng: 02/04/2026*

## **1. Introduction**

Stress is a significant aspect of the modern lifestyle, and in many nations today, workers often find their jobs to be more stressful than they were in the past. According to Selye [1], a completely stress-free life is neither possible nor desirable, and a healthy dose of tension and stress can serve as a motivator and improve productivity at work. Stress may inspire us when it remains within healthy bounds. It pushes us to prepare, to focus, to improve. Yet extreme stress seems to cause bodily discomfort and misery [2]. It also turns destructive, leading to exhaustion, disengagement, and burnout [3].

Teachers in the teaching profession, particularly those who teach English as a foreign language (EFL), frequently endure significant levels of stress. Colleges, students, and society are putting more pressure and expectations on instructors. As stated by Kyriacou [4], teaching is just as stressful as being a police officer, a jail guard, an air traffic controller, a doctor, or a nurse. EFL teachers' work-related stress can be a significant contributor to various health issues, which can then have a negative impact on students and society. Consequently, identifying the causes of work-related stress and addressing the problem will help accelerate the learning

process and support the teaching community in maintaining their well-being.

Stress among university instructors can be significantly influenced by various factors. Although institutional pressures, such as workload, student performance, and administrative demands, have been the subject of significant international research recently, there is growing recognition that personal circumstances also have a substantial impact on how university lecturers experience and manage stress. Family responsibilities, caregiving duties, personality traits, coping mechanisms, and gendered cultural norms are a few examples of these individual-level impacts. Nevertheless, there remains a dearth of research that methodically summarizes these individual stressors among EFL instructors in higher education.

To address this gap, the present study adopts a narrative literature review approach to examine existing research on personal factors contributing to stress among EFL lecturers and explores their mechanisms of influence. Since stress and professional performance are closely intertwined [5], some methods to improve the well-being of EFL teachers in higher education will be offered for instructors' and institutions' reference, based on the findings.

The reviewed literature was drawn from established academic sources, including peer-reviewed journal articles, academic books, and selected review studies indexed in databases such as Google Scholar. The selection primarily focused on studies published from the early 2000s to 2025 that examined stress, coping, and related personal characteristics among

teachers, lecturers, or closely related educational professionals. While research specifically targeting EFL lecturers was prioritized, studies on teachers were also included where theoretical frameworks or empirical findings were relevant and transferable to the EFL higher education context.

In this review, the term “EFL lecturer(s)” refers to instructor(s) who teach EFL in higher education settings, where English is not the primary language. Although the review prioritizes studies conducted with university-level EFL lecturers, relevant research involving school-level educators is also referenced where appropriate. Rather than suggesting direct equivalence across educational levels or instructional environments, these studies are used to shed light on broadly equivalent emotional and psychological processes, such as stress evaluation, emotional regulation, and coping styles.

## 2. Personal stressors

### 2.1 Marital status

The relationship between marital status and teacher stress is complex. Some studies, including those by [6], suggest that marital status may not have a statistically significant effect. However, other research indicates that being married can increase perceived stress, especially when teachers struggle to balance work and family demands [7;8]. Many married EFL instructors bear this invisible emotional burden, particularly those with dependent children. They often juggle multiple roles, including caregiver, researcher, spouse, teacher, and parent.

From a mechanistic perspective, marital status influences stress primarily

through role strain and role overload. The simultaneous demands of work and family increase cognitive and emotional load, limit opportunities for psychological detachment from work-related stressors, and reduce time available for recovery. Over time, sustained role multiplicity may lead to chronic emotional tension and heightened vulnerability to stress and emotional exhaustion.

As Poloski and Bogdanic [7] note, marital status is not merely a demographic factor. It is an emotional context that shapes how stress is experienced. Taken together, existing evidence suggests that marital status may be an important personal factor associated with workplace stress among EFL lecturers, particularly in contexts characterized by strong work-family role demands.

### ***2.2. Caregiving responsibilities - Genders - Cultural expectations***

Caregiving responsibilities are a significant source of stress for EFL lecturers, particularly in sociocultural contexts where caregiving roles are strongly gendered. In many Asian societies, including Vietnam, it is still widely believed that women should be the primary caregivers. *Xian qi liang mu*, meaning “a virtuous wife and caring mother,” is a cherished yet heavy ideal in Chinese culture. Similar ideas are also subtly reinforced by societal norms in Vietnam. As a result, female lecturers often bear a larger share of household and childcare responsibilities in addition to their academic work [9; 10; 11]. For example, research during the COVID-19 pandemic showed that female faculty members with young children reported much higher stress. At one U.S. university, 60% of women, compared to

49% of men, said their stress increased significantly due to caregiving and work demands [12]. As Moen and Roehling [13] explain, women in their 30s often face rising professional expectations while family responsibilities peak. When eldercare duties, such as supporting aging parents, are added, the pressure can become overwhelming [14].

Researchers note that caregiving responsibilities influence stress through the Family-to-Work spillover process, in which family obligations intersect with career responsibilities. This spillover constrains lecturers’ ability to psychologically disengage from stressors, shortens recovery time, and undermines emotional regulation. Long-term exposure to this stress may lead to emotional exhaustion, declines in overall well-being, and increased susceptibility to burnout. Studies have shown that female academics with heavy caregiving responsibilities often report reduced research time, lower productivity, and increased emotional fatigue [15; 16].

But to be clear, this is not just a woman’s issue. Some studies, such as Sun and Dapat [17], have found that in certain contexts, male EFL professors report higher overall job stress than their female colleagues. A study in Nigeria reported that male lecturers perceived stress differently from female lecturers, indicating a gender difference in stress perception [18]. However, empirical evidence specifically addressing family-related stressors among male lecturers remains limited [18], underscoring the need for further research.

It is clear that gendered and cultural expectations strongly shape how teachers experience and cope with stress. Female

lecturers, especially mothers, often face what is known as a “double shift”: professional work during the day and domestic work at night, which, over time, can increase the risk of emotional exhaustion.

### **2.3. Personality traits and Coping mechanisms**

Not every stressor originates in the external world. Some stem from teachers’ personalities, thought processes, and coping mechanisms.

Perfectionism, often regarded as a virtue in academia, is one such personality-related factor. However, perfectionism has two distinct facets. Self-critical perfectionism, characterized by a persistent fear of failure or self-doubt, has been associated with higher stress levels and emotional exhaustion [19]. Conversely, perfectionistic aspirations combined with high self-efficacy and perceived control may contribute to greater resilience and lower stress [20]. At the individual level, maladaptive perfectionism may intensify stress by amplifying self-evaluative pressure and emotional reactivity, whereas more adaptive forms of perfectionism tend to support a sense of competence and emotional regulation. Accordingly, the influence of perfectionism on stress depends not only on high standards but also on how these standards are cognitively and emotionally managed.

Coping mechanisms further shape lecturers’ experiences of stress by determining how stressors are addressed or avoided. Prior research distinguishes between active (problem-focused) and passive (avoidant) coping [21]. Active coping involves planning, problem-solving, and reframing challenges in a

positive, constructive manner. Lecturers who adopt an active coping style tend to show higher engagement and emotional resilience. Passive coping strategies, such as avoidance or withdrawal, delay stress resolution and prolong emotional arousal, thereby increasing vulnerability to anxiety, disengagement, and emotional exhaustion. Coping styles influence whether situational stressors are effectively managed or gradually develop into chronic stress.

These findings suggest that personality traits and coping mechanisms mediate the relationship between stressors and stress-related outcomes among EFL lecturers. Individual differences in perfectionism and coping styles shape stress appraisal and emotional regulation processes, thereby influencing lecturers’ susceptibility to sustained stress.

## **3. Consequences of teacher stress**

Teacher stress does not exist in isolation. It affects how educators instruct, how learners acquire knowledge, and how organizations operate. To put it another way, stress has the power to subtly influence every aspect of education, from the classroom to the university system. In this review, the consequences are considered at two levels: their effect on teaching and learning, and their longer-term implications for burnout and institutional functioning.

### **3.1 Effects on teaching and learning**

Research indicates that elevated stress among lecturers may undermine effective teaching by intensifying negative emotional states, such as frustration, anxiety, and emotional fatigue, as reported by Collie et al. [22] and Khan et al. [23]. When EFL lecturers experience sustained emotional exhaustion, they may struggle

to maintain patience, creativity, and responsiveness in classroom interactions. Teaching may consequently become less engaging, emotionally detached, and more mechanical.

Empirical studies suggest that prolonged stress may lead to more negative work-related attitudes. For example, Mariani et al. [24] found that instructors who experience ongoing stress tend to adopt a more pessimistic attitude toward their work. As stress accumulates, teaching practices may become increasingly routine and emotionally detached, diminishing lecturers' sense of fulfillment and potentially affecting students' classroom experiences. Overall, existing research suggests that stress among EFL lecturers can indirectly influence teaching quality and learning processes through emotional regulation, motivation, and instructional engagement.

### ***3.2. Burnout and its consequences***

Burnout is a more severe and cumulative consequence of prolonged occupational stress. According to Cunningham [25], burnout is characterized by physical, emotional, and mental fatigue and is triggered by prolonged exposure to work-related stress. Johnson et al.'s additional study [26] supports the idea that long-term stress eventually leads to emotional exhaustion and disengagement. In higher education, teaching is often identified as a profession with heightened vulnerability to burnout due to its emotional, cognitive, and relational demands [27].

As burnout develops, lecturers' connection to work, and often their sense of purpose, begins to weaken. Burnout

can lead people to emotionally and cognitively distance themselves from their work, as Maslach and Leiter [28] explain, resulting in a loss of motivation and energy. For EFL lecturers, such disengagement can affect not only their teaching performance but also their long-term commitment to the profession, thereby increasing the likelihood of job attrition.

At a broader level, burnout among lecturers carries institutional consequences. Weng and Xi [29] demonstrate that high turnover not only lowers staff morale but also increases recruitment and training costs. The departure of experienced lecturers may disrupt mentoring relationships, weaken instructional continuity, and undermine institutional stability. Similarly, Redding and Henry [30] found that losing a teacher midyear can significantly disrupt learning consistency and educational quality, particularly in skill-based courses such as language teaching. Beyond these practical costs, there is also an invisible toll, a decline in the emotional climate of the institution. When colleagues are constantly leaving or disengaged, team spirit and collaboration weaken. For this reason, teacher well-being should be regarded as an organizational priority, not merely a personal matter. If lecturers are emotionally healthy, the institution thrives; if they are not, the entire system suffers. Thus, teacher well-being is not only a matter of individual survival but also a form of institutional sustainability.

To enhance the systematicity and clarity of the literature review, Table 1 provides a concise summary of the key studies examined in this review.

*Table 1. Summary of Key Studies on Personal Stressors and Consequences among EFL Lecturers*

No.	Author(s)	Year	Context	Focus	Key Findings
1.	Selye	1976	Stress theory	Nature of stress	Moderate stress can be motivating; prolonged stress becomes harmful.
2.	Kyriacou	2010	Teachers	Teacher stress	Teaching is among the most stressful professions.
3.	Erden et al.	2023	Teachers (meta-analysis)	Marital status and stress	Effects of marital status on stress are context-dependent.
4.	Thébaud et al.	2024	University faculty	Caregiving and stress	Female faculty reported higher caregiving-related stress.
5.	Sun & Daphat	2024	Chinese EFL lecturers	Job stress	Male EFL lecturers reported high stress in some contexts.
6.	Stoerber & Rennert	2008	Teachers	Perfectionism	Self-critical perfectionism is associated with increased stress and burnout.
7.	Nazari et al.	2023	EFL teachers	Coping strategies	Active coping reduces stress; avoidance increases exhaustion.
8.	Collie et al.	2012	Teachers	Teaching outcomes	Stress undermines teaching efficacy and engagement.
9.	Cunningham	1983	Teachers	Burnout	Burnout results from prolonged occupational stress.
10.	Maslach & Leiter	2016	Education & Care professions	Burnout	Burnout leads to disengagement and loss of meaning.
11.	Redding & Henry	2018	Teachers	Turnover	Stress-related attrition disrupts teaching continuity.

**4. Pedagogical implications**

Stress will always be a part of life and teaching, but how individuals and institutions handle it will determine whether it becomes a transformative or destructive force.

**4.1 Individual strategies**

University EFL instructors may benefit from proactively developing coping mechanisms and self-care strategies to manage personal stress more effectively. Previous studies have shown that mindfulness-based interventions, such as yoga, deep breathing, and meditation, significantly reduce instructors’ stress and emotional fatigue [31]. Additionally, planning, problem-solving, and seeking social support from colleagues, friends, and family appear to be more adaptive than avoidance-based or passive coping techniques. Research indicates that

coping skills are not fixed traits but can be developed through practice, and that targeted interventions may enhance teachers’ resilience and psychological well-being [32].

Establishing clear boundaries between work and personal life also plays an important role in stress regulation. For example, avoiding checking work emails outside official hours or dedicating time solely to rest and personal activities may help lecturers psychologically detach from work. Moreover, a healthy lifestyle that includes regular exercise, enough sleep, a balanced diet, and scheduled breaks enhances overall well-being and improves stress management.

While high standards are common in academic settings, excessive self-critical perfectionism has been linked to heightened stress and emotional

exhaustion. Thus, it is essential to acknowledge and address perfectionistic tendencies by setting realistic goals and cultivating self-compassion, rather than engaging in harsh self-criticism, which can help individuals remain motivated and prevent burnout [33;34]. Additionally, to share experiences and ease psychological burdens, lecturers are encouraged to participate in peer support groups or seek professional counseling if stress becomes overwhelming. Through these strategies, educators can better manage personal stressors and sustain high levels of instructional effectiveness.

### **4.2. Institutional strategies**

Individual stress management efforts do not occur in isolation but are influenced by the institutional environments in which lecturers carry out their professional roles.

To support the mental health of EFL instructors, universities may consider implementing comprehensive policies and initiatives that promote mental well-being. First, these may include regular opportunities for professional development and stress-management training, such as stress-resilience workshops, cognitive-behavioral techniques, and mindfulness-based seminars. Comprehensive evaluations show that mindfulness therapies can significantly reduce teacher stress in various international settings [31]. Additionally, family-friendly practices such as flexible teaching schedules, caregiving leave, and access to childcare support can help lecturers better manage work-family demands. Research indicates that organizations that promote work-family balance see improvements in faculty burnout and well-being

[35]. Workload management is also an important institutional consideration. Regular reviews of teaching and administrative responsibilities, clearer role definitions, and more balanced task distribution may reduce role overload and role ambiguity, which are known contributors to occupational stress.

### **5. Conclusion**

This narrative literature review has examined stress among EFL lecturers in higher education, with a specific focus on personal-level determinants. The review highlights how marital status, caregiving responsibilities, personality traits, coping mechanisms, and gendered cultural expectations shape lecturers' experiences of stress in academic contexts. The evidence indicates that personal stressors operate through identifiable mechanisms, including role strain, Family-to-Work spillover, maladaptive emotion regulation, and ineffective coping styles. Married lecturers and those with caregiving responsibilities, particularly women in sociocultural contexts where caregiving is strongly gendered, often experience heightened stress due to competing professional and family demands. Additionally, individual differences in perfectionism and coping strategies influence stress appraisal and emotional regulation, which can lead to either emotional exhaustion or psychological resilience.

Sustained stress may have consequences that extend beyond individual well-being. As discussed, prolonged stress can undermine instructional engagement, reduce emotional availability in the classroom, and increase vulnerability to burnout among EFL lecturers. Over time, burnout

may weaken professional commitment and contribute to lecturer attrition, with potential implications for instructional continuity and institutional stability.

Based on these findings, this review emphasizes pedagogical implications at both individual and institutional levels. Developing adaptive coping strategies, setting boundaries between work and personal life, and addressing self-critical perfectionism can help lecturers manage personal stress more effectively. A supportive institutional environment can further enhance the effectiveness of these individual efforts.

This review also suggests that lecturer stress may be better understood when personal life circumstances, psychological resources, and professional demands are considered together rather than in isolation. To further develop this line of inquiry, further studies could consider longitudinal approaches and pay more attention to groups that have received less research attention, such as male lecturers' family-related stress, in order to gain a clearer picture of how personal stress may change over time.

## REFERENCES

1. Selye, H. (1976). *Stress in Health and Disease*. Butterworths, Boston.
2. Cedoline, A. J. (1982). *Job burnout in public education: symptoms, causes, and survival skills*. Teachers College Press.
3. Arbabisarjou, A., Ajdari, Z., Omeidi, K., & Jalalinejad, R. (2013). The relationship between Job stress and performance among the hospitals Nurses. *World Sci J*, 1, 181-8.
4. Kyriacou, C. (2010). Teacher Stress: Directions for future research. *Educational Review*, 53(1), 27–35. <https://doi.org/10.1080/00131910120033628>
5. Vargas Valdiviezo, Ma. A., & Guevara, R. (2025). A new perspective of work stress on teaching performance by competencies. *International Journal of Leadership in Education*, 28(5), 947–965. <https://doi.org/10.1080/13603124.2022.2052755>
6. Erden, A., Erden, H., & Aytaç, T. (2023). Teachers' well-being levels by gender and marital status: A meta-analysis study. *Asian Journal of Instruction*, 11(Special Issue), 38-60. <https://doi.org/10.47215/aji.1304646>
7. Poloski Vokic, Nina & Bogdanic, Ana. (2007). Individual Differences and Occupational Stress Perceived: A Croatian Survey. Faculty of Economics and Business, University of Zagreb, EFZG Working Papers Series. 11.
8. Sadeghi, K., & Sa'adatpourvahid, M. (2016). EFL Teachers' Stress and Job Satisfaction: What Contribution Can Teacher Education Make?. *Iranian Journal of Language Teaching Research*, 4(Issue 3 (Special Issue)), 75-96. <https://doi.org/10.30466/ijltr.2016.20355>
9. Chesley, N. (2017). What does it mean to be a “breadwinner” mother? *Journal of family issues*, 38 (18), 2594–2619. <https://doi.org/10.1177/0192513X16676857>
10. Donehower, G., Tovar, J.A., & Urdinola, B.P. (2019). Time use differences and similarities between developed and emerging economies in the Americas. In: B.P. Urdinola and J. A. Tovar, eds. *Time use and transfers in the Americas*. Springer International Publishing, 41–55. [https://doi.org/10.1007/978-1-4939-9888-8\\_3](https://doi.org/10.1007/978-1-4939-9888-8_3)

org/10.1007/978-3-030-11806-8\_3

11. Guo, Y. & Chu, L. (2019). Factors Hindering the Academic Career Development of Female Teachers in Agricultural and Forestry Universities of Yunnan Province: Empirical Study Based on Heckman's Two-Stage Selection Model[J]. *Journal of Yunnan Agricultural University (Social Science)*, 13(4), 142-150. DOI: 10.3969/j.issn.1004-390X(s).2019.04.023

12. Thébaud, S., Hoppen, C., David, J., & Boris, E. (2024). Understanding Gender Disparities in Caregiving, Stress, and Perceptions of Institutional Support among Faculty during the COVID-19 Pandemic. *Social Sciences*, 13(4), 181. <https://doi.org/10.3390/socsci13040181>

13. Moen, P., & Roehling, P. (2005). *The career mystique: Cracks in the American dream*. Rowman & Littlefield.

14. Duan, H. (2021). How Do Caregiving Responsibilities Affect Women's Work-Family Spillovers Over the Life Course? *Journal of Family Issues*, 43(8), 2181-2209. <https://doi.org/10.1177/0192513X211030032> (Original work published 2022)

15. Deryugina, Tatyana, Olga Shurchkov, and Jenna Stearns. "COVID-19 Disruptions Disproportionately Affect Female Academics." *AEA Papers and Proceedings* 111 (May 1, 2021): 164-68. <https://doi.org/10.1257/pandp.20211017>.

16. Davis, J. C., Li, E. P. H., Butterfield, M. S., DiLabio, G. A., Santhagunam, N., & Marcolin, B. (2022). Are we failing female and racialized academics? A Canadian national survey examining the impacts of the COVID-19 pandemic on

tenure and tenure-track faculty. *Gender, work, and organization*, 29(3), 703-722. <https://doi.org/10.1111/gwao.12811>

17. Sun, W., & Dapat, R. (2024). Unraveling Job Stress, Burnout, and Psychological Capital among Chinese EFL Teachers in Higher Institutions. *International Education Studies*, 17(5), 29-44.

18. Chukwuemeka, U. M., Okonkwo, U. P., Njoku, C. J., Igwe, S. E., Oyewumi, T. J., & Ugwuanyi, D. C. (2023). Work-related stress, quality of life, and coping mechanisms among lecturers in a tertiary educational institution in Anambra State, Nigeria. *BMC Psychology*, 11, 73. <https://doi.org/10.1186/s40359-023-01114-5>

19. Stoeber, J., & Rennert, D. (2008). Perfectionism in school teachers: relations with stress appraisals, coping styles, and burnout. *Anxiety, stress, and coping*, 21(1), 37-53. <https://doi.org/10.1080/10615800701742461>

20. Samfira, E. M., & Maricuțoiu, L. P. (2021). Not all Perfectionists Are as They Are Assessed: An Investigation of the Psychometric Properties of the Perfectionism Inventory in the Teaching Profession. *Frontiers in psychology*, 12, 624938. <https://doi.org/10.3389/fpsyg.2021.624938>

21. Nazari, F., Ghanizadeh, A. & Mirzaee, S. (2023). EFL teachers' coping strategies amidst the Covid-19 virtual education and their association with work engagement and teacher apprehension. *Educ Res Policy Prac* 22, 1-22. <https://doi.org/10.1007/s10671-022-09317-0>

22. Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social-emotional learning: Predicting

teacher stress, job satisfaction, and teaching efficacy. *Journal of educational psychology*, 104(4), 1189.

23. Khan, A., Shah, I. M., Khan, S., & Gul, S. (2012). Teachers' stress, performance & resources. *International Review of Social Sciences and Humanities*, 2(2), 10-23.

24. Mariani, A. M., Piccini, L., & Melchiori, F. M. (2020). Protective factors for teachers' work stress: Psychoeducational programs based on self-efficacy and hope to reinforce personal resources. *Italian Journal of Educational Research*, (25), 127-136.

25. Cunningham, W. G. (1983). Teacher burnout-solutions for the 1980s: A review of the literature. *The Urban Review*, 15(1), 37-51. <https://doi.org/10.1007/BF01112341>

26. Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P., & Millet, C. (2005). The experience of work-related stress across occupations. *Journal of managerial psychology*, 20(2), 178-187.

27. Skaalvik, E. M., & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological reports*, 114(1), 68-77.

28. Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, 15(2), 103-111. <https://doi.org/10.1002/wps.20311>

29. Weng, Q., & Xi, Y. (2010). Career growth and turnover intention: Moderating effect of vocational commitment and perceived opportunity. *Nankai Business Review*, 13, 119-131.

30. Redding, C., & Henry, G. T. (2018). Leaving School Early: An Examination

of Novice Teachers' Within- and End-of-Year Turnover. *American Educational Research Journal*, 56(1), 204-236. <https://doi.org/10.3102/0002831218790542> (Original work published 2019)

31. Agyapong, B., Brett-MacLean, P., Burbach, L., Agyapong, V. I. O., & Wei, Y. (2023). Interventions to Reduce Stress and Burnout among Teachers: A Scoping Review. *International journal of environmental research and public health*, 20(9), 5625. <https://doi.org/10.3390/ijerph20095625>

32. Samfira, E. M., & Paloş, R. (2021). Teachers' Personality, Perfectionism, and Self-Efficacy as Predictors for Coping Strategies Based on Personal Resources. *Frontiers in Psychology*, 12, 751930. <https://doi.org/10.3389/fpsyg.2021.751930>

33. Sauve, J. A. (2017). *Building teacher resilience: relations of self-compassion among teacher efficacy and burnout* (T). Electronic Theses and Dissertations (ETDs) 2008+. University of British Columbia. Retrieved October 21, 2025, from <https://open.library.ubc.ca/collections/ubctheses/24/items/1.0356624>

34. Hu, Y. (2023). Examining the effects of teacher self-compassion, emotion regulation, and emotional labor strategies as predictors of teacher resilience in EFL context. *Frontiers in Psychology*, 14, 1190837.

35. Koster, M., & McHenry, K. (2023). Areas of work-life that contribute to burnout among higher education health science faculty and perception of institutional support. *International Journal of Qualitative Studies on Health and Well-Being*, 18(1), 2235129. <https://doi.org/10.1080/17482631.2023.2235129>