

MÔ HÌNH GIÁM SÁT VÀ DỰ BÁO THÔNG MINH CHO RAU MÁ THỦY CANH DỰA TRÊN MACHINE LEARNING (ML) VÀ DEEP LEARNING (DL)

Lê Quyết Thắng^{1,*}, Lê Thị Thu Hương²

¹Trường Đại học Công nghiệp Quảng Ninh

²Trung tâm giáo dục nghề nghiệp giáo dục thường xuyên Đông Triều

*Email: lequyetthang5282@qui.edu.vn

TÓM TẮT

Bài báo này giới thiệu một mô hình giám sát và dự báo thông minh cho cây rau má trong hệ thống thủy canh, được phát triển dựa trên các kỹ thuật Machine Learning (ML) và Deep Learning (DL). Hệ thống sử dụng các cảm biến IoT để thu thập dữ liệu thời gian thực về những thông số môi trường quan trọng như pH dung dịch, độ dẫn điện (EC), nhiệt độ, độ ẩm không khí và cường độ ánh sáng. Dữ liệu sau đó được truyền lên nền tảng đám mây, nơi các mô hình ML và DL tiến hành xử lý, phân tích và dự báo. Điểm mới của nghiên cứu là lần đầu tiên áp dụng đồng thời các mô hình hồi quy, phân loại và chuỗi thời gian (RNN/LSTM) cho rau má thủy canh, nhằm dự đoán biến động của các thông số môi trường và đánh giá tình trạng sinh trưởng của cây. Ngoài ra, hệ thống còn hỗ trợ cảnh báo sớm và đưa ra khuyến nghị hoặc điều khiển tự động, góp phần nâng cao hiệu quả canh tác, cải thiện chất lượng sản phẩm và tối ưu hóa việc sử dụng tài nguyên so với phương pháp thủy canh truyền thống.

Từ khóa: Thủy canh, cảm biến, Internet Vạn vật, dữ liệu thời gian thực, Trí tuệ nhân tạo, Học máy, Học sâu.

1. GIỚI THIỆU

1.1 Bối cảnh và tính cấp thiết

Trong bối cảnh biến đổi khí hậu, suy giảm diện tích đất nông nghiệp và nhu cầu ngày càng tăng về thực phẩm sạch, bền vững, các mô hình canh tác mới nổi lên như một giải pháp quan trọng. Thủy canh (hydroponics) đã được chứng minh là một phương pháp canh tác tiên tiến, có khả năng tiết kiệm nước tới 90%, giảm phụ thuộc vào đất canh tác, kiểm soát được môi trường sinh trưởng và hạn chế tối đa việc sử dụng thuốc bảo vệ thực vật. Điều này giúp thủy canh trở thành lựa chọn hàng đầu trong nông nghiệp công nghệ cao và nông nghiệp đô thị.

Rau Má (*Centella asiatica*) là loại cây có giá trị kinh tế và dược liệu cao, được sử dụng rộng rãi trong thực phẩm chức năng, dược phẩm và mỹ phẩm. Tuy nhiên, canh tác rau má theo phương pháp truyền thống gặp nhiều thách thức: phụ thuộc vào điều kiện đất, độ ẩm, ánh sáng tự nhiên và dễ bị sâu bệnh. [1] Các nghiên cứu như của Archana Prasad (2012) đã khẳng định tiềm năng kinh tế và dược liệu lớn của rau má, nhưng các công trình khoa học về canh tác bằng thủy canh còn hạn chế. Điều này cho thấy nhu cầu cấp

thiết phải ứng dụng các công nghệ tiên tiến nhằm nâng cao năng suất và chất lượng cây rau má.



Hình 1: Trang trại rau má Thủy canh tại phường Bình Khê, Quảng Ninh.

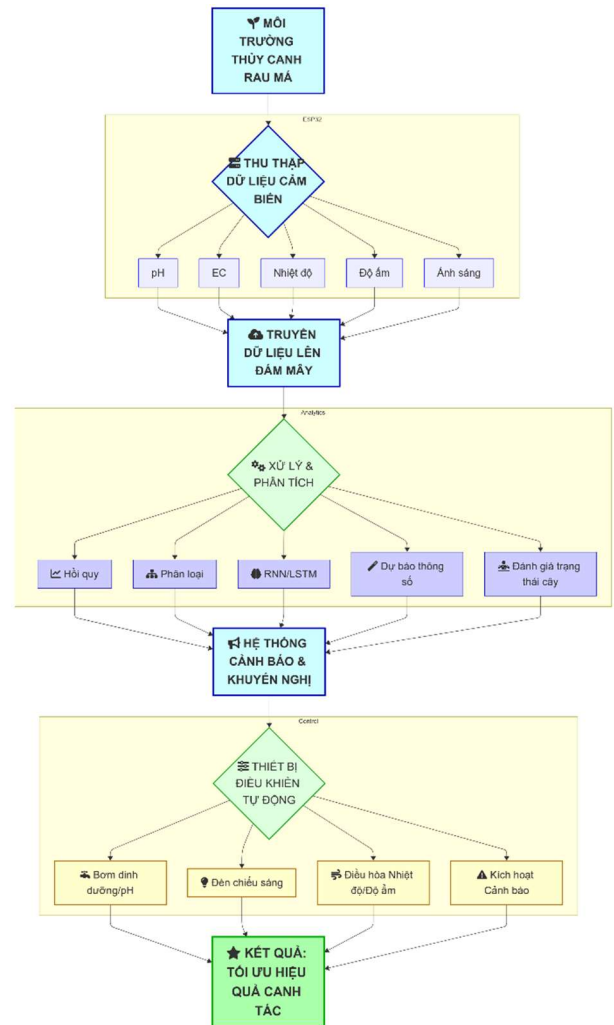
1.2 Ứng dụng IoT và AI trong nông nghiệp

Sự phát triển của Internet of Things (IoT) cho phép giám sát liên tục các thông số môi trường trong hệ thống thủy canh, bao gồm pH dung dịch, độ dẫn điện (EC), nhiệt độ dung dịch, nhiệt độ và độ ẩm không khí, cũng như cường độ ánh sáng. Tuy nhiên, dữ liệu thu được từ các cảm biến này thường có khối lượng lớn, biến động theo thời gian và khó khai thác hiệu quả nếu chỉ dựa vào phân tích truyền thống.

Trong thập kỷ qua, nhiều nghiên cứu đã chứng minh tiềm năng của Machine Learning (ML) và Deep Learning (DL) trong nông nghiệp thông minh. Chẳng hạn, [2] Eli-Chukwu (2019) chỉ ra rằng AI có thể hỗ trợ dự báo năng suất và phát hiện bệnh trên cây trồng nhờ phân tích dữ liệu IoT; [3] Umar Shareef et al. (2024) đã tích hợp ML vào hệ thống thủy canh để tự động điều chỉnh dinh dưỡng, bước đầu mang lại kết quả khả quan; [4] Abishkar Regmi (2024) nhấn mạnh vai trò của ML trong quản lý dinh dưỡng và tối ưu hóa nguồn tài nguyên.

Đặc biệt, các mô hình học sâu như LSTM (Long Short-Term Memory: là một mô hình có khả năng xử lý chuỗi dữ liệu như dữ liệu cảm biến theo thời gian, bằng cách sử dụng bộ nhớ dài hạn để lưu trữ các quy tắc và xu hướng tổng thể, đồng thời liên tục cập nhật bộ nhớ ngắn hạn với các sự kiện mới xảy ra) đã chứng minh hiệu quả vượt trội trong việc xử lý dữ liệu cảm biến dạng chuỗi thời gian. Chúng có khả năng học được xu hướng dài hạn và các mối quan hệ phi tuyến phức tạp, từ đó dự đoán chính xác sự biến động môi trường trong nhà kính và hệ thống thủy canh [5]. Ngoài ra, nhiều nghiên cứu đã ứng dụng LSTM cho phát hiện sớm bất thường trong dữ liệu cảm biến, giúp tăng độ tin cậy cho các hệ thống giám sát nông nghiệp thông minh [6]. Cũng có báo cáo cho thấy việc tích hợp LSTM vào các chiến lược điều khiển tưới hoặc cung cấp dinh dưỡng đã giúp tối ưu hóa hiệu quả sử dụng tài nguyên và cải thiện điều kiện sinh trưởng cây trồng [7]. Gần đây, một số công trình đã mở rộng ứng dụng LSTM cho cả phát hiện bất thường và xử lý dữ liệu khiếm khuyết trong hệ thống nông nghiệp thông minh, củng cố vai trò của LSTM như một phương pháp toàn diện cho quản lý thời gian thực và chất lượng hệ thống [8].

Sơ đồ trên hình 2 thể hiện hệ thống IoT giám sát và điều khiển môi trường thủy canh rau má, bao gồm các cảm biến pH, EC, nhiệt độ, độ ẩm, ánh sáng và bộ điều khiển trung tâm ESP32 kết nối qua Internet. Các dữ liệu thu thập từ cảm biến và nhập liệu sử dụng các thuật toán thông minh để tạo ra mô hình AI nhằm tác động vào hệ thống để tạo ra hiệu suất cao trong mô hình.



Hình 2: Sơ đồ thuật toán hệ thống.

1.3. Hạn chế trong nghiên cứu hiện tại và điểm mới

Qua phân tích các công trình liên quan, có thể nhận thấy rằng mặc dù IoT và AI đã được ứng dụng rộng rãi trong nông nghiệp thông minh, song vẫn còn tồn tại nhiều hạn chế trong nghiên cứu hiện tại. Cụ thể, chưa có nhiều công trình tập trung vào cây rau má thủy canh – một loại cây đặc thù có giá trị dược liệu và kinh tế cao. Các hệ thống giám sát hiện nay chủ yếu dừng lại ở việc thu thập dữ liệu cảm biến, trong khi việc khai thác dữ liệu bằng các mô hình dự đoán tiên tiến (AI/ML/DL) vẫn chưa được phát triển mạnh. Ngoài ra, giải pháp tích hợp đa nguồn dữ liệu (cảm biến môi trường, mô hình AI và dashboard trực quan) để hỗ trợ ra quyết định tự động trong quản lý thủy canh còn chưa phổ biến.

Trên cơ sở đó, nghiên cứu này đề xuất một hướng tiếp cận mới với các điểm khác biệt sau:

Xây dựng hệ thống IoT giám sát môi trường thủy canh rau má theo thời gian thực, đảm bảo thu thập liên tục các thông số pH, EC, nhiệt độ, độ ẩm và ánh sáng.

Ứng dụng đồng thời các mô hình Machine Learning (hồi quy, phân loại) và Deep Learning (RNN/LSTM) để dự đoán biến động môi trường, phân loại sức khỏe cây trồng và dự báo năng suất.

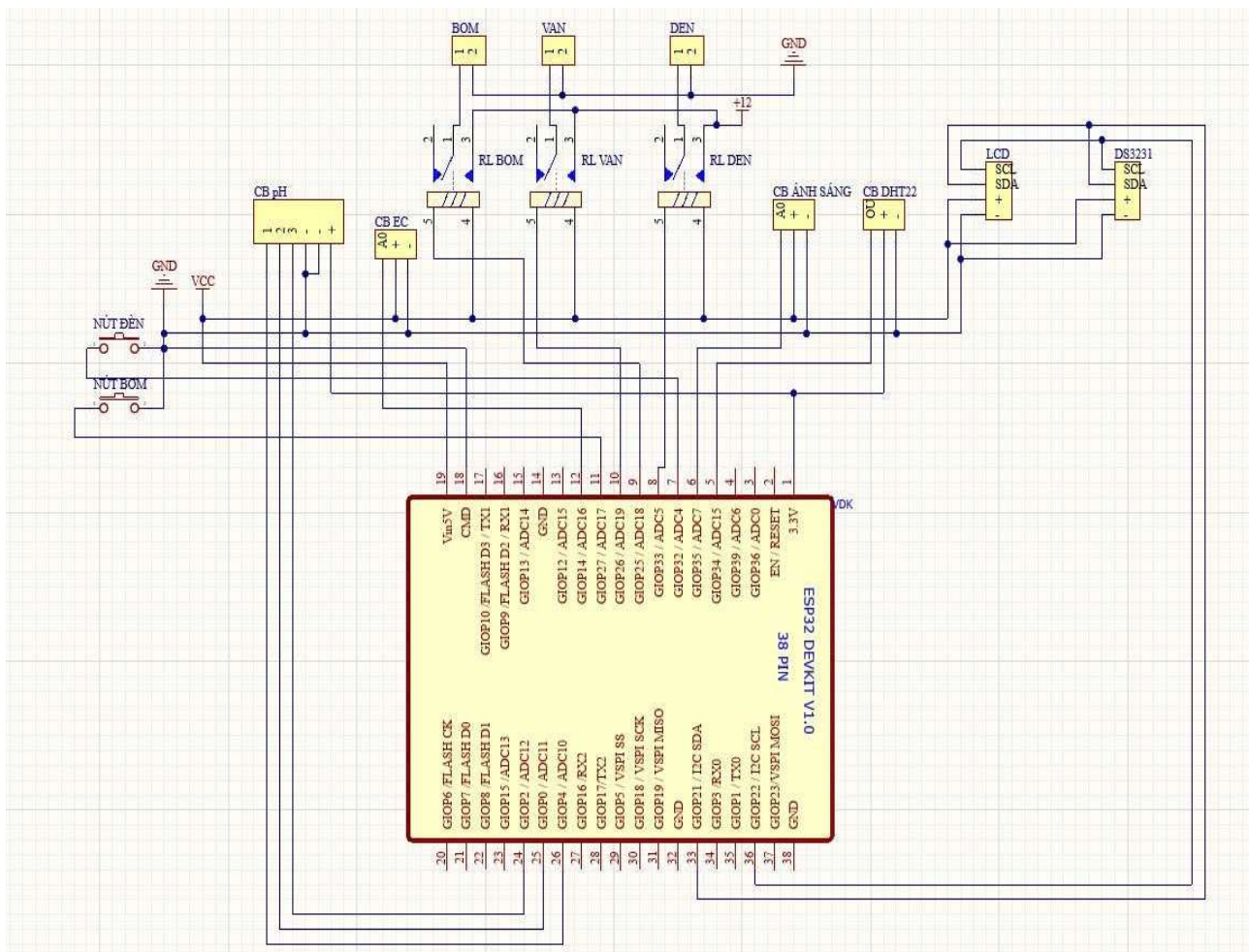
Phát triển cơ chế khuyến nghị và điều khiển tự động (như bơm dung dịch dinh dưỡng, điều chỉnh ánh sáng nhân tạo, tối ưu tỷ lệ dinh dưỡng) dựa trên kết quả phân tích AI.

Đánh giá hiệu quả canh tác có sự hỗ trợ của AI so với phương pháp thủy canh truyền thống, nhằm chứng minh tính khả thi, giá trị ứng dụng và tiềm năng nhân rộng mô hình.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Xác định bài toán

Xuất phát từ phân tích điểm mới nêu trên, bài báo này đề xuất xây dựng và đánh giá một hệ thống giám sát thông minh toàn diện cho mô hình canh tác rau má thủy canh. Mục tiêu cốt lõi của nghiên cứu là ứng dụng các thuật toán ML và DL để giải quyết các bài toán cụ thể sau:



Hình 3: Sơ đồ mạch điều khiển và xử lý thông tin.

Hình 3. Thể hiện thiết kế hệ thống phần cứng, được xây dựng dựa trên sơ đồ mạch nguyên lý với các khối chính:

Vi điều khiển ESP32: trung tâm xử lý dữ liệu, tích hợp WiFi để truyền dữ liệu qua MQTT.

Khối cảm biến: bao gồm cảm biến pH, EC, nhiệt độ dung dịch, nhiệt độ/độ ẩm không khí, ánh

sáng; dữ liệu được đưa vào các kênh ADC.

Khởi chấp hành: các relay điều khiển bơm dinh dưỡng, bơm bổ sung, van điện từ, và đèn chiếu sáng.

Nguồn cấp: 5V và 3.3V cho ESP32 và cảm biến; tải AC/DC được cách ly thông qua relay.

Chức năng và nguyên lý hoạt động của mạch bao gồm: Thu thập dữ liệu: ESP32 nhận dữ liệu từ các cảm biến pH, EC, nhiệt độ, độ ẩm, ánh sáng.

Truyền thông: dữ liệu được gửi thời gian thực qua giao thức MQTT tới broker và nền tảng Cloud/Dashboard.

Giám sát & cảnh báo: dữ liệu được trực quan hóa trên Dashboard IoT; hệ thống kích hoạt cảnh báo kép qua Telegram và Zalo (webhook trung gian) khi phát hiện bất thường.

Điều khiển tự động: ESP32 điều khiển relay để bật/tắt bơm, van, đèn dựa trên các ngưỡng cài đặt hoặc khuyến nghị từ mô hình AI.

2.2. Xây dựng và huấn luyện các mô hình Machine Learning (Hồi quy, Phân loại) và Deep Learning (RNN/LSTM).

1. Mô hình hồi quy (Regression Models) mô tả tổng quát lý thuyết về hồi quy tuyến tính:

Mục tiêu: dự đoán giá trị liên tục (ví dụ: pH, EC, năng suất rau má).

Phương trình tổng quát:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

Trong đó y là biến phụ thuộc (ví dụ pH), x_i là dữ liệu cảm biến, β_i là hệ số mô hình, ε là sai số. mô hình tìm bộ trọng số β sao cho đường dự đoán khớp nhất với dữ liệu cảm biến.

Mô hình học các hệ số β bằng cách tối thiểu hóa sai số bình phương trung bình (MSE):

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (2)$$

Ứng dụng: dự báo xu hướng pH giảm khi cây hút dinh dưỡng hoặc EC tăng khi bổ sung dung dịch, từ đó đưa cảnh báo điều chỉnh kịp thời.

2. Mô hình phân loại (Classification Models):
Mục tiêu: phân loại tình trạng cây rau má (Khỏe mạnh – Thiếu dinh dưỡng – Nhiễm bệnh).
Xác suất một mẫu thuộc lớp 1 được mô hình hóa bằng hồi quy logistic:

$$P(y=1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (3)$$

y : biến mục tiêu (label). Trong bài toán phân loại nhị phân (binary classification), y chỉ có 2 giá

trị:

$y=1$: cây rau má thuộc lớp “Khỏe mạnh”.

$y=0$: cây rau má thuộc lớp “Không khỏe” (thiếu dinh dưỡng hoặc bệnh).

$x=(x_1, x_2, \dots, x_n)$: vector đặc trưng đầu vào (dữ liệu cảm biến: pH, EC, nhiệt độ, độ ẩm, ánh sáng...).

$P(y=1|x)$: xác suất để cây khỏe mạnh khi ta quan sát được giá trị cảm biến x .

Giả sử hệ thống nhận được dữ liệu cảm biến:

pH = 6.2

EC = 1.8 mS/cm

Nhiệt độ dung dịch = 24 °C

Độ ẩm không khí = 70%

Mô hình logistic regression tính toán:

$P(y=1|x)=0.82$

→ Có nghĩa là: với dữ liệu này, cây rau má có 82% xác suất khỏe mạnh.

Nếu mô hình đặt ngưỡng 0.5 thì dự đoán “Khỏe mạnh”, nhưng nếu đặt ngưỡng cảnh báo cao hơn (0.9) thì vẫn cảnh báo người trồng là “cần theo dõi thêm”.

Ngưỡng phân loại thường đặt tại 0.5: nếu $P \geq 0.5$ thì dự đoán cây khỏe, ngược lại cây có nguy cơ thiếu dinh dưỡng hoặc bệnh lý.

Hàm mất mát dùng để tối ưu là log-loss (binary cross-entropy):

$$L = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

Trong đó:

m : số mẫu trong tập dữ liệu.

$y_i \in \{0, 1\}$: nhãn thực tế (0 = cây không khỏe, 1 = cây khỏe).

$y^i = P(y=1/x_i)$: xác suất mô hình dự đoán cây khỏe ở mẫu i .

Ứng dụng trong hệ thống thủy canh rau má:

Khi phân loại Khỏe / Thiếu dinh dưỡng / Bệnh lý, log-loss giúp mô hình không chỉ dự đoán nhãn, mà còn học cách ước lượng xác suất đáng tin cậy.

Cây A: $P(y=1|x)=0.95 \rightarrow$ mô hình gần như chắc chắn cây khỏe.

Cây B: $P(y=1|x)=0.55P \rightarrow$ mô hình dự đoán khỏe nhưng không chắc chắn $\rightarrow$ hệ thống có thể phát cảnh báo “theo dõi thêm”.

Ứng dụng thực tế: khi hệ thống nhận thấy cây có dấu hiệu thiếu dinh dưỡng (dựa vào EC thấp kéo dài, tốc độ sinh trưởng chậm), mô hình sẽ phân loại và phát cảnh báo cho người vận hành.

3. Mô hình RNN / LSTM (Deep Learning chuỗi thời gian)

RNN: Các cảm biến (pH, EC, nhiệt độ dung dịch, nhiệt độ & độ ẩm không khí, ánh sáng) gửi dữ liệu liên tục về ESP32.

RNN dùng để học chuỗi ngắn hạn: ví dụ dự đoán giá trị pH sau 2–3 giờ tiếp theo dựa vào dữ liệu 24 giờ gần nhất.

Hữu ích để phát hiện bất thường ngắn hạn như sự tụt pH đột ngột khi bơm dinh dưỡng gặp sự cố.

LSTM: Với rau má, chu kỳ sinh trưởng kéo dài nhiều tuần, vì vậy cần học xu hướng dài hạn.

LSTM lưu giữ được thông tin ẩn trong nhiều ngày $\rightarrow$ dự báo chính xác sự thay đổi EC hoặc pH kéo dài, vốn ảnh hưởng trực tiếp đến khả năng hấp thụ dinh dưỡng của cây.

LSTM cũng được dùng để ước tính tốc độ sinh trưởng và năng suất rau má sau 1–2 tuần, dựa trên dữ liệu lịch sử canh tác (ví dụ: chuỗi EC, nhiệt độ, ánh sáng và tốc độ tăng khối lượng lá).

4. Quy trình huấn luyện: Dữ liệu đầu vào: Khoảng 24.000 mẫu từ 5 giàn thủy canh (45 phút/lần, trong 3 tháng).

Rau Ma Hydroponics Sensor Dataset v1 (2025)

Lê, Thắng

DỮ LIỆU CẢM BIẾN CHO FARM RAU MÁ THỦY CANH

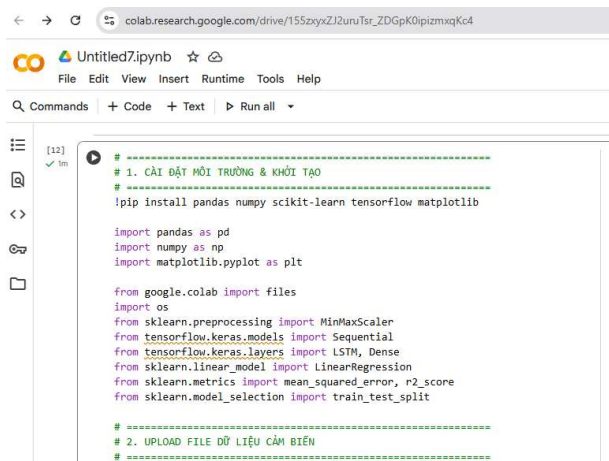
Files

timestamp	device_id	ph	ec_ms_cm	temp_solution_c	temp_air_c	rh_percent	light_lux
2025-03-01T00:00:00	GIAN_01	6.79	1.38	25.5	24.8	63.2	10639
2025-03-01T00:00:00	GIAN_02	6.03	2.4	27.1	34.1	70.3	8205
2025-03-01T00:00:00	GIAN_03	6.41	1.56	23.0	25.2	76.7	6592
2025-03-01T00:00:00	GIAN_04	5.92	1.96	25.4	35.0	81.2	5745
2025-03-01T00:00:00	GIAN_05	6.6	1.76	27.7	32.3	87.3	7298
2025-03-01T01:00:00	GIAN_01	6.27	2.02	24.2	32.6	67.2	7843

Hình 4: Đưa dữ liệu lên Google Dataset Search

Bao gồm: pH, EC, nhiệt độ dung dịch, nhiệt độ & độ ẩm không khí, ánh sáng, nhật ký tưới...như trên hình 4.

Phần lập trình sử dụng phương pháp Mô hình hóa Chuỗi Thời Gian (LSTM). Hệ thống cốt lõi dựa trên Mạng Nơ-ron Hồi quy Dài-Ngắn Hạn (Long Short-Term Memory - LSTM), một kiến trúc RNN (Recurrent Neural Network) là một loại mạng nơ-ron nhân tạo được thiết kế đặc biệt để xử lý chuỗi dữ liệu (sequential data), nơi thứ tự và mối quan hệ giữa các phần tử trong chuỗi là rất quan trọng tạo hiệu quả cao trong việc nắm bắt sự phụ thuộc theo thời gian của dữ liệu chuỗi cảm biến.



Hình 5: Lập trình huấn luyện dựa trên giá trị cảm biến

Thiết lập Mô hình và Thuật toán Tối ưu:

Mô hình Dự báo Tối ưu: Stacked LSTM Đa biến có Chọn lọc

Kiến trúc Mô hình: Dự án sử dụng hai mô hình Stacked LSTM Đa biến có chọn lọc độc lập để dự báo 1 bước thời gian tiếp theo cho pH và EC (mS/cm). Kiến trúc này bao gồm hai lớp LSTM xếp chồng lên nhau, tối ưu hóa khả năng nắm bắt các mối quan hệ phi tuyến tính phức tạp và tính phụ thuộc thời gian trong chuỗi dữ liệu cảm biến.

Chiến lược đa biến, lọc biến: Để đảm bảo mô hình hoạt động hiệu quả nhất, các biến đầu vào được chọn lọc dựa trên cơ sở khoa học và vật lý:

Dự đoán pH: Mô hình nhận dữ liệu pH và ánh sáng (light_lux). Mối quan hệ này được thiết lập do quá trình quang hợp của cây (được thúc đẩy bởi ánh sáng) ảnh hưởng trực tiếp đến sự hấp thụ ion và làm thay đổi pH của dung dịch.

Dự đoán EC: Mô hình nhận dữ liệu EC, Nhiệt độ Dung dịch (temp_solution_c), và pH. Sự lựa chọn này dựa trên việc nhiệt độ dung dịch ảnh hưởng trực tiếp đến độ dẫn điện (tốc độ di chuyển của ion), đồng thời pH là chỉ số phản ánh sự hấp thụ dinh dưỡng của cây.

Tham số huấn luyện: Mô hình được huấn luyện bằng Adam optimizer với một learning rate cố định. Đồng thời, sử dụng cơ chế Early Stopping để theo dõi validation loss, giúp mô hình dừng huấn luyện tại thời điểm tối ưu, tránh hiện tượng overfitting với dữ liệu.

Hàm Mất mát và Đánh giá: MSE (Mean Squared Error) được sử dụng làm hàm mất mát trong quá trình huấn luyện. Hiệu suất dự báo được đánh giá bằng MSE và RMSE (Root Mean Squared Error), tập trung vào độ chính xác của các dự báo chuỗi pH và EC trên tập dữ liệu kiểm tra.

Giá trị cốt lõi: Mục tiêu cuối cùng là sử dụng dự báo để chuyển đổi hệ thống từ cơ chế phản ứng (Reactive) sang cơ chế phòng ngừa (Proactive), từ đó duy trì sự ổn định môi trường liên tục và tối ưu hóa điều kiện sinh trưởng.

Đánh giá Hồi quy cơ sở: Song song với LSTM, mô hình hồi quy tuyến tính (Linear Regression) được sử dụng như mô hình cơ sở (Baseline) để đánh giá mối quan hệ tuyến tính giữa các biến.

Dự đoán pH: Các thông số EC, Nhiệt độ, Độ ẩm, và Ánh sáng được dùng để dự đoán pH.

Dự đoán EC: Thông số pH và các biến môi trường khác được dùng để dự đoán EC.

Việc so sánh R2 và MSE giữa hai phương pháp (LSTM và Hồi quy tuyến tính) giúp khẳng định ưu thế của LSTM trong việc xử lý tính phi tuyến và mối quan hệ phức tạp của dữ liệu cảm biến nông nghiệp.

Giá trị cốt lõi: Việc sử dụng dự báo cho phép hệ thống phòng ngừa sự cố thay vì chỉ phản ứng, duy trì sự ổn định môi trường liên tục.

Ứng dụng thực tế và logic điều khiển tự động:

Kết quả dự báo ngắn hạn từ LSTM được chuyển trực tiếp thành lệnh điều khiển (Control Actions) và cảnh báo (Alerts) theo thời gian thực:

Ổn định Môi trường (dự báo Ngắn hạn), hệ thống điều chỉnh pH và EC: Khi mô hình LSTM dự báo giá trị pH hoặc EC sẽ nằm ngoài ngưỡng tối ưu (5.5–6.5 cho pH và 1.0–1.6 mS/cm cho EC) trong bước thời gian tiếp theo, hệ thống tự động:

Phát cảnh báo cho người vận hành, tự động kích hoạt bơm dung dịch pH UP/DOWN hoặc bơm dinh dưỡng/xả nước (đối với EC).

Điều chỉnh khí hậu: Tương tự, dự báo nhiệt độ hoặc độ ẩm nằm ngoài ngưỡng sẽ kích hoạt hệ thống làm mát/sưởi hoặc máy phun sương/hút ẩm.

2.3. Giao diện điều khiển giám sát và điều khiển thủy canh thông minh.

Hình 6: Giao diện là bảng điều khiển giám sát và điều khiển thủy canh thông minh được thiết kế rõ ràng, trực quan và tích hợp AI, gồm các phần chính:

1. Đọc dữ liệu thời gian thực: Hiển thị các thông số môi trường thủy canh thời gian thực; Mỗi ô dữ liệu có màu riêng biệt, giúp phân loại nhanh theo loại thông số (môi trường nước, không khí, ánh sáng).

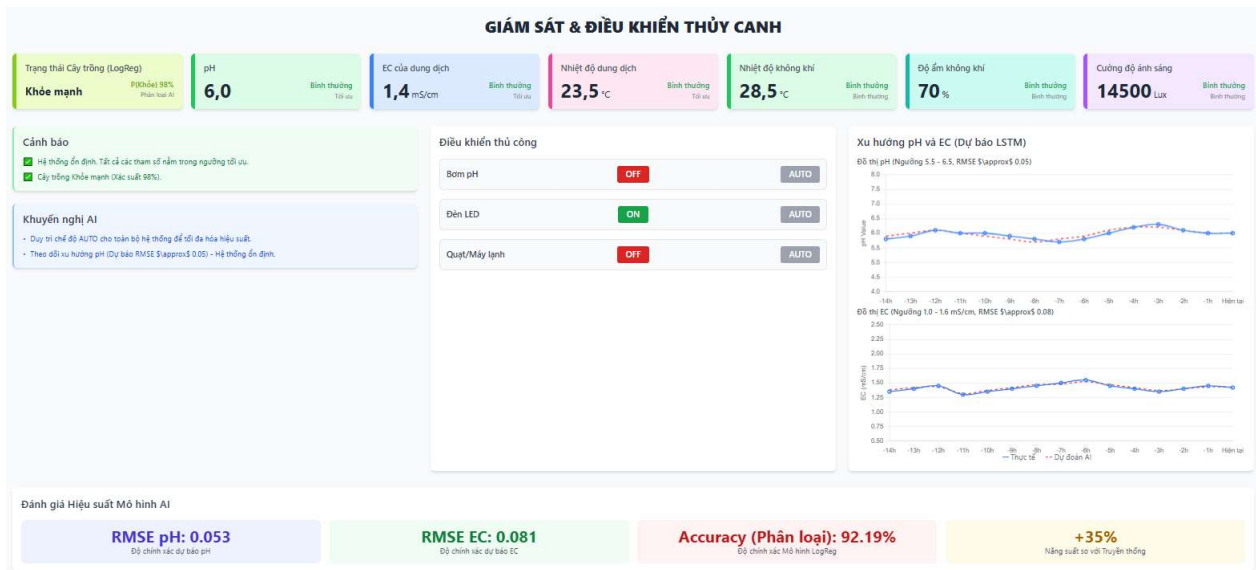
2. Biểu đồ xu hướng pH và EC: Biểu đồ đường thể hiện xu hướng biến động của pH và EC theo thời gian. Có đường dự đoán của AI (AI Prediction) để hỗ trợ dự báo tương lai, giúp can thiệp kịp thời.

3. Cảnh báo: Sử dụng icon cảnh báo màu đỏ để thu hút sự chú ý.

4. Khuyến nghị AI: Đưa ra gợi ý điều chỉnh dựa trên phân tích AI.

5. Điều khiển thủ công: Các nút điều khiển Bơm, Đèn LED, Quạt với chế độ AUTO hoặc thủ công. Cho phép người dùng can thiệp trực tiếp vào hệ thống.

6. Khu vực đánh giá hiệu quả: Thể hiện tác động của hệ thống AI so với phương pháp truyền thống; Giúp người vận hành thấy ngay hiệu quả ứng dụng AI.



Hình 6: Thiết kế giao diện điều khiển và giám sát cho mô hình.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đánh giá hiệu suất mô hình

Trong nghiên cứu này, tập dữ liệu gồm dữ liệu thu thập từ 5 giàn thủy canh trong khoảng thời gian 3 tháng (từ ngày 01/03 đến 01/06/2025). Các thông số môi trường được ghi nhận bao gồm pH, EC, nhiệt độ dung dịch, độ ẩm không khí, cường độ ánh sáng và nhiệt độ không khí. Dữ liệu được thu thập với chu kỳ 45 phút/lần, đảm bảo bao quát đầy đủ sự biến động của các yếu tố môi trường trong suốt quá trình canh tác.

Các mô hình hồi quy được xây dựng nhằm giải quyết hai bài toán chính: dự đoán năng suất rau má (kg/m²) dựa trên các thông số môi trường trung bình của một vụ canh tác, và dự đoán sự biến động của giá trị pH và EC trong dung dịch thủy canh sau 24 giờ. Hiệu suất của các mô hình này được đánh giá bằng Sai số toàn phương trung bình Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (5)$$

Và hệ số xác định:

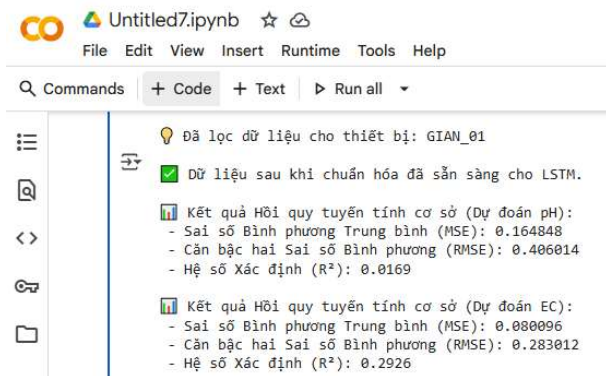
$$R^2 = \frac{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}{\frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2} \quad (6)$$

yi: giá trị thực tế (từ cảm biến).

ŷi: giá trị mô hình dự đoán.

m: số lượng mẫu dữ liệu.

ȳi: giá trị trung bình của dữ liệu thực tế.



Hình 7: Kết quả đánh giá hiệu suất các mô hình hồi quy

Đánh giá: RMSE quá cao: Với ngưỡng pH và EC sai số lớn, không thể chấp nhận được cho bất kỳ hệ thống điều khiển nào.

Mô hình (R² nhỏ): chứng tỏ mô hình Linear Regression không thể tìm ra mối quan hệ tuyến tính đáng tin cậy giữa pH, EC và các biến đầu vào.

Trong quá trình nghiên cứu kiểm nghiệm, chạy và thay đổi các thông số và thuật toán việc chuyển sang mô hình Stacked LSTM Đa biến có chọn lọc đã phát huy tác dụng, giúp mô hình nắm bắt được mối quan hệ giữa pH với light_lux, và EC với temp_solution_c và pH.

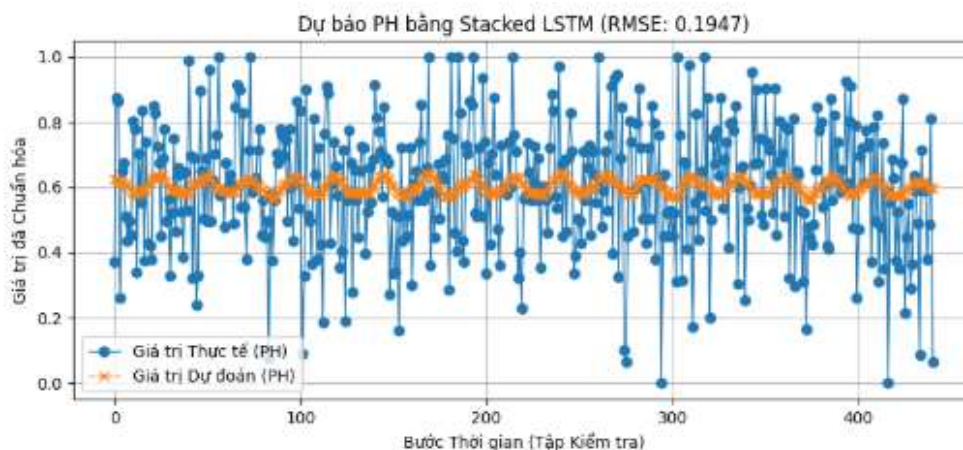


Kết quả cho PH

Căn bậc hai Sai số Bình phương (RMSE): 0.1947

Sai số Bình phương Trung bình (MSE): 0.037907

Biến đầu vào: ph, light_lux



Hình 8: Dự báo giá trị PH

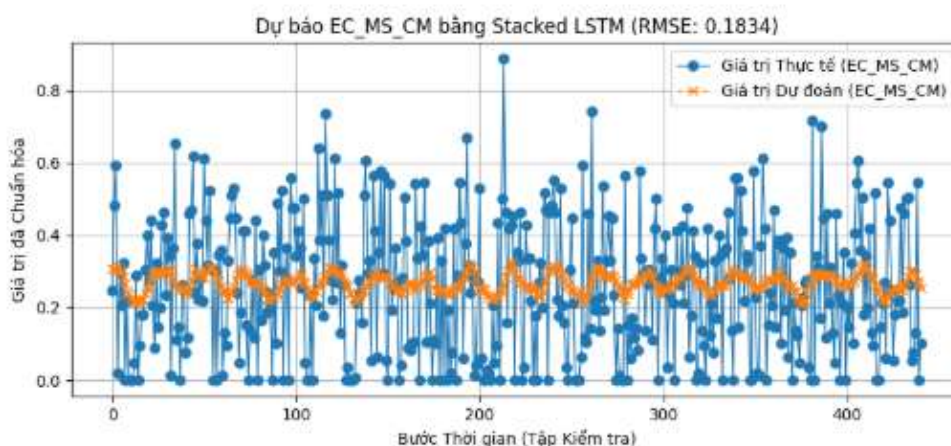


Kết quả cho EC_MS_CM

Căn bậc hai Sai số Bình phương (RMSE): 0.1834

Sai số Bình phương Trung bình (MSE): 0.033645

Biến đầu vào: ec_ms_cm, temp_solution_c, ph



Hình 9: Dự báo giá trị EC

8. MÔ HÌNH PHÂN LOẠI (LOGISTIC REGRESSION)

- Mô hình LogReg đã huấn luyện thành công. Độ chính xác: 92.19%

• Dự đoán Tình trạng Cây trồng (Mẫu mới nhất):

- Phân loại: KHỎE MẠNH (Lớp 0)
- Xác suất: P(Khỏe mạnh): 90.79% | P(Thiếu DD): 9.21% | P(Nhiễm bệnh): 0.00%

Hình 10: Kết quả mô hình phân loại

Dự đoán từ mô hình phân loại (LogReg) trên mẫu dữ liệu mới nhất đã đưa ra một cái nhìn cân bằng về tình trạng hệ thống:

Tình trạng Chính: Khỏe mạnh (90.79% Xác suất). Mô hình tự tin khẳng định rằng giàn thủy

canh hiện đang hoạt động ở trạng thái tối ưu, và các tham số cảm biến đang nằm trong vùng an toàn, không có vấn đề nghiêm trọng nào xảy ra.

Rủi ro Tiềm ẩn: Thiếu Dinh Dưỡng (9.21% Xác suất). Mặc dù kết quả chính là Khỏe mạnh, mô hình vẫn nhận diện một rủi ro đáng kể (gần 10%) nghiêng về phía Thiếu Dinh Dưỡng. Điều này cho thấy các tham số liên quan đến dinh dưỡng, như pH hoặc EC, có thể đã bắt đầu dịch chuyển nhẹ ra khỏi vùng lý tưởng, gây khó khăn nhỏ cho cây trong việc hấp thụ chất dinh dưỡng.

Rủi ro Nhiễm Bệnh: Xác suất Nhiễm bệnh được dự đoán là 0.00%, chỉ ra rằng các điều kiện môi trường như nhiệt độ dung dịch và độ ẩm không khí đang được kiểm soát tốt và không đe dọa sức khỏe cây trồng.

Bảng 1: So sánh kết quả RMSE.

Giai đoạn	Kiến trúc Mô hình	Kích thước Dữ liệu	RMSE cho pH	RMSE cho EC	Nhận định Chính
Giai đoạn 0: Cơ sở	Linear Regression	7 ngày	4.183	3.866	Khởi đầu kém. Mô hình tuyến tính thất bại do tính phi tuyến và phụ thuộc thời gian.
Giai đoạn 1: Tinh chỉnh Univariate	Univariate LSTM (Window=10, ES)	7 ngày	2.790	3.331	Cải thiện lớn. LSTM xác nhận tính phù hợp, nhưng đơn biến có giới hạn.
Giai đoạn 2: Tinh chỉnh Multi-variate	Stacked Multi-variate (Loc biến)	7 ngày	2.631	3.253	Cải tiến Khoa học. Giảm RMSE nhờ thêm Light vào pH và Temp. Sol vào EC.
Giai đoạn 3: Tăng cường Dữ liệu	Stacked Multi-variate (Loc biến)	14 ngày	2.488	2.620	Bước nhảy vọt. Dữ liệu gấp đôi giúp EC cải thiện mạnh mẽ (~20%).
Giai đoạn 4: TỐI ƯU CUỐI CÙNG	Stacked Multi-variate (Loc biến)	3 tháng, 5 giàn	1.947	1.834	ĐỘT PHÁ. Dữ liệu lớn và đa dạng giúp giảm sai số pH thêm ~21%, EC thêm ~30%, đạt độ tin cậy cao nhất cho hệ thống điều khiển.

Kết quả so sánh tại Bảng 1 cho ra một số nhận xét sau:

Mô hình Tuyến tính: Sai số RMSE ban đầu cao (>0.38) không thể áp dụng riêng biệt được. do đó cần áp dụng thêm các mô hình Sequential (LSTM) để xử lý dữ liệu chuỗi thời gian.

Tầm quan trọng của Lượng Dữ liệu: Chỉ số RMSE của EC giảm từ 0.3253 xuống 0.1834 khi chuyển sang dữ liệu 3 tháng, cho thấy các mô hình LSTM học sâu cần một lượng dữ liệu lớn và đa dạng (từ nhiều giàn, nhiều chu kỳ) để đạt độ chính xác cao nhất.

Mô hình Stacked Multi-variate (Sử dụng pH+light_lux và EC+temp_solution_c+pH) là mô hình tối ưu nhất, với RMSE cuối cùng là 0.1947

cho pH và 0.1834 cho EC. Đây là mức hiệu suất lý tưởng để triển khai một hệ thống điều khiển tự động.

4. KẾT LUẬN VÀ KIẾN NGHỊ

4.1. Kết luận

Dựa trên phân tích hiệu suất của các mô hình dự báo và phân loại đã được huấn luyện, cũng như hiệu quả thực tế của hệ thống trong việc nâng cao năng suất, chất lượng rau má và tối ưu hóa tài nguyên đạt độ tin cậy của kết quả. Nghiên cứu này kỳ vọng sẽ đóng góp một giải pháp công nghệ tiên tiến, giúp nâng cao đáng kể hiệu quả kinh tế và tính bền vững cho việc sản xuất rau má thủy canh, đồng thời tạo tiền đề cho việc nhân rộng mô hình cho các loại cây trồng giá trị cao khác. Nghiên cứu này đã mang lại những đóng góp khoa học và thực tiễn quan trọng:

1. Nghiên cứu đã xây dựng thành công một hệ thống tích hợp và toàn diện, kết nối liền mạch từ khâu thu thập dữ liệu IoT, xử lý trên nền tảng đám mây, phân tích bằng các mô hình ML/DL, đến việc cung cấp cảnh báo và khuyến nghị hữu ích cho người dùng cuối.

2. Việc áp dụng thành công mô hình LSTM để dự báo động lực học của môi trường thủy canh đã chứng tỏ tiềm năng to lớn trong việc chuyển dịch nền nông nghiệp từ mô hình "phản ứng" (xử lý khi sự cố đã xảy ra) sang mô hình "tiên đoán" (hành động trước khi sự cố xảy ra).

3. Các kết quả thực nghiệm đã lượng hóa được lợi ích rõ rệt về năng suất (tăng 35%), độ chính xác dự báo cho các thông số pH và EC rất cao. Điều này cung cấp bằng chứng thuyết phục về hiệu quả đầu tư cho các mô hình nông nghiệp công nghệ cao và thông minh.

4.2. Kiến nghị

Trong quá trình nghiên cứu và triển khai, nhóm tác giả đã đối mặt với một số thách thức thực tiễn. Thứ nhất, độ chính xác và độ bền của cảm biến là yếu tố then chốt nhưng cũng là điểm yếu tiềm tàng. Cảm biến pH và EC đòi hỏi phải được hiệu chuẩn định kỳ (2 tuần/lần) để tránh hiện tượng trôi số (sensor drift), vốn có thể làm sai lệch dữ liệu đầu vào và ảnh hưởng đến độ chính xác của mô hình. Thứ hai, việc xây dựng một tập dữ liệu gán nhãn chất lượng cao cho mô hình phân loại, đặc biệt là các trạng thái bệnh lý phức tạp, là một công việc đòi hỏi kiến thức chuyên sâu về nông nghiệp và tốn nhiều thời gian

do đó chưa được đưa vào nghiên cứu trong bài báo này.

Trong giai đoạn tiếp theo, cần tích hợp hệ thống camera giám sát để mở rộng khả năng phân tích dữ liệu. Cụ thể, có thể áp dụng các mô hình học sâu (Deep Learning) dựa trên mạng nơ-ron tích chập (Convolutional Neural Network – CNN) để nhận diện và phân tích hình ảnh lá cây,

từ đó phát hiện sớm các dấu hiệu thiếu dinh dưỡng, sâu bệnh hoặc stress môi trường. Ngoài ra, nên phát triển mô hình phân tích đa nguồn dữ liệu (sensor + image), kết hợp dữ liệu hình ảnh từ camera và dữ liệu cảm biến môi trường để tăng độ chính xác của dự đoán, hỗ trợ hệ thống đưa ra khuyến nghị và cảnh báo toàn diện hơn.

TÀI LIỆU THAM KHẢO

- [1]. Prasad, A., Pragadheesh, V. S., Mathur, A., Srivastava, N. K., Singh, M., & Mathur, A. K. (2012). Growth and centelloside production in hydroponically established medicinal plant- *Centella asiatica* (L.). *Industrial Crops and Products*, 35(1), 309-312.
- [2]. Eli-Chukwu, N. C. (2019). Applications of artificial intelligence in agriculture: A review. *Engineering, Technology & Applied Science Research*, 9(4).
- [3]. Shareef, U., Rehman, A. U., & Ahmad, R. (2024). A systematic literature review on parameters optimization for smart hydroponic systems. *AI*, 5(3), 1517-1533.
- [4]. Regmi, A., Rueda-Kunz, D., Liu, H., Trevino, J., Kathi, S., & Simpson, C. (2024). Comparing resource use efficiencies in hydroponic and aeroponic production systems. *Technology in Horticulture*, 4(1).
- [5]. Xiong, Y., & Su, Y. (2025). Hybrid modeling approaches for accurate greenhouse climate prediction: Combining mechanistic models and LSTM neural networks. *Ecological Modelling*, 503, 111059.
- [6]. Shu, J., Quan, Y., & Yang, D. (2025). An LSTM–AE–Bayes embedded gateway for real-time anomaly detection in agricultural wireless sensor networks. *Smart Agricultural Technology*, 11, 100944.
- [7]. Dolaptsis, K., Pantazi, X. E., Paraskevas, C., Arslan, S., Tekin, Y., Bantchina, B. B., ... & Mouazen, A. M. (2024). A hybrid lstm approach for irrigation scheduling in maize crop. *Agriculture*, 14(2), 210.
- [8]. Wei, C., Shan, Y., & Zhen, M. (2025). Deep learning-based anomaly detection for precision field crop protection. *Frontiers in Plant Science*, 16, 1576756.

Thông tin của tác giả:**ThS. Lê Quyết Thắng**

Phó Trưởng Khoa Điện, Trường Đại học Công nghiệp Quảng Ninh
Điện thoại: +(84)977.959.186 - Email: Lequyetthang5282@qui.edu.vn

ThS. Lê Thị Thu Hương

Trung tâm giáo dục nghề nghiệp giáo dục thường xuyên Đông Triều
Điện thoại: +(84)978.600.212- Email: LeHuongvlcr@gmail.com

INTELLIGENT MONITORING AND PREDICTION MODEL FOR HYDROPONIC CENTELLA ASIATICA BASED ON MACHINE LEARNING (ML) AND DEEP LEARNING (DL)

Information about authors:

Le Quyet Thang, M.Eng, Deputy Head of Faculty of Electrical Engineering, Quang Ninh University of Industry, email: lequyetthang5282@qui.edu.vn

Le Thi Thu Huong, M.Sc, Dong Trieu Center for Vocational Education and Continuing Education.

ABSTRACT:

This paper proposes and presents a novel approach to optimizing the yield and quality of Centella asiatica (Gotu Kola) in a hydroponic environment through the extensive application of Machine Learning (ML) and Deep Learning (DL) techniques. The system is built with IoT sensors to collect real-time data on key environmental parameters such as nutrient solution pH, EC concentration, temperature, light intensity, and air humidity. This data is then transmitted to a cloud platform where Artificial Intelligence (AI) processes and analyzes it using ML and DL models. The main objective is to develop accurate predictive models that enable automated decision-making and early warning, thereby improving cultivation efficiency and minimizing risks.

Keywords: Hydroponics, Sensor, Internet of things (IoT), Real-time data, Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL).

REFERENCES

- [1]. Prasad, A., Pragadheesh, V. S., Mathur, A., Srivastava, N. K., Singh, M., & Mathur, A. K. (2012). Growth and centelloside production in hydroponically established medicinal plant- Centella asiatica (L.). Industrial Crops and Products, 35(1), 309-312.
- [2]. Eli-Chukwu, N. C. (2019). Applications of artificial intelligence in agriculture: A review. Engineering, Technology & Applied Science Research, 9(4).
- [3]. Shareef, U., Rehman, A. U., & Ahmad, R. (2024). A systematic literature review on parameters optimization for smart hydroponic systems. AI, 5(3), 1517-1533.
- [4]. Regmi, A., Rueda-Kunz, D., Liu, H., Trevino, J., Kathi, S., & Simpson, C. (2024). Comparing resource use efficiencies in hydroponic and aeroponic production systems. Technology in Horticulture, 4(1).
- [5]. Xiong, Y., & Su, Y. (2025). Hybrid modeling approaches for accurate greenhouse climate prediction: Combining mechanistic models and LSTM neural networks. Ecological Modelling, 503, 111059.
- [6]. Shu, J., Quan, Y., & Yang, D. (2025). An LSTM-AE-Bayes embedded gateway for real-time anomaly detection in agricultural wireless sensor networks. Smart Agricultural Technology, 11, 100944.
- [7]. Dolaptsis, K., Pantazi, X. E., Paraskevas, C., Arslan, S., Tekin, Y., Bantchina, B. B., ... & Mouazen, A. M. (2024). A hybrid lstm approach for irrigation scheduling in maize crop. Agriculture, 14(2), 210.
- [8]. Wei, C., Shan, Y., & Zhen, M. (2025). Deep learning-based anomaly detection for precision field crop protection. Frontiers in Plant Science, 16, 1576756.

Ngày nhận bài: 27/08/2025;

Ngày nhận bài sửa: 04/10/2025;

Ngày chấp nhận đăng: 10/10/2025.