

ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ
MINH

TRƯỜNG ĐẠI HỌC SÀI GÒN



**NHẬN DIỆN BIỂU CẢM KHUÔN
MẶT TRONG ĐIỀU KIỆN ÁNH
SÁNG YẾU SỬ DỤNG CNN NHE
KẾT HỢP KỸ THUẬT TĂNG CƯỜNG
DỮ LIỆU THÍCH ỨNG**

**ĐỀ CƯƠNG LUẬN VĂN MÔN HỌC NCKH TRONG
CNTT**

NGÀNH: CÔNG NGHỆ THÔNG TIN

Nhóm sinh viên thực hiện:

Họ và tên	MSSV
Văn Tuấn Kiệt	3122410202
Mai Phúc Lâm	3122410207
Nguyễn Đức Duy Lâm	3122410208
Nguyễn Hữu Lộc	3122410213

Giáo viên hướng dẫn: Đỗ Như Tài

TP.HCM, 2025

ĐỀ CƯƠNG LUẬN VĂN

1 Lý do chọn đề tài

Trong thời đại công nghệ số và trí tuệ nhân tạo phát triển mạnh mẽ, các hệ thống nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) ngày càng đóng vai trò quan trọng trong nhiều lĩnh vực như giám sát an ninh, giáo dục thông minh, chăm sóc sức khỏe, và tương tác người - máy. Tuy nhiên, một trong những thách thức lớn mà các hệ thống FER hiện nay gặp phải là điều kiện ánh sáng không ổn định, đặc biệt là ánh sáng yếu – tình huống phổ biến trong môi trường thực tế như ban đêm, nhà thông minh hoặc thiết bị giám sát có cấu hình thấp.

Nhiều nghiên cứu hiện tại đã áp dụng các mô hình học sâu như CNN, GAN hay các phương pháp tiền xử lý hình ảnh phức tạp để cải thiện hiệu suất trong điều kiện ánh sáng yếu. Tuy nhiên, phần lớn các phương pháp này đòi hỏi tài nguyên tính toán cao, khó triển khai trên thiết bị thực tế như camera nhúng, điện thoại cấu hình thấp hoặc hệ thống giám sát nhỏ gọn.

Xuất phát từ thực tiễn đó, nhóm chúng em lựa chọn nghiên cứu và xây dựng một hệ thống FER đơn giản, hiệu quả và phù hợp với thiết bị hạn chế tài nguyên. Đề tài tập trung vào việc kết hợp mô hình CNN nhẹ (MobileNetV3) với kỹ thuật tăng cường dữ liệu thích ứng theo mức độ sáng nhằm nâng cao khả năng nhận diện biểu cảm trong điều kiện ánh sáng yếu. Với hướng tiếp cận này, đề tài kỳ vọng có thể góp phần vào việc phát triển các hệ thống nhận diện thông minh có tính khả thi cao trong ứng dụng thực tiễn.

2 Tổng quan nghiên cứu

2.1 Tình hình nghiên cứu hiện tại

Nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) là một nhánh quan trọng trong lĩnh vực trí tuệ nhân tạo (AI) và thị giác máy tính. Trong những năm gần đây, nhờ sự phát triển của học sâu (Deep Learning), đặc biệt là mạng nơ-ron tích chập (Convolutional Neural Networks - CNN), hiệu suất của các hệ thống FER đã được cải thiện đáng kể. Các mô hình như VGGNet, ResNet, Inception, và gần đây là EfficientNet và Vision Transformer đã đạt độ chính xác cao trên nhiều tập dữ liệu chuẩn như FER-2013, CK+, JAFFE, và AffectNet.

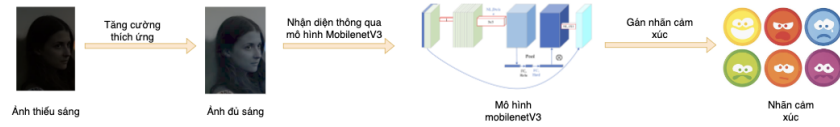
Tuy nhiên, đa số các nghiên cứu tập trung vào điều kiện ánh sáng chuẩn, trong khi điều kiện ánh sáng yếu vẫn còn là một thách thức lớn. Trong môi trường ánh sáng yếu, đặc trưng khuôn mặt bị mất thông tin, độ tương phản thấp, dẫn đến độ chính xác giảm đáng kể. Để giải quyết vấn đề này, một số nghiên cứu gần đây đã đề xuất sử dụng các kỹ thuật như tăng cường dữ liệu ánh sáng yếu (Low-Light Data Augmentation), cải thiện độ sáng bằng GAN (EnlightenGAN, RetinexNet), hoặc sử dụng bộ tăng cường ánh sáng học sâu. Mặc dù các phương pháp này cho kết quả tốt, chúng thường đòi hỏi mô hình phức tạp và tài nguyên tính toán lớn.

Bên cạnh đó, xu hướng nghiên cứu gần đây cũng tập trung vào các mô hình nhẹ (Lightweight CNNs) như MobileNet, ShuffleNet, hoặc SqueezeNet nhằm đáp ứng yêu cầu triển khai thực tế trên thiết bị di động hoặc hệ thống nhúng. Một số công trình kết hợp mô hình nhẹ với chiến lược huấn luyện thông minh hoặc tăng cường dữ liệu thích ứng để cải thiện hiệu quả trong điều kiện môi trường thay đổi, bao gồm ánh sáng yếu.

2.2 Hướng tiếp cận của đề tài

Đề tài tập trung vào việc giải quyết bài toán nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu thông qua hai hướng tiếp cận

chính:



Hình 1: Sơ đồ tổng quan quá trình nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu sử dụng mô hình MobileNetV3 kết hợp tăng cường dữ liệu thích ứng.

- **Sử dụng mô hình CNN nhẹ:** MobileNetV3 được lựa chọn làm kiến trúc mạng nơ-ron nền tảng vì có dung lượng nhỏ, tốc độ suy luận nhanh và phù hợp triển khai trên thiết bị giới hạn tài nguyên. Điều này giúp đảm bảo tính khả thi trong các ứng dụng thực tế như camera an ninh, robot xã hội, hay thiết bị đeo thông minh.
- **Tăng cường dữ liệu thích ứng:** Khác với các phương pháp tăng cường dữ liệu cố định, đề tài đề xuất kỹ thuật tăng cường dựa trên mức độ sáng của từng ảnh đầu vào. Các phép biến đổi như điều chỉnh gamma, contrast stretching sẽ được áp dụng linh hoạt để mô phỏng đa dạng điều kiện ánh sáng yếu, giúp mô hình học được các đặc trưng biểu cảm một cách ổn định hơn.

Hướng tiếp cận này vừa đảm bảo hiệu quả nhận diện trong điều kiện ánh sáng khó khăn, vừa phù hợp với yêu cầu triển khai thực tế. Đề tài không chỉ có ý nghĩa trong việc nâng cao chất lượng hệ thống FER, mà còn mở ra khả năng mở rộng cho các bài toán thị giác máy tính khác trong môi trường bất lợi.

Câu hỏi nghiên cứu

Từ thực trạng và hướng tiếp cận đã trình bày, đề tài tập trung trả lời các câu hỏi nghiên cứu chính sau:

- Việc áp dụng kỹ thuật tăng cường dữ liệu thích ứng theo mức độ sáng

có giúp cải thiện hiệu suất nhận diện biểu cảm khuôn mặt trong điều kiện ánh sáng yếu hay không?

- Mô hình CNN nhẹ (MobileNetV3) khi kết hợp với dữ liệu đã được tăng cường thích ứng có thể đạt độ chính xác trên 70% và đảm bảo tốc độ xử lý đủ nhanh để triển khai thực tế trên thiết bị tài nguyên thấp hay không?

3 Mục đích và nhiệm vụ nghiên cứu

3.1 Mục đích nghiên cứu

Mục đích của đề tài là xây dựng một hệ thống nhận diện biểu cảm khuôn mặt hiệu quả trong điều kiện ánh sáng yếu, đảm bảo tính chính xác và tốc độ xử lý, đồng thời phù hợp để triển khai trên các thiết bị có giới hạn tài nguyên như camera giám sát, thiết bị di động hoặc hệ thống nhúng. Thông qua việc kết hợp mô hình CNN nhẹ (MobileNetV3) với kỹ thuật tăng cường dữ liệu thích ứng theo mức độ sáng, đề tài hướng đến giải quyết các hạn chế về hiệu suất trong môi trường ánh sáng không ổn định.

Đề tài kỳ vọng chứng minh rằng với cách tiếp cận đơn giản nhưng hợp lý, có thể đạt được độ chính xác chấp nhận được ($> 70\%$) trong điều kiện ánh sáng yếu mà không cần sử dụng các mô hình phức tạp hoặc chi phí cao.

3.2 Nhiệm vụ nghiên cứu

Để đạt được mục đích trên, đề tài cần thực hiện các nhiệm vụ chính sau:

- Tìm hiểu tổng quan về các mô hình CNN nhẹ, đặc biệt là MobileNetV3, và kỹ thuật tăng cường dữ liệu liên quan đến ánh sáng yếu.
- Khảo sát và phân tích các phương pháp tăng cường dữ liệu thích ứng hiện có, từ đó đề xuất kỹ thuật phù hợp với bài toán FER.

- Tiền xử lý và xây dựng tập dữ liệu mô phỏng điều kiện ánh sáng yếu dựa trên FER-2013 hoặc các tập dữ liệu FER công khai khác.
- Thiết kế và huấn luyện mô hình MobileNetV3 kết hợp với kỹ thuật tăng cường dữ liệu thích ứng để nhận diện biểu cảm khuôn mặt.
- Đánh giá hiệu suất mô hình qua các chỉ số như Accuracy, Precision, Recall, F1-score và thời gian suy luận trên CPU.
- So sánh kết quả giữa mô hình có và không áp dụng tăng cường thích ứng để chứng minh hiệu quả của phương pháp đề xuất.

4 Đối tượng và phạm vi nghiên cứu

4.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là các kỹ thuật nhận diện biểu cảm khuôn mặt (Facial Expression Recognition - FER) trong điều kiện ánh sáng yếu, với trọng tâm là mô hình học sâu nhẹ (lightweight deep learning) và kỹ thuật tăng cường dữ liệu thích ứng theo mức độ sáng.

Cụ thể, đề tài tập trung vào:

- Mô hình mạng nơ-ron tích chập nhẹ (MobileNetV3) trong bài toán phân loại cảm xúc khuôn mặt.
- Các kỹ thuật tăng cường dữ liệu ảnh thích ứng dựa trên đặc trưng ánh sáng của từng ảnh (gamma correction, contrast adjustment, histogram equalization,...).
- Ứng dụng các kỹ thuật trên trong điều kiện ánh sáng yếu nhằm cải thiện hiệu suất phân loại biểu cảm.

4.2 Phạm vi nghiên cứu

Phạm vi nghiên cứu của đề tài được giới hạn trong các nội dung sau:

- Tập trung xử lý hình ảnh tĩnh (không xử lý video hoặc chuỗi hình ảnh thời gian thực).
- Sử dụng tập dữ liệu công khai FER-2013, với các ảnh được biến đổi để mô phỏng điều kiện ánh sáng yếu.
- Không sử dụng các mô hình phức tạp như GAN, Vision Transformer hoặc các kiến trúc mạng lớn có yêu cầu phần cứng cao.
- Hạn chế ở việc đánh giá hiệu suất mô hình dựa trên độ chính xác, thời gian suy luận và các chỉ số cơ bản (Accuracy, Precision, Recall, F1-score), không mở rộng sang các khía cạnh tâm lý học hay biểu cảm vi mô.
- Triển khai và đánh giá mô hình trên thiết bị CPU mô phỏng điều kiện tài nguyên thấp.

5 Phương pháp nghiên cứu

Để thực hiện đề tài, nhóm áp dụng kết hợp ba nhóm phương pháp chính: phương pháp lý thuyết, phương pháp thực nghiệm và phương pháp so sánh đánh giá. Mỗi phương pháp đóng vai trò hỗ trợ từng giai đoạn cụ thể trong quá trình nghiên cứu, từ khảo sát tài liệu cho đến đánh giá hiệu quả mô hình đề xuất.

5.1 Phương pháp lý thuyết

- Nghiên cứu tổng quan các tài liệu khoa học, công trình nghiên cứu và bài báo liên quan đến nhận diện biểu cảm khuôn mặt (FER), đặc biệt là trong điều kiện ánh sáng yếu.
- Tìm hiểu và phân tích các mô hình CNN nhẹ, trong đó trọng tâm là MobileNetV3 — một mô hình có khả năng triển khai thực tế trên thiết bị tài nguyên thấp.

- Khảo sát các kỹ thuật tăng cường dữ liệu ảnh có liên quan đến ánh sáng, như gamma correction, adaptive histogram equalization, contrast stretching... từ đó đề xuất hướng tăng cường thích ứng theo mức sáng ảnh.

5.2 Phương pháp thực nghiệm

- Sử dụng tập dữ liệu FER-2013 làm dữ liệu huấn luyện và kiểm thử. Tiền xử lý dữ liệu bằng cách mô phỏng các điều kiện ánh sáng yếu thông qua phép biến đổi độ sáng và tương phản.
- Áp dụng kỹ thuật tăng cường dữ liệu thích ứng: sử dụng đặc trưng ánh sáng (ví dụ: độ sáng trung bình, histogram) của từng ảnh để áp dụng các phương pháp tăng cường phù hợp.
- Huấn luyện mô hình MobileNetV3 trên tập dữ liệu đã được tăng cường. Thực hiện fine-tuning mô hình để đạt được hiệu suất tốt nhất.
- Đánh giá mô hình theo các chỉ số chuẩn: Accuracy, Precision, Recall, F1-score và thời gian suy luận trên CPU.

5.3 Phương pháp so sánh đánh giá

- Thực hiện so sánh giữa hai mô hình: mô hình cơ bản (không tăng cường dữ liệu thích ứng) và mô hình có áp dụng kỹ thuật tăng cường thích ứng.
- Phân tích kết quả định lượng (qua chỉ số) và định tính (quan sát vùng biểu cảm dễ nhận diện hơn sau tăng cường) để đánh giá hiệu quả của phương pháp đề xuất.
- Đánh giá mức độ cải thiện về độ chính xác và tốc độ xử lý, từ đó xác định tính khả thi của mô hình trong ứng dụng thực tế.

6 Giả thuyết khoa học

Dựa trên các nghiên cứu trước và định hướng của đề tài, nhóm đưa ra các giả thuyết khoa học như sau:

- **Giả thuyết 1:** Mô hình học sâu nhẹ (MobileNetV3) có thể đạt hiệu suất nhận diện biểu cảm khuôn mặt tốt trong điều kiện ánh sáng yếu nếu được huấn luyện với tập dữ liệu được tăng cường thích ứng theo mức độ sáng.
- **Giả thuyết 2:** Kỹ thuật tăng cường dữ liệu ảnh thích ứng theo đặc trưng ánh sáng đầu vào giúp cải thiện độ chính xác và khả năng khái quát hóa của mô hình so với các phương pháp tăng cường cố định truyền thống.
- **Giả thuyết 3:** Hệ thống kết hợp giữa MobileNetV3 và tăng cường dữ liệu thích ứng có thể duy trì độ chính xác trên 70% và tốc độ suy luận dưới 0.1 giây/ảnh trên CPU, phù hợp với yêu cầu triển khai thực tế trên thiết bị giới hạn tài nguyên.

Các giả thuyết trên sẽ được kiểm chứng thông qua thực nghiệm trên tập dữ liệu FER-2013 đã xử lý, bằng các chỉ số đánh giá như Accuracy, Precision, Recall, F1-score, và thời gian xử lý.

7 Những đóng góp mới của đề tài

Đề tài không chỉ tiếp cận bài toán nhận diện biểu cảm khuôn mặt theo hướng tối ưu tài nguyên mà còn đề xuất cách tiếp cận hiệu quả trong điều kiện ánh sáng yếu – một thách thức phổ biến trong ứng dụng thực tiễn. Những đóng góp mới của đề tài bao gồm:

- **Đề xuất kỹ thuật tăng cường dữ liệu thích ứng theo mức độ sáng:** Khác với các phương pháp tăng cường cố định truyền thống, nhóm đã thiết kế cơ chế tăng cường ảnh linh hoạt dựa trên đặc trưng

ánh sáng đầu vào (trung bình độ sáng, histogram, v.v.). Kỹ thuật này giúp cải thiện khả năng học và tổng quát hóa của mô hình trong các tình huống ánh sáng không đồng đều.

- **Ứng dụng mô hình CNN nhẹ (MobileNetV3) trong điều kiện ánh sáng yếu:** Đề tài chứng minh rằng, với phương pháp xử lý dữ liệu phù hợp, các mô hình CNN nhẹ hoàn toàn có thể đạt hiệu suất tốt mà vẫn đảm bảo tốc độ suy luận nhanh – mở ra khả năng triển khai thực tế trong các hệ thống giám sát, thiết bị IoT, hoặc môi trường có giới hạn tài nguyên.
- **Tạo tập dữ liệu ánh sáng yếu có tính đại diện cao:** Nhóm đã xây dựng tập dữ liệu mô phỏng ánh sáng yếu từ FER-2013 với nhiều cấp độ sáng khác nhau, phục vụ cho huấn luyện và đánh giá mô hình. Đây có thể là tài nguyên tham khảo hữu ích cho các nghiên cứu liên quan trong tương lai.
- **Thực nghiệm so sánh có định hướng rõ ràng:** Đề tài không chỉ triển khai mô hình mà còn thực hiện đánh giá định lượng giữa các phương pháp có và không có tăng cường thích ứng, giúp làm rõ hiệu quả thực sự của hướng tiếp cận đề xuất.

Những đóng góp trên, tuy ở mức ứng dụng, nhưng có ý nghĩa thực tiễn cao và có thể trở thành tiền đề cho các nghiên cứu chuyên sâu hơn trong lĩnh vực nhận diện biểu cảm trong điều kiện ánh sáng không lý tưởng.

8 Hạn chế và rủi ro của đề tài

Mặc dù đề tài đã được định hướng rõ ràng, phù hợp với năng lực thực hiện và có tính khả thi cao, nhóm nghiên cứu vẫn nhận thấy một số hạn chế và rủi ro có thể ảnh hưởng đến kết quả cuối cùng như sau:

- **Dữ liệu ánh sáng yếu chỉ mang tính mô phỏng:** Do hạn chế về nguồn dữ liệu thực tế, nhóm sử dụng ảnh FER-2013 và áp dụng kỹ

thuật giảm sáng để tạo ra tập dữ liệu ánh sáng yếu. Điều này có thể không phản ánh đầy đủ các tình huống ánh sáng yếu thực tế ngoài môi trường.

- **Khả năng tổng quát hóa của mô hình còn hạn chế:** Mô hình được huấn luyện trên dữ liệu tăng cường có thể chưa phản ánh được đầy đủ các biểu cảm phức tạp, đặc biệt là biểu cảm vi mô hoặc biểu cảm có sự che khuất khuôn mặt.
- **Giới hạn về tài nguyên tính toán:** Việc huấn luyện và thử nghiệm mô hình chủ yếu được thực hiện trên CPU hoặc môi trường giả lập, chưa kiểm nghiệm trên các thiết bị thực tế như camera nhúng hoặc hệ thống IoT.
- **Rủi ro về thời gian triển khai:** Việc xử lý tập dữ liệu lớn, tinh chỉnh mô hình và đánh giá kết quả có thể kéo dài hơn dự kiến nếu gặp lỗi kỹ thuật hoặc thiếu tài nguyên máy tính.

Nhóm sẽ cố gắng khắc phục và giảm thiểu các hạn chế trên bằng cách tối ưu quy trình huấn luyện, kiểm thử cẩn thận, và chuẩn bị phương án dự phòng trong quá trình triển khai.

9 Dự kiến kế hoạch nghiên cứu

Kế hoạch thực hiện đề tài được chia thành các giai đoạn rõ ràng, từ nghiên cứu lý thuyết, thu thập dữ liệu, đến thực nghiệm và viết báo cáo. Nhóm dự kiến hoàn thành trong thời gian 8 tuần như sau:

STT	Nội dung công việc	Thời gian
1	Tìm hiểu tổng quan về nhận diện biểu cảm khuôn mặt, điều kiện ánh sáng yếu và các mô hình CNN nhẹ	Tuần 1
2	Khảo sát kỹ thuật tăng cường dữ liệu và đề xuất phương pháp tăng cường thích ứng theo ánh sáng ảnh	Tuần 2
3	Tiền xử lý dữ liệu FER-2013 và xây dựng tập dữ liệu mô phỏng điều kiện ánh sáng yếu	Tuần 3
4	Huấn luyện mô hình MobileNetV3 cơ bản và tinh chỉnh tham số	Tuần 4
5	Tích hợp kỹ thuật tăng cường thích ứng vào pipeline huấn luyện	Tuần 5
6	Đánh giá mô hình: so sánh với mô hình không tăng cường; phân tích kết quả qua các chỉ số	Tuần 6
7	Viết báo cáo, chuẩn bị biểu đồ, bảng số liệu, hình ảnh minh họa	Tuần 7
8	Rà soát, chỉnh sửa báo cáo, hoàn thiện đề cương và chuẩn bị bảo vệ	Tuần 8

Kế hoạch có thể được điều chỉnh linh hoạt tùy theo tiến độ thực tế, tuy nhiên nhóm sẽ đảm bảo hoàn thành đúng hạn và chất lượng tốt nhất.

10 Dự kiến nội dung của luận văn

Luận văn dự kiến được chia thành 5 chương chính như sau:

- **Chương 1 – Giới thiệu đề tài**
 - Bối cảnh, lý do chọn đề tài
 - Mục tiêu, đối tượng, phạm vi nghiên cứu
 - Ý nghĩa khoa học và thực tiễn
- **Chương 2 – Cơ sở lý thuyết và tổng quan nghiên cứu**
 - Các kỹ thuật nhận diện biểu cảm khuôn mặt
 - Kiến trúc CNN nhẹ (đặc biệt là MobileNetV3)
 - Kỹ thuật tăng cường dữ liệu và các công trình liên quan
- **Chương 3 – Phương pháp đề xuất**
 - Chi tiết kỹ thuật tăng cường dữ liệu thích ứng
 - Ứng dụng trên tập dữ liệu FER-2013
 - Tích hợp với mô hình MobileNetV3
- **Chương 4 – Thực nghiệm và đánh giá**
 - Thiết lập huấn luyện và tham số mô hình
 - Đánh giá theo các chỉ số: Accuracy, F1-score, v.v.
 - So sánh mô hình có và không có tăng cường thích ứng
- **Chương 5 – Kết luận và hướng phát triển**
 - Tổng kết nội dung đã thực hiện
 - Đánh giá ưu, nhược điểm
 - Đề xuất hướng phát triển trong tương lai

11 Danh mục tài liệu tham khảo

- [1] S. Kusal *et al.*, “A review on text-based emotion detection—techniques, applications, datasets, and future directions,” *arXiv preprint*, arXiv:2205.03235, 2022.
- [2] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, “URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement,” in *Proc. IEEE CVPR*, 2022, pp. 5901–5910.
- [3] M. Bie *et al.*, “DA-FER: Domain adaptive facial expression recognition,” *Appl. Sci.*, vol. 13, no. 10, p. 6314, 2023, doi: 10.3390/app13106314.
- [4] L. A. Al Hak, W. A. Ali, and S. J. Saba, “Facial expression recognition using data augmentation and transfer learning,” *Ingénierie des Systèmes d’Information*, vol. 29, no. 3, pp. 1219–1225, 2024, doi: 10.18280/isi.290338.
- [5] A. G. Howard *et al.*, “Searching for MobileNetV3,” in *Proc. IEEE ICCV*, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.
- [6] X. Liang, J. Liang, T. Yin, and X. Tang, “A lightweight method for face expression recognition based on improved MobileNetV3,” *IET Image Process.*, vol. 17, no. 8, pp. 2375–2384, 2023, doi: 10.1049/ipe2.12798.
- [7] S. B. R. Prasad and B. S. Chandana, “MobileNetV3: A deep learning technique for human face expressions identification,” *Int. J. Inf. Technol.*, 2023, doi: 10.1007/s41870-023-01380-x.