

CẢI TIẾN PHÂN CỤM MỜ VỚI MULTI-VERSE OPTIMIZER ĐỂ PHÂN LOẠI BỆNH HÔ HẤP: CÚM, COVID-19 VÀ VIÊM PHỔI

Nguyễn Thị Thùy Trang, Đinh Nguyễn Trọng Nghĩa*

Trường Đại học Công Thương Thành phố Hồ Chí Minh

*Email: nghiadnt@huit.edu.vn

Ngày nhận bài: 19/02/2025; Ngày nhận bài sửa: 07/7/2025; Ngày chấp nhận đăng: 24/10/2025

TÓM TẮT

Phân loại chính xác các bệnh hô hấp như cúm, COVID-19 và viêm phổi là yếu tố quan trọng trong chẩn đoán và điều trị, nhưng do sự tương đồng về triệu chứng, việc phân nhóm bệnh nhân gặp nhiều thách thức. Nghiên cứu này đề xuất cải tiến thuật toán phân cụm mờ bằng cách tích hợp Multi-Verse Optimizer (MVO) nhằm tối ưu vị trí tâm cụm, giúp nâng cao chất lượng phân cụm và giảm nhầm lẫn giữa các nhóm bệnh. Việc kết hợp khả năng tìm kiếm tối ưu toàn cục của MVO với phương pháp Fuzzy C-Means (FCM) giúp mô hình phân tách các cụm bệnh chính xác hơn. Kết quả thực nghiệm cho thấy phương pháp đề xuất vượt trội so với FCM truyền thống, cải thiện đáng kể độ chính xác phân cụm và khả năng phân biệt giữa các nhóm bệnh. Phương pháp này có tiềm năng ứng dụng trong hỗ trợ chẩn đoán bệnh hô hấp, góp phần nâng cao hiệu quả trong thực hành y khoa.

Từ khóa: Phân cụm mờ, Multi-Verse Optimizer, Fuzzy C-Means, phân loại bệnh hô hấp, chẩn đoán y khoa.

1. GIỚI THIỆU

Các bệnh lý đường hô hấp như cúm mùa, COVID-19 và viêm phổi hiện đang là thách thức lớn đối với sức khỏe cộng đồng do đặc điểm lây lan nhanh và nguy cơ biến chứng cao. Mặc dù khác biệt về cơ chế gây bệnh, các bệnh này thường biểu hiện các triệu chứng lâm sàng tương đồng như ho, sốt, khó thở, gây khó khăn trong việc chẩn đoán và phân loại chính xác.

Trong bối cảnh đó, các hệ thống hỗ trợ chẩn đoán ứng dụng phương pháp phân tích dữ liệu, đặc biệt là các mô hình học máy kết hợp với thuật toán phân cụm, đang được khai thác nhằm nâng cao độ chính xác và tính khách quan trong đánh giá lâm sàng, qua đó hỗ trợ bác sĩ đưa ra quyết định điều trị hiệu quả hơn.

Trong số các phương pháp phân cụm truyền thống, K-Means [1] và K-Prototype [2] được sử dụng rộng rãi trong lĩnh vực y học nhờ ưu điểm về đơn giản và hiệu quả tính toán. Tuy nhiên, cả hai đều là các phương pháp phân cụm cứng, không có khả năng xử lý sự giao thoa giữa các lớp – một hiện tượng phổ biến trong dữ liệu bệnh nhân – do đó thường không đạt hiệu quả tối ưu trong các bài toán phân tích y sinh học.

Fuzzy C-Means được phát triển như một phương pháp phân cụm mềm (soft clustering), cho phép mỗi điểm dữ liệu có thể đồng thời thuộc về nhiều cụm với các mức độ thành viên khác nhau [3]. Tính linh hoạt này đặc biệt phù hợp với dữ liệu không chắc chắn, chẳng hạn như hồ sơ bệnh án – nơi các biểu hiện bệnh lý có thể chồng lấn hoặc không đặc trưng rõ ràng. Nhờ đó, FCM đã được ứng dụng hiệu quả trong nhiều nghiên cứu y sinh, bao gồm phân cụm bệnh nhân COVID-19 dựa trên các chỉ số xét nghiệm và triệu chứng lâm sàng [4], hoặc kết hợp với mạng nơ-ron tích chập (CNN) để phát hiện tổn thương trên ảnh CT phổi [5].

Một hạn chế đáng kể của FCM là dễ rơi vào các cực tiểu cục bộ, đặc biệt khi dữ liệu có nhiễu hoặc khi tâm cụm được khởi tạo không phù hợp. Điều này làm chậm quá trình hội tụ và có thể ảnh hưởng đến chất lượng phân cụm. Nhằm khắc phục vấn đề này, nhiều biến thể của FCM đã được phát triển, điển hình là mô hình Multiple FCM, trong đó cải thiện khả năng hội tụ và độ ổn định thông qua việc khởi tạo nhiều cấu hình ban đầu và mở rộng không gian tìm kiếm giải pháp [6].

Một hướng tiếp cận tiềm năng đang được nghiên cứu mạnh mẽ là kết hợp FCM với các thuật toán tối ưu hóa metaheuristic nhằm tăng cường khả năng tìm kiếm toàn cục và hạn chế hiện tượng hội tụ sớm tại các cực tiểu cục bộ. Các thuật toán như Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) và Genetic Algorithm (GA) đã được tích hợp vào quá trình tối ưu hóa tâm cụm – yếu tố có ảnh hưởng trực tiếp đến chất lượng phân cụm. Việc kết hợp này đã chứng minh hiệu quả trong việc cải thiện độ chính xác và độ ổn định của mô hình trên nhiều loại dữ liệu khác nhau [7].

Ngoài ra, một số mô hình lai như PSO-FCM [8] hoặc WOA-FCM – kết hợp giữa FCM và Whale Optimization Algorithm (WOA) [9] – cũng đã được đề xuất và ứng dụng trong các bài toán y sinh học, đặc biệt là khi dữ liệu có phân bố không đồng đều hoặc chứa nhiễu. Kết quả thực nghiệm cho thấy các mô hình này mang lại những cải tiến đáng kể về độ chính xác, tốc độ hội tụ và tính ổn định trong các tình huống thực tế.

Trong số các thuật toán tối ưu hóa hiện đại, Multi-Verse Optimizer lấy cảm hứng từ lý thuyết vật lý về vũ trụ song song, được đánh giá cao nhờ khả năng duy trì đa dạng quần thể và điều chỉnh linh hoạt trọng số trong quá trình tiến hóa. Những đặc tính này giúp MVO đạt hiệu quả tốt trong việc tránh bẫy cực tiểu cục bộ và tăng khả năng tìm kiếm lời giải toàn cục [10].

Dựa trên các cơ sở trên, nghiên cứu này đề xuất mô hình phân cụm tích hợp giữa FCM và MVO, trong đó FCM thực hiện phân cụm bệnh nhân dựa trên triệu chứng và chỉ số lâm sàng, còn MVO đóng vai trò tối ưu hóa vị trí khởi tạo tâm cụm nhằm nâng cao hiệu quả phân nhóm. Sự kết hợp này hướng đến cải thiện chất lượng phân cụm, giảm thiểu nguy cơ hội tụ sớm và mở rộng tính ứng dụng trong phân tích dữ liệu y tế có đặc điểm phức tạp, không đồng nhất và có tính mờ cao.

Nghiên cứu này đề xuất mô hình phân cụm tích hợp giữa Fuzzy C-Means và Multi-Verse Optimizer nhằm nâng cao độ chính xác và khả năng hội tụ trong xử lý dữ liệu y tế. Trên cơ sở đó, nhóm tác giả xây dựng hệ thống hỗ trợ phân loại các bệnh hô hấp có triệu chứng lâm sàng tương đồng như cúm mùa, COVID-19 và viêm phổi. Thực nghiệm trên tập dữ liệu gồm 600 bệnh nhân cho thấy mô hình MVO-FCM đạt hiệu suất vượt trội so với FCM truyền thống, đồng thời khẳng định tiềm năng ứng dụng rộng rãi trong nhiều bài toán phân cụm y sinh với dữ liệu phức tạp, không đồng nhất và mang tính mờ cao.

Bài báo được bố cục như sau: Phần 1 giới thiệu bối cảnh nghiên cứu, nhấn mạnh thách thức từ các bệnh lý đường hô hấp và sự cần thiết của các phương pháp phân cụm trong hỗ trợ chẩn đoán y tế. Phần 2 trình bày cơ sở lý thuyết về phân cụm mờ, bao gồm khái niệm, ma trận thành viên mờ, hệ số làm mờ và công thức tính tâm cụm. Phần 3 mô tả chi tiết mô hình MVO-FCM được đề xuất, cơ chế hoạt động và hàm đánh giá chất lượng cụm. Phần 4 trình bày và thảo luận kết quả thực nghiệm, đồng thời so sánh với phương pháp FCM truyền thống để đánh giá hiệu quả mô hình. Cuối cùng, Phần 5 đưa ra kết luận, tổng hợp các đóng góp chính và định hướng nghiên cứu trong tương lai.

2. FUZZY CLUSTERING

2.1. Giới thiệu

Fuzzy Clustering (phân cụm mờ) là một kỹ thuật phân cụm trong đó mỗi điểm dữ liệu có thể đồng thời thuộc về nhiều cụm với các mức độ thành viên khác nhau, thay vì chỉ được gán duy nhất vào một cụm như trong các phương pháp phân cụm cứng truyền thống. Phương pháp này đặc biệt phù hợp với các bài toán có dữ liệu không rõ ràng hoặc tồn tại sự chồng chéo giữa các cụm [3].

Fuzzy Logic được ứng dụng rộng rãi trong nhiều lĩnh vực như y tế, tài chính, phân tích thị trường và xử lý ảnh, nhờ khả năng mô hình hóa và xử lý thông tin không chắc chắn. Trong phân tích thị trường, Fuzzy Logic cho phép phân loại khách hàng vào nhiều phân khúc đồng thời, đồng thời định lượng mức độ thuộc về của từng đối tượng đối với các nhóm thị trường khác nhau. Trong lĩnh vực y tế, Fuzzy Logic hỗ trợ chẩn đoán bệnh bằng cách đánh giá mức độ liên kết giữa hồ sơ bệnh nhân và các nhóm bệnh, đặc biệt hiệu quả khi một triệu chứng có thể liên quan đến nhiều loại bệnh khác nhau. Cơ chế suy luận theo mức độ này góp phần giảm thiểu tính mơ hồ trong quá trình chẩn đoán, từ đó nâng cao độ chính xác và hỗ trợ các quyết định lâm sàng có căn cứ hơn.

2.2. Các thành phần trong Fuzzy Clustering

a. Ma trận thành viên mờ (Fuzzy Membership Matrix)

Ma trận thành viên mờ U là một ma trận kích thước $n \times C$, trong đó n là số điểm dữ liệu và C là số cụm. Mỗi phần tử u_{ik} trong ma trận biểu diễn độ thuộc về của điểm dữ liệu x_i đối với cụm C_k . Các giá trị này nằm trong khoảng $[0,1]$, phản ánh mức độ dữ liệu thuộc về từng cụm một cách linh hoạt thay vì gán cứng vào một cụm duy nhất như trong phân cụm truyền thống [3].

Mỗi điểm dữ liệu x_i có một tập hợp các giá trị u_{ik} (mỗi k đại diện cho một cụm) biểu thị mức độ "thuộc về" của điểm đó đối với các cụm. Các giá trị này thỏa mãn điều kiện được biểu diễn bởi công thức (1):

$$\sum_{k=1}^C u_{ik} = 1 \quad (\forall i = 1, 2, \dots, n) \quad (1)$$

Ví dụ:

Giả sử có một tập hợp bệnh nhân $X = \{BN_1, BN_2, BN_3\}$ và ba cụm mờ G_1 (Cúm), G_2 (COVID-19), G_3 (Viêm phổi). Ma trận U thể hiện mức độ "thuộc về" của mỗi bệnh nhân đối với từng cụm bệnh như sau:

$$U = \begin{pmatrix} 0,2 & 0,6 & 0,2 \\ 0,7 & 0,2 & 0,1 \\ 0,1 & 0,2 & 0,7 \end{pmatrix}$$

BN_1 có mức độ thuộc về nhóm Cúm là 0,2; COVID-19 là 0,6 và Viêm phổi là 0,2. Các giá trị này cho thấy bệnh nhân có khả năng cao mắc COVID-19 nhưng vẫn có một phần triệu chứng liên quan đến Cúm và Viêm phổi. Điều này giúp bác sĩ không chỉ đưa ra chẩn đoán sơ bộ mà còn đánh giá khả năng bệnh nhân có thể thuộc về nhiều nhóm bệnh khác nhau, từ đó hỗ trợ quyết định lâm sàng chính xác hơn.

b. Hệ số làm mờ (fuzzification parameter: m)

Hệ số m là một giá trị điều khiển độ mờ trong phân cụm. Khi $m=1$, phân cụm trở thành phân cụm cứng, tức là mỗi đối tượng chỉ thuộc về một cụm duy nhất. Khi $m>1$, sự phân cụm trở nên mờ hơn, tức là mỗi đối tượng có thể thuộc về nhiều cụm với mức độ khác nhau [3].

c. Trung tâm của cụm (centroids)

Trong phân cụm mờ, mỗi cụm có một trung tâm, tương tự như phân cụm truyền thống. Tuy nhiên, điểm khác biệt là một đối tượng có thể thuộc về nhiều cụm với mức độ khác nhau. Trung tâm cụm được tính dựa trên trọng số mờ của các đối tượng trong cụm.

Giả sử x_i là các đối tượng và u_{ij} là trọng số mờ của đối tượng x_i đối với cụm j , trung tâm cụm C_j được tính bằng công thức (2):

$$C_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (2)$$

Trong đó:

- u_{ij} : Mức độ thuộc về của đối tượng x_i trong cụm j . Đây là trọng số mờ, nằm trong khoảng $[0,1]$
- u_{ij}^m : Trọng số mờ này được nâng lên lũy thừa m , với $m>1$ thường được sử dụng (thường là $m=2$) để tăng cường ảnh hưởng của các đối tượng có mức độ thuộc về cao.

3. PHƯƠNG PHÁP ĐỀ XUẤT

Nghiên cứu đề xuất một phương pháp tích hợp giữa thuật toán FCM và MVO nhằm cải thiện hiệu quả phân cụm bệnh nhân mắc các bệnh hô hấp bao gồm cúm, COVID-19 và viêm phổi, dựa trên dữ liệu triệu chứng lâm sàng và kết quả xét nghiệm.

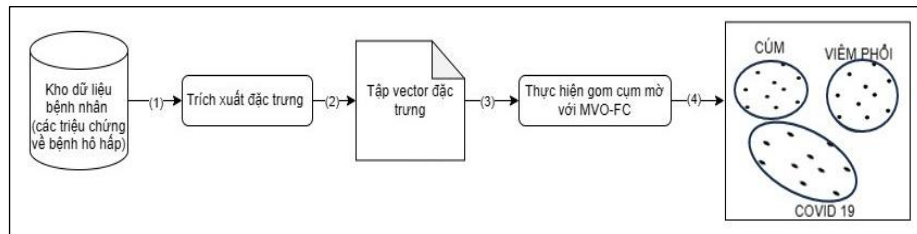
Việc phân loại bệnh nhân dựa trên triệu chứng thường gặp nhiều khó khăn do sự chồng lấn biểu hiện giữa các nhóm bệnh. Ngoài ra, dữ liệu y tế thường chứa nhiễu, phân bố không đồng đều và mang tính mờ cao, ảnh hưởng đáng kể đến độ chính xác của các phương pháp phân cụm truyền thống. Trong bối cảnh này, FCM đóng vai trò xử lý dữ liệu mờ bằng cách cho phép một đối tượng có thể đồng thời

thuộc về nhiều cụm với các mức độ thành viên khác nhau. Điều này giúp phản ánh đúng hơn thực tế lâm sàng, nơi mà các biểu hiện bệnh lý có thể giao thoa và không hoàn toàn đặc trưng cho một nhóm bệnh duy nhất.

Tuy nhiên, hiệu suất của FCM phụ thuộc mạnh vào việc khởi tạo tâm cụm ban đầu, và thuật toán có xu hướng hội tụ tại cực tiểu cục bộ thay vì cực tiểu toàn cục. Nhược điểm này dẫn đến kết quả phân cụm thiếu ổn định và độ chính xác thấp trong các trường hợp dữ liệu phức tạp. Để khắc phục vấn đề trên, MVO được tích hợp nhằm tối ưu hóa quá trình khởi tạo tâm cụm, thông qua việc mô phỏng cơ chế hấp dẫn giữa các vũ trụ trong không gian tìm kiếm lời giải. MVO cho phép duy trì sự đa dạng quần thể và mở rộng phạm vi tìm kiếm, từ đó giảm thiểu nguy cơ rơi vào cực tiểu cục bộ và tăng cường khả năng hội tụ đến nghiệm tối ưu toàn cục.

Phương pháp MVO-FC tận dụng ưu điểm của FCM trong xử lý dữ liệu mờ và khả năng tối ưu toàn cục của MVO, nhằm nâng cao hiệu quả phân cụm bệnh nhân, tăng độ chính xác trong phân loại và hỗ trợ ra quyết định lâm sàng chính xác hơn trong môi trường dữ liệu không chắc chắn.

Mô hình hệ thống phân nhóm bệnh hô hấp dựa trên MVO-FC được minh họa tại Hình 1.



Hình 1. Mô hình hệ thống phân nhóm bệnh hô hấp dựa trên MVO-FC

Quy trình đề xuất gồm bốn bước chính, bắt đầu từ việc thu thập dữ liệu triệu chứng cho đến giai đoạn phân cụm và phân loại bệnh nhân. Cụ thể:

(1) Trích xuất đặc trưng:

Dữ liệu triệu chứng được thu thập từ kho bệnh án và trải qua bước xử lý sơ bộ nhằm loại bỏ nhiễu, chuẩn hóa định dạng và đảm bảo tính toàn vẹn dữ liệu. Các đặc trưng lâm sàng quan trọng được trích xuất bao gồm: nhiệt độ cơ thể (biểu hiện sốt), mức độ ho, khó thở, đau họng, nồng độ C-reactive protein (CRP, đơn vị mg/L), số lượng bạch cầu (tế bào/ μ L) và độ bão hòa oxy trong máu SpO₂ (%). Các đặc trưng này được lựa chọn dựa trên tần suất xuất hiện và tính phân biệt giữa các bệnh hô hấp phổ biến.

(2) Mã hóa vector đặc trưng:

Sau khi trích xuất, các đặc trưng được chuyển đổi thành dạng vector đặc trưng số hóa, đại diện cho hồ sơ triệu chứng của từng bệnh nhân. Tập hợp các vector này tạo thành ma trận đặc trưng đầu vào cho thuật toán phân cụm.

(3) Phân cụm bằng thuật toán MVO-FC:

Thuật toán phân cụm MVO-FC được áp dụng nhằm tối ưu hóa quá trình phân nhóm bệnh nhân. Trong đó, MVO đảm nhiệm vai trò tìm kiếm vị trí khởi tạo tối ưu cho các tâm cụm, giúp FCM đạt hiệu quả phân cụm cao hơn cả về độ chính xác lẫn tốc độ hội tụ. Việc kết hợp này góp phần cải thiện khả năng phân biệt giữa các nhóm bệnh có triệu chứng lâm sàng tương đồng.

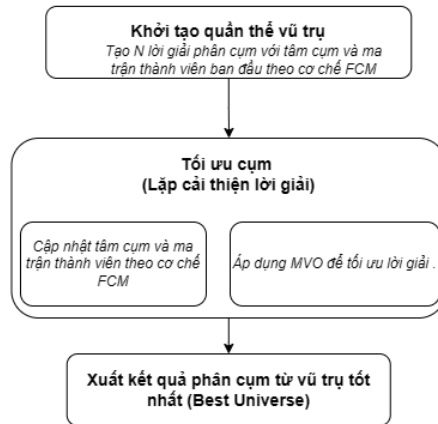
(4) Xác định nhóm bệnh:

Dựa trên kết quả phân cụm thu được, mỗi bệnh nhân được gán nhãn vào một trong ba nhóm bệnh chính gồm: cúm mùa, COVID-19 và viêm phổi. Việc gán nhãn được thực hiện thông qua phân tích trọng số thành viên (membership degree) trong từng cụm, kết hợp với thống kê mô tả và tham chiếu từ tập dữ liệu có nhãn kiểm chứng.

3.1. Cơ chế hoạt động của Multi-Verse Optimizer Fuzzy Clustering

Mô hình MVO-FC kế thừa các thành phần cốt lõi từ thuật toán FCM, bao gồm: ma trận thành viên mờ, quá trình cập nhật tâm cụm và hệ số fuzzification m . Tuy nhiên, điểm khác biệt chính nằm ở việc sử dụng thuật toán MVO để tối ưu hóa vị trí khởi tạo và cập nhật tâm cụm. Cơ chế tích hợp này giúp

cải thiện khả năng hội tụ, tăng độ tách biệt giữa các cụm và giảm thiểu ảnh hưởng của nhiễu, từ đó nâng cao độ chính xác trong phân loại.



Hình 2. Cơ chế hoạt động của MVO-FC

Cơ chế hoạt động tổng thể của MVO-FC được minh họa trong Hình 2, thể hiện luồng xử lý chính: bắt đầu từ việc khởi tạo quần thể vũ trụ (tạo lời giải phân cụm ban đầu), sau đó tối ưu cụm thông qua MVO để cải thiện lời giải, và cuối cùng xuất ra kết quả phân cụm từ vũ trụ tốt nhất.

Như được minh họa trong Hình 2, cơ chế hoạt động của MVO-FC bao gồm ba bước chính:

(1) Tính toán ma trận thành viên:

Thuật toán FCM được triển khai để tính toán mức độ thành viên mờ (membership degree) của từng đối tượng dữ liệu đối với mỗi cụm. Mỗi bệnh nhân được gán một giá trị thành viên cho từng cụm bệnh, phản ánh mức độ liên quan theo ngữ cảnh lâm sàng.

(2) Tối ưu hóa tâm cụm bằng MVO:

MVO được áp dụng để tối ưu hóa vị trí các tâm cụm trong không gian đặc trưng. Quá trình này mô phỏng cơ chế chuyển động giữa các vũ trụ dựa trên mật độ vật chất và hằng số hấp dẫn, cho phép mở rộng không gian tìm kiếm và tránh hiện tượng hội tụ sớm. Việc tối ưu tâm cụm nhằm nâng cao độ tách biệt giữa các nhóm bệnh và tăng khả năng phân biệt trong điều kiện dữ liệu có nhiễu hoặc phân bố không đồng đều.

(3) Kết quả phân cụm:

Kết quả của thuật toán là tập hợp các tâm cụm tối ưu cùng với ma trận thành viên mờ, thể hiện mức độ liên kết của từng bệnh nhân với mỗi nhóm bệnh. Những thông tin này được sử dụng để xác định nhãn phân loại cho từng đối tượng, từ đó hỗ trợ quá trình ra quyết định trong chẩn đoán lâm sàng.

3.2. Đánh giá chất lượng phân cụm trong thuật toán MVO-FC

Trong mô hình MVO-FC, hàm đánh giá (fitness function) được sử dụng nhằm định lượng chất lượng của quá trình phân cụm mờ, qua đó hướng dẫn cơ chế tối ưu hóa của thuật toán MVO trong việc tìm kiếm các tâm cụm tối ưu. Việc lựa chọn hàm đánh giá phù hợp có vai trò then chốt trong việc đảm bảo hiệu quả phân cụm, đặc biệt trong bối cảnh dữ liệu y sinh học thường có tính mờ và phân bố không đồng đều.

Hàm đánh giá trong MVO-FC được xây dựng dựa trên ba tiêu chí chính:

(1) Độ tách biệt giữa các cụm (cluster separation): Phản ánh khoảng cách giữa các tâm cụm, càng lớn thì khả năng phân biệt giữa các nhóm bệnh càng cao.

(2) Mức độ mờ (fuzziness) của điểm dữ liệu: Đo lường mức độ không chắc chắn trong việc gán điểm dữ liệu vào các cụm, từ đó phản ánh tính linh hoạt và khả năng mô hình hóa của thuật toán trong môi trường dữ liệu không rõ ràng.

(3) Độ đồng nhất nội tại của cụm (intra-cluster compactness): Phản ánh mức độ gần nhau giữa các điểm dữ liệu trong cùng một cụm, càng nhỏ thì cụm càng chặt và phân cụm càng hiệu quả.

Hàm mục tiêu tổng quát, ký hiệu là $Fc(P)$, được thiết kế nhằm cân bằng giữa khả năng phân tách cụm và tính đồng nhất nội tại. Trong đó, P đại diện cho một cấu hình cụm cụ thể (tập các tâm cụm được mã hóa trong một "vũ trụ" theo mô hình MVO). Giá trị của $Fc(P)$ càng nhỏ thì chất lượng phân cụm càng cao, do biểu thị sự phân tách rõ ràng giữa các cụm và mức độ liên kết chặt chẽ trong từng cụm.

Hàm đánh giá này đóng vai trò quan trọng trong quá trình tiến hóa của MVO, giúp điều hướng quá trình tìm kiếm đến các vùng giải pháp có cấu trúc cụm hợp lý và phản ánh đúng đặc điểm của dữ liệu bệnh nhân. Việc sử dụng hàm đánh giá này cũng góp phần tăng tính hợp lệ (validity) của kết quả phân cụm trong bối cảnh dữ liệu thực tế.

$$Fc(P) = \frac{\text{Tách biệt giữa các cụm}}{\text{Sự đồng nhất trong mỗi cụm}} \quad (3)$$

$$Fc(P) = \frac{\sum_{k=1}^C \sum_{i=1}^N u_{ik}^m \|x_i - c_k\|^2}{\sum_{k=1}^C \sum_{i=1}^N u_{ik}^m}$$

Trong đó:

- N là số lượng điểm dữ liệu.
- C là số lượng cụm.
- u_{ik} là độ "thuộc về" của điểm dữ liệu x_i đối với cụm k .
- m là hệ số mờ.
- x_i là vector dữ liệu thứ i .
- c_k là trung tâm của cụm k .
- $\|x_i - c_k\|^2$ là khoảng cách bình phương giữa điểm dữ liệu x_i và trung tâm cụm c_k .

3.3. Thuật toán MVO-FC

Việc tích hợp thuật toán MVO với FCM giúp cải thiện vị trí tâm cụm, giảm sự chồng lấn giữa các nhóm và nâng cao độ chính xác trong phân loại. Dưới đây là mô tả chi tiết về thuật toán MVO-FC.

Thuật toán MVO-FC sử dụng các tham số đầu vào như sau:

- C : Số lượng cụm trong dữ liệu.
- m : Hệ số làm mờ kiểm soát mức độ phân tách giữa các cụm, thường được đặt là 2.
- Tập hợp các giải pháp ngẫu nhiên $U = \{U_1, U_2, \dots, U_N\}$
- Các tham số MVO $\omega(\omega_s, \omega_p, \omega_m, \omega_t)$
- + ω_s (Search Boundaries): Giới hạn trên và dưới của từng biến trong không gian tìm kiếm.
- + ω_p (Population Size – N): Số lượng vũ trụ (giải pháp) trong mỗi vòng lặp.
- + ω_m (Max Iterations): Số vòng lặp tối đa của thuật toán.
- + ω_t (Threshold – Ngưỡng xác suất Wormhole): Xác suất ngưỡng để quyết định có áp dụng cơ chế Wormhole hay không.

Những tham số này giúp MVO-FC tối ưu hóa quá trình phân cụm mờ, đảm bảo tìm ra lời giải tối ưu nhất trong không gian tìm kiếm.

Thuật toán MVO-FC (Multi-Verse Optimizer Fuzzy Clustering)

Bước 1: Khởi tạo quần thể vũ trụ

- Tạo quần thể gồm N vũ trụ, ký hiệu tập hợp các ma trận thành viên ngẫu nhiên là: $U = \{U_1, U_2, \dots, U_n\}$.
- Đối với mỗi vũ trụ $Universe_i$:
 - + Xác định tâm cụm.
 - + Tính giá trị hàm mục tiêu (fitness function).
- Xác định Best_Universe – vũ trụ có giá trị hàm mục tiêu tốt nhất.

Bước 2: Lặp lại cho đến khi đạt số vòng lặp tối đa hoặc hội tụ

- a. Cập nhật từng vũ trụ $Universe_i$ trong quần thể:

- Cập nhật ma trận thành viên.
 - Cập nhật vị trí tâm cụm.
 - Tính giá trị hàm thích nghi.
 - Nếu giá trị fitness của $Universe_i$ tốt hơn $Best_Universe$, cập nhật $Best_Universe$.
 - b. Cập nhật vị trí vũ trụ mới bằng cơ chế MVO:
 - Đối với mỗi $Universe_i$ (ngoại trừ $Best_Universe$):
 - + White Hole Mechanism: Chọn ngẫu nhiên một vũ trụ $Universe_i$ để trao đổi một phần thông tin với $Universe_i$.
 - + Wormhole Mechanism: Nếu xác suất $P < \omega_i$:
 - Điều chỉnh vị trí ngẫu nhiên của $Universe_i$ bằng cách dịch chuyển tới một vị trí mới gần $Best_Universe$, tạo cơ hội khám phá không gian tìm kiếm mới.
 - Cơ chế này mô phỏng lỗ sâu (wormhole) trong vũ trụ, giúp $Universe_i$ thoát khỏi các cực tiểu cục bộ, tăng khả năng tìm kiếm các nghiệm tối ưu toàn cục.
- Bước 3: Xuất kết quả
- Xuất vị trí cụm tối ưu từ $Best_Universe$.
 - Gán nhãn phân cụm cho dữ liệu dựa trên ma trận thành viên $U_{Best_Universe}$.

4. KẾT QUẢ VÀ THẢO LUẬN

4.1. Dữ liệu thực nghiệm

Các thí nghiệm được tiến hành trên một máy tính cá nhân có cấu hình gồm bộ vi xử lý Intel Core i3-7100U @ 2.40GHz, bộ nhớ RAM 8.0 GB và hệ điều hành Windows 10 Pro 64-bit. Quá trình thực nghiệm được triển khai bằng Python 3.11, sử dụng các thư viện khoa học dữ liệu như NumPy và Pandas để xử lý dữ liệu, Scikit-learn để tiền xử lý và đánh giá mô hình.

Để đánh giá hiệu quả của thuật toán MVO-FC trong việc phân loại ba loại bệnh hô hấp cấp tính phổ biến gồm COVID-19, cúm mùa và viêm phổi, bộ dữ liệu mô phỏng đã được xây dựng và sử dụng trong các thí nghiệm. Việc lựa chọn dữ liệu mô phỏng thay cho dữ liệu bệnh nhân thực tế xuất phát từ các hạn chế trong việc tiếp cận dữ liệu bệnh nhân thực tế cùng các yêu cầu nghiêm ngặt về bảo mật y tế.

Tập dữ liệu gồm tổng cộng 600 mẫu bệnh nhân giả định, được phân bổ đều cho ba nhóm bệnh, mỗi nhóm gồm 200 mẫu. Việc phân bổ đồng đều giúp đảm bảo tính cân bằng giữa các lớp, từ đó giảm sai lệch trong quá trình huấn luyện và đánh giá mô hình phân cụm. Mỗi mẫu được mô tả bởi bảy đặc trưng lâm sàng phổ biến trong chẩn đoán bệnh hô hấp: nhiệt độ cơ thể, mức độ ho (thang 0–10), tình trạng khó thở (có/không), tình trạng đau họng (có/không), nồng độ CRP (mg/L), số lượng bạch cầu (WBC, tế bào/ μ L) và độ bão hòa oxy trong máu SpO_2 (%).

Các đặc trưng trên được sinh bằng phương pháp lấy mẫu ngẫu nhiên từ các phân phối xác suất được thiết kế phù hợp với bản chất sinh lý học và đặc điểm thống kê thực tế của từng loại bệnh. Cách tiếp cận này cho phép mô phỏng sát với đặc điểm biểu hiện lâm sàng, đồng thời tạo ra không gian đặc trưng có tính chồng lấn và mờ – những yếu tố quan trọng để đánh giá năng lực của thuật toán phân cụm mờ.

Cụ thể, các đặc trưng liên tục như nhiệt độ, CRP, WBC và SpO_2 được mô phỏng từ phân phối chuẩn (Gaussian distribution) [11], với các tham số trung bình và độ lệch chuẩn được điều chỉnh riêng biệt cho từng nhóm bệnh. Điều này giúp tái hiện sự khác biệt về mặt sinh lý giữa các loại bệnh. Ví dụ, bệnh nhân viêm phổi có xu hướng có CRP cao và SpO_2 thấp hơn so với bệnh nhân cúm mùa, đặc điểm này được phản ánh trực tiếp qua các thông số phân phối tương ứng.

Mức độ ho, do là đặc trưng rời rạc và mang tính chủ quan cao, được sinh ngẫu nhiên từ phân phối rời rạc đồng đều (discrete uniform distribution) trong khoảng 0–10 [11]. Phương pháp này bảo đảm sự đa dạng của dữ liệu và tránh hiện tượng thiên lệch do giả định giá trị trung tâm.

Các biến nhị phân như khó thở và đau họng được mô hình hóa thông qua phân phối Bernoulli [11], với xác suất xuất hiện được điều chỉnh theo từng nhóm bệnh. Ví dụ, triệu chứng khó thở được gán xác suất cao hơn cho nhóm viêm phổi so với nhóm cúm mùa, phản ánh đặc điểm lâm sàng thường thấy.

Để tăng mức độ mờ trong không gian đặc trưng, khoảng 10% mẫu dữ liệu được thiết kế dưới dạng “lai ghép” – nghĩa là các đặc trưng của một mẫu có thể đến từ nhiều nhóm bệnh khác nhau, như nhiệt độ từ nhóm cúm nhưng chỉ số CRP từ nhóm COVID-19. Thiết kế này phù hợp với nguyên lý hoạt động của phân cụm mờ, nơi mỗi đối tượng dữ liệu có thể đồng thời thuộc về nhiều cụm với các mức độ khác nhau.

Toàn bộ quá trình sinh dữ liệu được thực hiện với seed ngẫu nhiên cố định nhằm bảo đảm tính nhất quán và khả năng tái lập trong các lần thực nghiệm. Điều này góp phần nâng cao tính khách quan và minh bạch khi đánh giá hiệu quả của mô hình phân cụm.

Mặc dù dữ liệu được xây dựng dưới dạng mô phỏng, các thông số và đặc trưng sử dụng trong quá trình thiết kế đều được tham chiếu từ các nguồn y học đáng tin cậy. Cụ thể, các nghiên cứu lâm sàng của Chen et al. về bệnh nhân COVID-19 [12], hướng dẫn điều trị cúm mùa của IDSA (2018) [13], và hướng dẫn xử trí viêm phổi từ ATS/IDSA (2019) [14], cùng với các phân tích của Wang et al. và Liu et al. [15, 16], đã được sử dụng để xác lập các tham số và mô hình hóa đặc điểm lâm sàng một cách phù hợp.

Bảng 1 trình bày một phần dữ liệu mô phỏng được sử dụng trong các thí nghiệm, với mỗi mẫu được gán nhãn bệnh tương ứng để hỗ trợ quá trình đánh giá hiệu quả phân cụm.

Bảng 1. Tập dữ liệu bệnh nhân hô hấp

Fever	Cough (0-10)	Difficulty Breathing (0/1)	Sore Throat (0/1)	CRP (mg/L)	WBC	SpO2	Disease
38,2	5	1	0	36,1	6252,5	87,6	Covid-19
38,1	7	1	1	16,1	5676,3	95,1	Flu
37,7	7	1	1	10,1	6189,1	96,4	Flu
37,7	8	1	0	30	5132,5	92,4	Covid-19
37,8	5	1	1	49,2	6173,5	88,9	Covid-19
38,2	9	1	1	69,6	16709,4	96,2	Pneumonia
37,6	7	0	1	36	5791,1	92,8	Covid-19
37,7	3	1	0	11,4	5405	95,2	Flu

4.2. Kết quả

Để đánh giá rõ hơn hiệu quả của việc tích hợp thuật toán MVO vào quá trình tối ưu hóa vị trí tâm cụm, nhóm nghiên cứu đã tiến hành so sánh chất lượng phân cụm giữa FCM truyền thống và MVO-FC thông qua một số chỉ số đánh giá khách quan. Kết quả được trình bày trong Bảng 2, bao gồm các thước đo: Fuzzy Silhouette, Accuracy, Adjusted Rand Index (ARI) và Normalized Mutual Information (NMI).

Bảng 2. So sánh chất lượng cụm giữa FCM và MVO-FC

Phương pháp	Fuzzy Silhouette ↑	Accuracy ↑	Adjusted Rand Index ↑	NMI ↑
FCM	0,52	0,80	0,79	0,72
MVO-FC	0,54	0,98	0,94	0,91

Kết quả thực nghiệm cho thấy việc tích hợp thuật toán MVO vào quá trình tối ưu hóa vị trí tâm cụm trong phương pháp FCM đã mang lại những cải thiện đáng kể về chất lượng phân cụm so với FCM truyền thống. Cụ thể:

Chỉ số Fuzzy Silhouette Index đạt giá trị 0,54 đối với MVO-FC, cao hơn mức 0,52 của FCM, phản ánh khả năng phân tách cụm tốt hơn, với khoảng cách giữa các cụm lớn hơn và mức độ chồng lấn giảm rõ rệt. Độ chính xác phân loại (Accuracy) cũng được cải thiện mạnh mẽ, tăng từ 0,80 lên 0,98 cho thấy khả năng gán nhãn đúng cho các bệnh nhân được nâng cao rõ rệt nhờ tối ưu hóa vị trí tâm cụm.

Chỉ số ARI ghi nhận mức tăng từ 0,79 lên 0,94 thể hiện mức độ tương đồng cao hơn giữa kết quả phân cụm và nhãn thực tế, đồng thời cho thấy khả năng giảm thiểu sự nhầm lẫn giữa các nhóm bệnh.

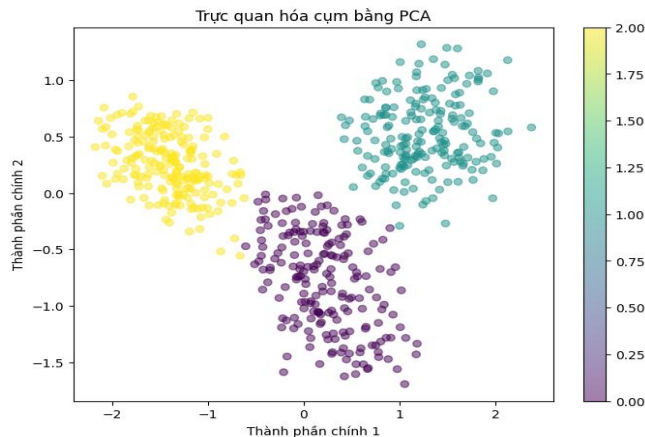
Bên cạnh đó, NMI tăng từ 0,72 lên 0,91, phản ánh mức độ bảo toàn thông tin giữa kết quả phân cụm và phân loại thực tế được cải thiện rõ rệt.

Về khả năng nhận diện đúng và đầy đủ các mẫu thuộc nhóm mục tiêu, chỉ số Precision đạt 0,92 và Recall đạt 0,91. Những kết quả này cho thấy MVO-FC không chỉ đảm bảo độ chính xác cao mà còn duy trì tính ổn định và khả năng bao phủ tốt trong bối cảnh dữ liệu lâm sàng phức tạp và dễ gây nhầm lẫn.

4.3. Thảo luận

Phân tích chi tiết cho thấy FCM truyền thống dễ gặp sai sót trong các tình huống mà các triệu chứng đặc trưng (chẳng hạn như sốt cao và ho nặng) xuất hiện đồng thời ở nhiều nhóm bệnh. Ví dụ, một bệnh nhân có nhiệt độ cơ thể 39°C, mức độ ho đạt 7 điểm và có biểu hiện khó thở đã bị FCM gán nhầm vào nhóm COVID-19 thay vì cúm mùa, do thuật toán không phân biệt hiệu quả các đặc trưng tinh tế như SpO₂ = 95,1% và CRP = 16,1 mg/L. Ngược lại, MVO-FC với khả năng tối ưu hóa toàn cục vị trí tâm cụm đã khai thác hiệu quả các đặc trưng phân biệt, giúp tăng cường độ chính xác và khả năng phân tách các nhóm bệnh có biểu hiện tương đồng.

Trong những trường hợp biểu hiện lâm sàng rõ rệt, ví dụ như bệnh nhân có nhiệt độ 39,5°C, ho ở mức 8 điểm, SpO₂ = 89% và CRP = 110 mg/L – điển hình cho thể COVID-19 nặng – cả FCM và MVO-FC đều cho kết quả phân cụm chính xác. Tuy nhiên, MVO-FC vẫn thể hiện ưu thế nổi bật nhờ khả năng tối ưu hóa toàn cục, góp phần nâng cao độ chính xác và độ tin cậy của phân cụm trong những tình huống có biên độ biến thiên lớn.



Hình 3. Đánh giá trực quan phân cụm bằng PCA

Hình 3 trình bày kết quả trực quan hóa không gian đặc trưng thông qua Phân tích Thành phần Chính (PCA), cho thấy ba cụm dữ liệu được tách biệt rõ ràng, xác nhận khả năng phân biệt hiệu quả giữa các nhóm bệnh của phương pháp MVO-FC. Khoảng cách giữa các cụm lớn và hiện tượng chồng lấn giữa các nhóm gần như không đáng kể, phản ánh mức độ chính xác và tính hiệu quả của mô hình trong việc xác định ranh giới giữa các nhóm dữ liệu.

Dựa trên các tâm cụm được xác định, nhãn bệnh tương ứng của từng cụm được gán như sau: Cụm 1 ứng với nhóm bệnh COVID-19, Cụm 2 đại diện cho bệnh Viêm phổi và Cụm 3 tương ứng với bệnh Cúm mùa. Ma trận thành viên mờ U thể hiện mức độ thuộc về của từng bệnh nhân đối với các cụm, với kết quả cụ thể cho sáu bệnh nhân đầu tiên trong Bảng 1 như sau: BN1 được phân vào nhóm COVID-19, BN2 và BN3 vào nhóm Cúm, BN4 và BN5 vào nhóm COVID-19, BN6 thuộc nhóm Viêm phổi. Khi đối chiếu với nhãn bệnh thực tế (Disease) trong Bảng 1, mô hình MVO-FC đạt độ chính xác phân loại (Accuracy) ở mức 0,98.

Kết quả này cho thấy việc tích hợp thuật toán MVO vào quá trình tối ưu hóa vị trí tâm cụm của FCM đã cải thiện đáng kể khả năng hội tụ, giảm nhiễu trong phân cụm, đồng thời hạn chế tình trạng nhầm lẫn giữa các nhóm bệnh có biểu hiện lâm sàng tương đồng. Những đặc điểm này góp phần khẳng định tiềm năng ứng dụng của mô hình MVO-FC trong bài toán hỗ trợ chẩn đoán và phân loại bệnh hô hấp phức tạp.

$$U = \begin{bmatrix} 0,67 & 0,15 & 0,18 \\ 0,27 & 0,11 & 0,62 \\ 0,24 & 0,12 & 0,64 \\ 0,57 & 0,17 & 0,26 \\ 0,58 & 0,21 & 0,21 \\ 0,22 & 0,66 & 0,12 \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

5. KẾT LUẬN

Nghiên cứu này đã đề xuất một hướng tiếp cận nhằm nâng cao hiệu quả của phương pháp phân cụm mờ thông qua việc tích hợp thuật toán MVO. Phương pháp kết hợp MVO-FC đã được áp dụng trong bài toán phân cụm dữ liệu bệnh nhân mắc các bệnh hô hấp, bao gồm cúm, COVID-19 và viêm phổi. Kết quả thực nghiệm cho thấy mô hình MVO-FC vượt trội so với phương pháp FCM truyền thống, với các chỉ số đánh giá như Fuzzy Silhouette Index, Adjusted Rand Index và Normalized Mutual Information đều đạt giá trị cao hơn đáng kể.

Việc ứng dụng thuật toán MVO đã nâng cao khả năng tối ưu toàn cục, giúp phương pháp tránh được hiện tượng hội tụ vào các điểm cực tiểu cục bộ và cải thiện rõ rệt khả năng phân biệt ranh giới giữa các cụm. Nhờ đó, độ chính xác trong việc phân loại bệnh nhân dựa trên đặc điểm triệu chứng và kết quả xét nghiệm được cải thiện một cách đáng kể. Các kết quả đạt được khẳng định tiềm năng ứng dụng thực tiễn của phương pháp MVO-FC trong các hệ thống hỗ trợ chẩn đoán bệnh, đặc biệt là đối với các bệnh hô hấp có triệu chứng tương đồng dễ gây nhầm lẫn.

Trong các hướng phát triển tiếp theo, việc tích hợp thêm các kỹ thuật tối ưu hóa khác, chẳng hạn như thuật toán di truyền, sẽ được xem xét nhằm tiếp tục nâng cao độ chính xác và tính ổn định của mô hình. Bên cạnh đó, việc mở rộng thử nghiệm trên các tập dữ liệu có quy mô lớn hơn và mức độ đa dạng cao hơn cũng là một định hướng quan trọng, nhằm đánh giá tính khả thi và năng lực áp dụng của phương pháp trong các tình huống lâm sàng phức tạp ngoài môi trường kiểm soát.

TÀI LIỆU THAM KHẢO

- [1] P. D. P. Silitonga, "Clustering of Patient Disease Data by Using K-Means Clustering," *International Journal of Computer Science and Information Security*, vol. 15, no. 7, pp. 219–221, 2017.
- [2] O. Pasin and S. Gonenc, "An Investigation into Epidemiological Situations of COVID-19 with Fuzzy K-Means and K-Prototype Clustering Methods," *Scientific Reports*, vol. 13, no. 1, Art. no. 6255, 2023, doi: <https://doi.org/10.1038/s41598-023-33214-y>
- [3] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press 1981, doi: <https://doi.org/10.1007/978-1-4757-0450-1>
- [4] W. S. Utami, S. Artika, and R. Aldiansyah, "Data Clustering of Confirmed COVID-19 Patients Using Fuzzy C-Means," *International Journal of Engineering, Technology and Natural Sciences*, vol. 5, no. 1, pp. 37–44, 2023.
- [5] J. K. Deen, R. Ganesan, and A. Merline, "Fuzzy-C-Means Clustering Based Segmentation and CNN-Classification for Accurate Segmentation of Lung Nodules," *Asian Pacific Journal of Cancer Prevention*, vol. 18, no. 7, pp. 1869–1873, 2017, doi: <https://doi.org/10.22034/APJCP.2017.18.7.1869>
- [6] Y. Wu, H. Duan, and S. Du, "Multiple Fuzzy C-Means Clustering Algorithm in Medical Diagnosis," *Technology and Health Care*, vol. 23, Suppl. 2, pp. S519–S527, 2015, doi: <https://doi.org/10.3233/THC-150989>
- [7] K. Bhalla and A. Gosain, "Performance Analysis of Hybridized Fuzzy Clustering Algorithms Using Metaheuristic Algorithms," in *Lecture Notes in Networks and Systems*, vol. 680, pp. 445–461, 2023, doi: https://doi.org/10.1007/978-981-99-2602-2_34

- [8] J. Zhang and Z. Ma, “Hybrid Fuzzy Clustering Method Based on FCM and Enhanced Logarithmical PSO (ELPSO),” *Computational Intelligence and Neuroscience*, vol. 2020, no. 1, Art. no. 1386839, 2020, doi: <https://doi.org/10.1155/2020/1386839>
- [9] S. Tongbram *et al.*, “A Novel Image Segmentation Approach Using FCM and Whale Optimization Algorithm,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 11565–11580, 2021, doi: <https://doi.org/10.1007/s12652-020-02762-w>
- [10] S. Mirjalili, S. M. Mirjalili, and A. Hatamlou, “Multi-Verse Optimizer: A Nature-Inspired Algorithm for Global Optimization,” *Neural Computing and Applications*, vol. 27, no. 2, pp. 495–513, 2016, doi: <https://doi.org/10.1007/s00521-015-1870-7>
- [11] S. M. Ross, *Introduction to Probability Models*, New York, Academic Press, 2014.
- [12] N. Chen *et al.*, “Epidemiologic and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China,” *The Lancet*, vol. 395, no. 10223, pp. 507–513, 2020, doi: [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- [13] T. M. Uyeki *et al.*, “Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenza,” *Clinical Infectious Diseases*, vol. 68, no. 6, pp. 895–902, 2019, doi: <https://doi.org/10.1093/cid/ciy874>
- [14] J. P. Metlay *et al.*, “Diagnosis and treatment of adults with community-acquired pneumonia: An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America,” *American Journal of Respiratory and Critical Care Medicine*, vol. 200, no. 7, pp. e45–e67, 2019, doi: <https://doi.org/10.1164/rccm.201908-1581ST>
- [15] D. Wang *et al.*, “Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China,” *JAMA*, vol. 323, no. 11, pp. 1061–1069, 2020, doi: <https://doi.org/10.1001/jama.2020.1585>
- [16] Y. Liu *et al.*, “Association between age and clinical characteristics and outcomes of COVID-19,” *European Respiratory Journal*, vol. 55, no. 5, Art. no. 2001112, 2020, doi: <https://doi.org/10.1183/13993003.01112-2020>

ABSTRACT

IMPROVING FUZZY CLUSTERING WITH MULTI-VERSE OPTIMIZER FOR CLASSIFYING RESPIRATORY DISEASES: INFLUENZA, COVID-19, AND PNEUMONIA

Nguyen Thi Thuy Trang, Dinh Nguyen Trong Nghia*

Ho Chi Minh City University of Industry and Trade

*Email: nghiadnt@huit.edu.vn

Accurately classifying respiratory diseases such as influenza, COVID-19, and pneumonia is crucial for diagnosis and treatment. However, due to the similarity in symptoms, grouping patients into distinct categories presents significant challenges. This study proposes an improvement to the fuzzy clustering algorithm by integrating the Multi-Verse Optimizer (MVO) to optimize cluster center positions, enhancing clustering quality and reducing misclassification among disease groups. By combining MVO's global optimization capability with the FCM method, the model achieves more precise disease cluster separation. Experimental results demonstrate that the proposed method outperforms traditional FCM, significantly improving clustering accuracy and disease group differentiation. This approach has the potential to be applied in supporting respiratory disease diagnosis, contributing to increased efficiency in medical practice.

Keywords: Fuzzy clustering, Multi-Verse Optimizer, Fuzzy C-Means, respiratory disease classification, medical diagnosis.