

Thư Tòa soạn

TÌNH HÌNH NGHIÊN CỨU KHOA HỌC TẠI VIỆT NAM TRONG GIAI ĐOẠN 2019-2023 VÀ ĐỊNH HƯỚNG NGHIÊN CỨU TRONG Y KHOA

Nguyễn Minh Tâm^{1*}, Nguyễn Văn Nhơn², Nguyễn Văn Thành²

¹Khoa Cơ bản, Trường Đại học Phan Thiết, 225 Nguyễn Thông, Phan Thiết, Bình Thuận, Việt Nam

²Khoa Khoa học Sức khỏe, Trường Đại học Phan Thiết, 225 Nguyễn Thông, Phan Thiết, Bình Thuận, Việt Nam

Tóm tắt: Nghiên cứu này tiến hành phân tích xu hướng công bố khoa học tại Việt Nam trong giai đoạn 2019-2023, với trọng tâm là mối quan hệ giữa số lượng công bố, số lượng trích dẫn và chỉ số JNCI (Journal Normalized Citation Impact). Dữ liệu được thu thập từ cơ sở dữ liệu InCites của Web of Science, sau đó được xử lý và phân tích nhằm rút trích các kết quả có ý nghĩa khoa học quan trọng. Một trong những kết quả phân tích cho thấy rằng, mặc dù một số ít cơ sở nghiên cứu có số lượng công bố thấp, nhưng chỉ số JNCI của họ lại cao. Ngược lại, phần lớn các đơn vị có số lượng công bố nhiều hơn nhưng chỉ số JNCI lại thấp hơn. Phương pháp nghiên cứu kết hợp giữa các công cụ thống kê truyền thống và các thuật toán phân cụm hiện đại như K-means clustering và DBSCAN. Ngoài ra, kỹ thuật word cloud được sử dụng để minh họa trực quan, giúp phân tích sâu các đặc trưng của từng nhóm nghiên cứu. Kết quả cho thấy, các cơ sở có sự hợp tác quốc tế và công bố trên các tạp chí thuộc nhóm Q1, Q2 có mức độ ảnh hưởng cao hơn đáng kể, thể hiện qua số lượng trích dẫn và chỉ số JNCI vượt trội. Đặc biệt, các lĩnh vực về y tế và khoa học sức khỏe trở nên nổi bật nhờ sự quan tâm ở cấp độ quốc tế và tính cấp thiết của các nghiên cứu trong bối cảnh đại dịch COVID-19. Những phát hiện này nhấn mạnh vai trò của chiến lược hợp tác quốc tế và việc công bố trên các tạp chí khoa học uy tín nhằm tăng cường chất lượng và tầm ảnh hưởng của nghiên cứu khoa học Việt Nam trên bình diện thế giới.

Từ khóa: Việt Nam, nghiên cứu khoa học, y khoa, sức khỏe, InCites, WoS, trích dẫn, ...

1. GIỚI THIỆU

Cùng với bước tiến của khoa học kỹ thuật của nhân loại, hoạt động công bố khoa học đang diễn ra sôi nổi hơn bao giờ hết. Trên nền tảng các công trình khoa học đã được công bố trước đó, người làm nghiên cứu có thể theo đuổi và phát triển hướng nghiên cứu, thực hiện các nghiên cứu có giá trị và lại tiếp tục đóng góp tri thức cho cộng đồng thông qua việc công bố nghiên cứu của mình. Công bố khoa học đã trở thành tiêu chí quan trọng để đánh giá năng lực không chỉ của cá nhân các nhà nghiên cứu mà còn của các cơ sở nghiên cứu khoa học, bao gồm các trường đại học và viện nghiên cứu trên toàn thế giới (Binswanger, 2014).

Bên cạnh đó, số lượng các ấn phẩm khoa học trong từng lĩnh vực cụ thể cũng phản ánh được các yếu tố quan trọng như quy mô của lĩnh vực, sự quan tâm của cộng đồng khoa học, nhận thức của xã hội, tính thời sự của vấn đề được nghiên cứu. Những yếu tố này lại tác động lên các chính sách quản lý khoa học từ cấp độ quốc gia cho đến từng cơ sở nghiên cứu. Một ví dụ điển hình là số lượng các công bố khoa học liên quan đến đại dịch COVID-19. Theo báo cáo của Nature, trong 6 tháng đầu năm 2020 ước tính đã có hơn 23.500 ấn phẩm khoa học về COVID-19 được công bố, bao gồm các bài báo nghiên cứu, thư, bài xã luận, ghi chú và bài đánh giá (Teixeira da Silva, 2020). Tuy nhiên, cho tới tháng 10 năm 2020, con số

này đã tăng lên hơn 87.000 ấn phẩm và vẫn tiếp tục tăng nhanh sau đó khi các nhà khoa học trên toàn thế giới vẫn tiếp tục nghiên cứu virus Corona, là tác nhân gây đại dịch, và tác động của nó đối với sức khỏe người nhiễm bệnh (Grabmeier, 2021). So sánh với số lượng ấn phẩm công bố khoa học trong lĩnh vực khoa học nano, là một trong những chủ đề khoa học nóng nhất những năm 1990, thì các nghiên cứu trong lĩnh vực khoa học nano phải mất hơn 19 năm để tăng từ 4.000 lên 90.000 bài báo khoa học trong khi số lượng công trình nghiên cứu về virus corona đã đạt đến mức đó chỉ trong khoảng năm tháng (Teixeira da Silva, 2020). Các số liệu này đã chứng minh một cách rõ ràng tính thời sự và sự quan tâm sâu sắc của cộng đồng đối với đại dịch COVID-19.

Trong những năm gần đây, nghiên cứu khoa học tại Việt Nam đã đạt được những bước tiến đáng kể cả về số lượng công bố và mức độ ảnh hưởng (Nguyễn Minh Quân, 2020; Lương Đình Hải, 2021). Đáng chú ý, các lĩnh vực liên quan đến y tế và khoa học sức khỏe nhận được sự quan tâm rất lớn của cộng đồng, đặc biệt là trong bối cảnh đại dịch COVID-19 (Van Luong Nguyen et.al., 2022; Bach Xuan Tran et.al., 2022). Nhiều cơ sở nghiên cứu tại Việt Nam đã và đang tăng cường các hoạt động hợp tác cả trong nước lẫn quốc tế với mục đích có thể công bố các công trình nghiên cứu của mình trên các tạp chí khoa học uy tín có quy trình bình duyệt chặt chẽ thuộc hai cơ sở dữ liệu khổng lồ là Scopus và WoS (Web of Science) (Pranckutė, 2021). Cả hai cơ sở dữ liệu này đều cung cấp các chỉ số trắc lượng khoa học để đánh giá sự phát triển và uy tín của các bài báo khoa học, của nhà khoa học - là tác giả của các bài báo khoa học được công bố, và của cơ sở nghiên cứu - nơi mà nghiên cứu được thực hiện. Việc hiểu rõ các chỉ số trắc lượng về công bố khoa học có ý nghĩa rất quan trọng vì thông qua đó có thể đánh giá chất lượng của công bố khoa học cũng như tác động của nó đối với cộng đồng

nghiên cứu. Các chỉ số này cung cấp thông tin về tần suất được trích dẫn, uy tín của tạp chí hoặc tổ chức phát hành, và sự ảnh hưởng của công bố đó đến lĩnh vực nghiên cứu cụ thể. Điều này giúp cho các nhà nghiên cứu, cơ quan quản lý hoặc tổ chức tài trợ có thể đưa ra các quyết định phù hợp dựa trên dữ liệu và thông tin đáng tin cậy. Tuy nhiên, việc đánh giá năng lực của một cơ sở nghiên cứu thông qua sự tăng trưởng về số lượng bài báo khoa học, tổng số trích dẫn, và các chỉ số trắc lượng khác qua các năm hoặc từng nhóm đối tượng quan tâm vẫn đang là một vấn đề chưa được giải đáp hoàn toàn đối với các cơ sở nghiên cứu tại Việt Nam. Để trả lời cho câu hỏi này, nhóm nghiên cứu đã thực hiện nghiên cứu này bằng cách thu thập dữ liệu và phân tích xu hướng công bố khoa học tại Việt Nam từ năm 2019 đến 2023, tập trung vào mối liên hệ giữa số lượng công bố, trích dẫn, và chỉ số JNCI, cả từ góc độ toàn bộ và theo từng giai đoạn cụ thể của từng nhóm quan sát.

2. PHƯƠNG PHÁP NGHIÊN CỨU

InCites là một sản phẩm của hệ thống Clarivate, cung cấp cơ sở dữ liệu liên quan đến công bố của các nhà khoa học, tổ chức nghiên cứu, tạp chí khoa học,... theo phạm vi quốc gia, khu vực về các lĩnh vực khoa học tự nhiên và kỹ thuật (theo chỉ mục SCIE - *Science Citation Index Expanded*) và khoa học xã hội (theo chỉ mục SSCI - *Social Sciences Citation Index*), cả hai chỉ mục này đều nằm trong cơ sở dữ liệu (CSDL) của WoS (Clarivate, 2024). Đồng thời là công cụ hỗ trợ cho các nhà khoa học hay bộ phận hoạch định nghiên cứu nhằm mục đích phân tích và đánh giá xu hướng nghiên cứu trong nội bộ đơn vị hoặc trong các đơn vị khác quan tâm. Do đó, trong phạm vi nghiên cứu này, nhóm tác giả quan tâm đến tình hình nghiên cứu khoa học của các cơ sở nghiên cứu được InCites ghi nhận, dữ liệu truy xuất vào ngày 20/08/2024. Dữ liệu thu thập được

tiến hành xử lý và phân tích bằng ngôn ngữ lập trình Python, là một trong những công cụ phổ biến trong phân tích dữ liệu hiện nay và được sử dụng rộng rãi trong lĩnh vực phát triển phần mềm, khoa học dữ liệu, trí tuệ nhân tạo và phát triển web (McKinney, 2017).

Trong quá trình mô tả và phân tích dữ liệu, một số phương pháp trực quan hóa khác nhau được áp dụng nhằm biểu diễn các đặc điểm của tập dữ liệu. Cụ thể, một số biểu đồ sử dụng gồm biểu đồ cột (bar chart) để so sánh giá trị giữa các nhóm; biểu đồ hộp (boxplot) để tóm tắt sự phân tán của dữ liệu, dễ dàng quan sát tứ phân vị, giá trị trung bình và xác định trường hợp ngoại lệ (outliers); biểu đồ tần suất (histogram) minh họa phân phối xác suất của dữ liệu liên tục, cung cấp thông tin về độ lệch (skewness) và mức độ phân tán. Đối với dữ liệu chuỗi thời gian, biểu đồ đường (line chart) thể hiện sự biến động của các giá trị qua các năm nhằm làm rõ xu hướng dữ liệu theo thời gian. Cuối cùng, word cloud (đám mây từ) được áp dụng để minh họa tần suất xuất hiện của các chủ đề xuất hiện trong công trình công bố, trong đó các từ xuất hiện nhiều hơn được hiển thị với kích thước lớn hơn, giúp dễ dàng nhận biết các từ khóa quan trọng (DePaolo và Wilkinson, 2014).

Đồng thời, nhóm áp dụng 2 phương pháp phân cụm thông dụng là K-Means (Jain, 2010) và DBSCAN (Ester et. al., 1996). K-means là bài toán phân cụm không giám sát, có mục tiêu phân chia dữ liệu thành k nhóm dựa trên sự tương đồng giữa các điểm dữ liệu, với ý tưởng là chia tập dữ liệu thành k cụm sao cho tổng bình phương khoảng cách giữa các điểm dữ liệu và trung tâm của cụm là nhỏ nhất. Ngược lại, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) là thuật toán phân dựa trên mật độ của dữ liệu, tự động xác định các cụm dựa trên mật độ điểm trong không gian, có thể

phát hiện các điểm nhiễu. Tuy nhiên, một nhược điểm của K-means là độ nhạy với các centroid k khởi tạo vì nó có xu hướng hội tụ đến nghiệm cục bộ, điều này có thể dẫn đến kết quả không tối ưu. Ngược lại, DBSCAN có lợi thế trong việc phân cụm dữ liệu không đều nhau, nhưng tham số ϵ (bán kính) và số điểm tối thiểu trong cụm có ảnh hưởng đến việc phát hiện cụm.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Tổng quan về nghiên cứu khoa học tại Việt Nam

Từ hệ thống dữ liệu InCites được truy xuất vào ngày 20/08/2024, CSDL thu thập trong 05 năm (2019-2023) bao gồm các yếu tố (InCites Help Center, 2021) như sau:

+ Số lượng công trình công bố (*Document*): là số lượng bài báo khoa học được xuất bản bởi các tạp chí có trong CSDL WoS.

+ Lượt trích dẫn (*Times Cited*): là tổng số trích dẫn của các công trình công bố.

+ Chỉ số ảnh hưởng trích dẫn được chuẩn hóa của tạp chí (*Journal Normalized Citation Impact - JNCI*): là tỷ lệ giữa số lượng trích dẫn thực tế với tỷ lệ trích dẫn trung bình của các công trình công bố trong cùng một tạp chí trong cùng một năm và cùng loại tài liệu. Chỉ số này cho thấy hiệu suất của ấn phẩm liên quan đến cách các nhà nghiên cứu khác thực hiện khi họ xuất bản công trình của mình trên tạp chí. Nếu giá trị số của JNCI vượt quá một, thì công trình nghiên cứu được đánh giá trên mức trung bình và ngược lại.

+ Vị trí của tác giả: có 3 vị trí tác giả, bao gồm tác giả đầu tiên (*first author*), đồng tác giả (*co-author*) và tác giả liên hệ (*corresponding author*).

+ Hệ số ảnh hưởng của tạp chí (*Journal Impact Factor - JIF*) là tỷ lệ giữa tổng tất cả các trích dẫn trong năm hiện tại của các bài báo được công bố trong hai năm trước và tổng số công trình công bố (bao gồm các

bài báo - articles, bài đánh giá - reviews và bài báo kỷ yếu - proceedings papers) được xuất bản trên tạp chí trong hai năm trước.

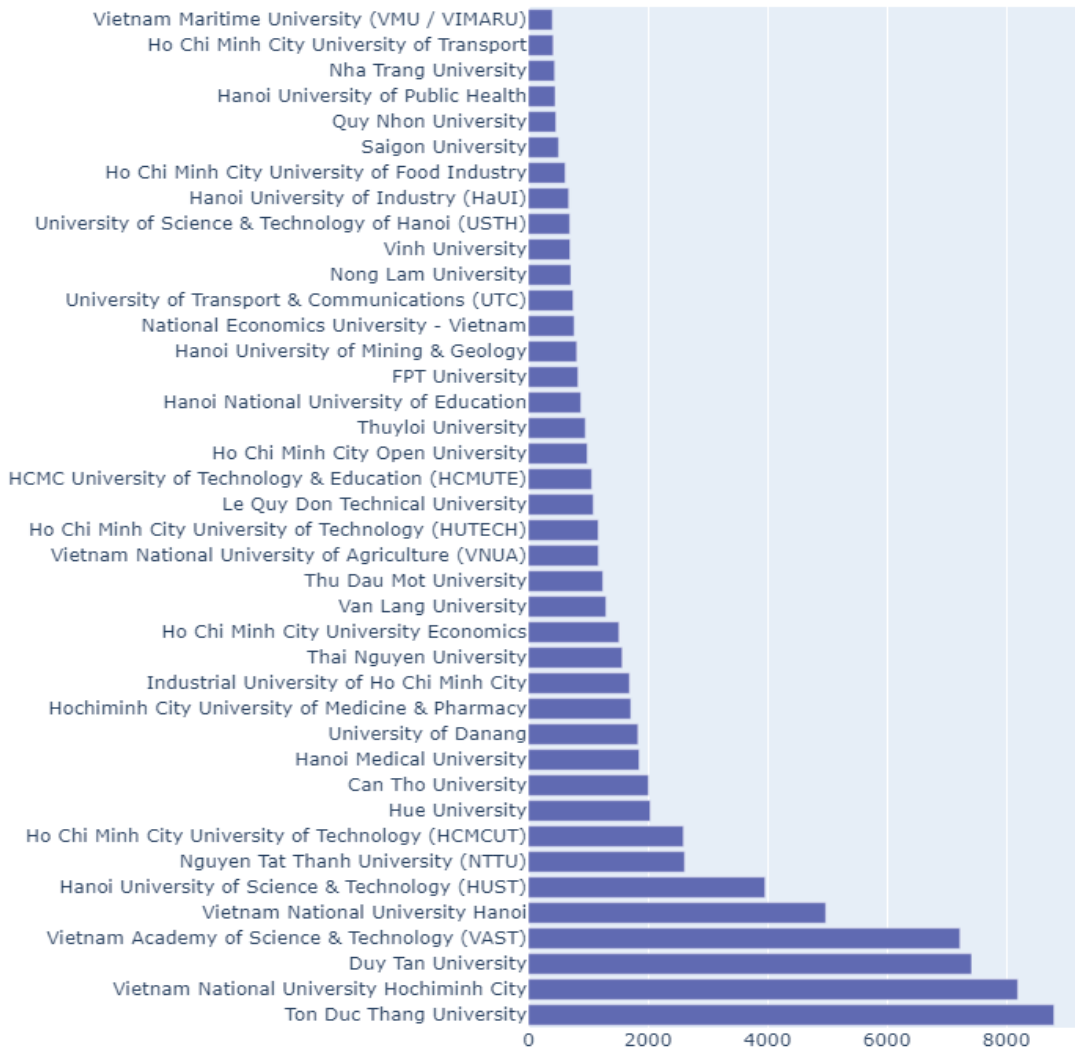
+ Chỉ số trích dẫn của tạp chí (*Journal Citation Indicator – JCI*) là số lần trích dẫn trung bình mà các bài báo của một tạp chí so với số trích dẫn trung bình của các bài báo tương tự trong cùng lĩnh vực và cùng năm. Giá trị này thể hiện tác động trích dẫn chuẩn hóa theo danh mục trung bình đối với các bài báo được xuất bản trong giai đoạn ba năm trước đó.

Từ dữ liệu thu thập trong 05 năm của 75 cơ sở có công trình công bố được hệ thống

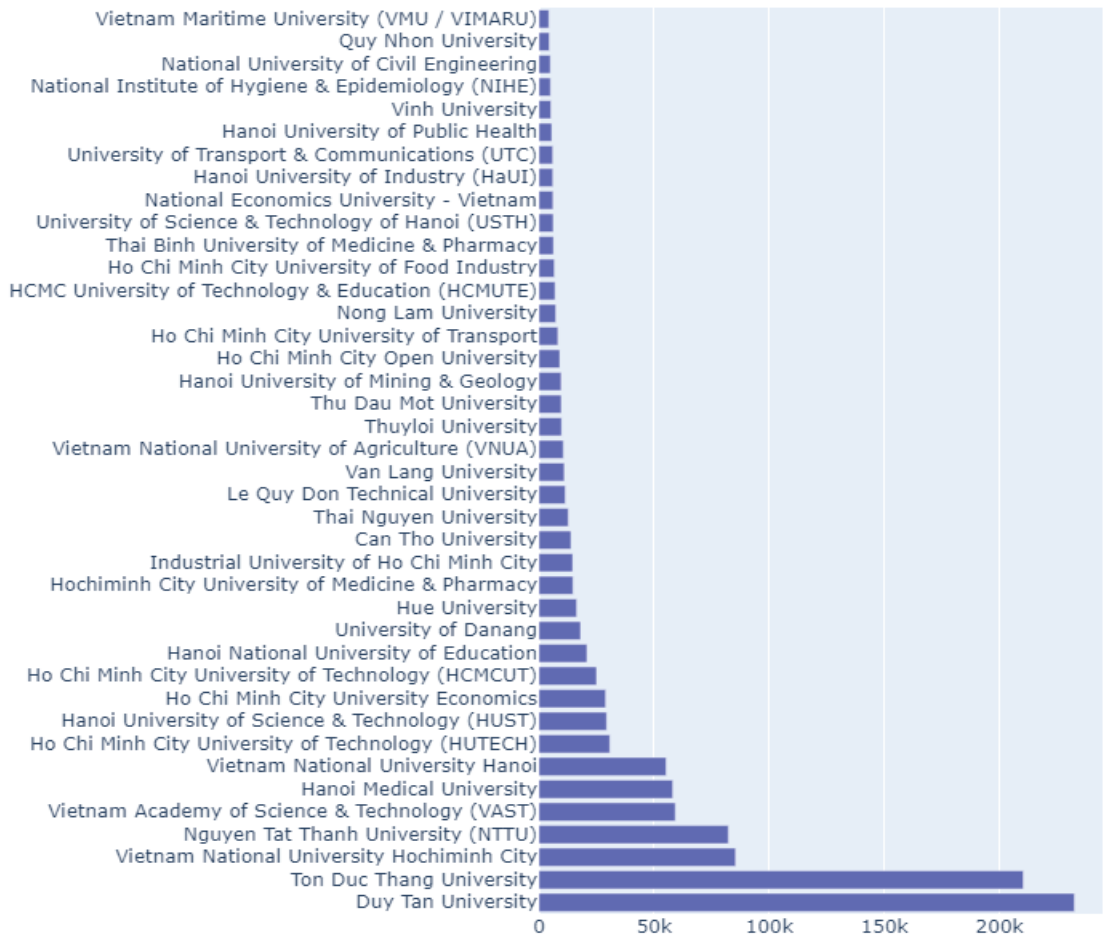
InCites ghi nhận, nhóm nghiên cứu tiến hành quan sát, phân tích và có một số nhận xét như sau:

3.1.1 Hiệu suất nghiên cứu

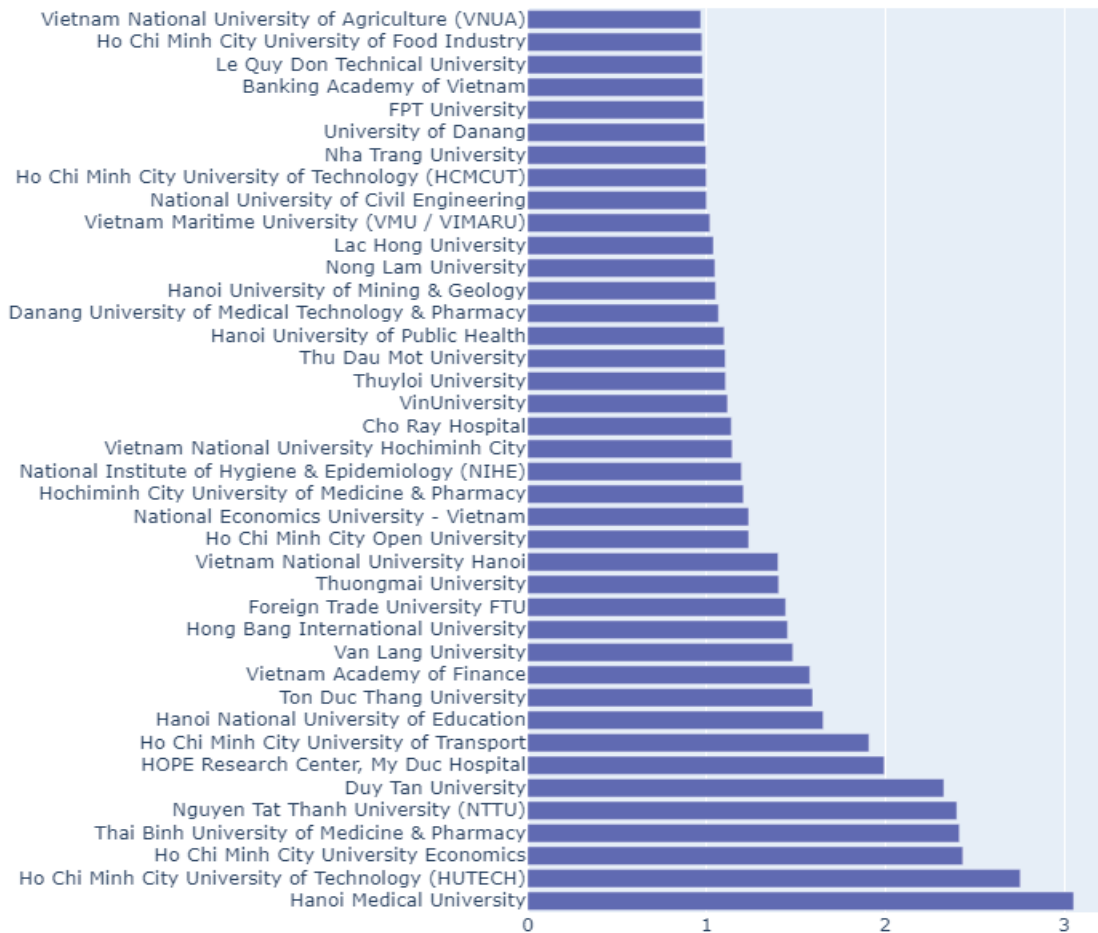
Quan sát phân bố dữ liệu của các đơn vị nghiên cứu cho thấy mỗi đơn vị áp dụng chiến lược phát triển nghiên cứu khoa học khác nhau, được thể hiện qua sự đa dạng các chỉ số. Hình 1, Hình 2, và Hình 3 lần lượt thể hiện 40 đơn vị nghiên cứu đứng đầu về các chỉ số trích lượng, bao gồm số lượng bài báo (Hình 1), tổng số trích dẫn (Hình 2), và chỉ số JNCI (Hình 3).



Hình 1. Top 40 đơn vị có số lượng bài báo cao nhất trong giai đoạn 2019-2023



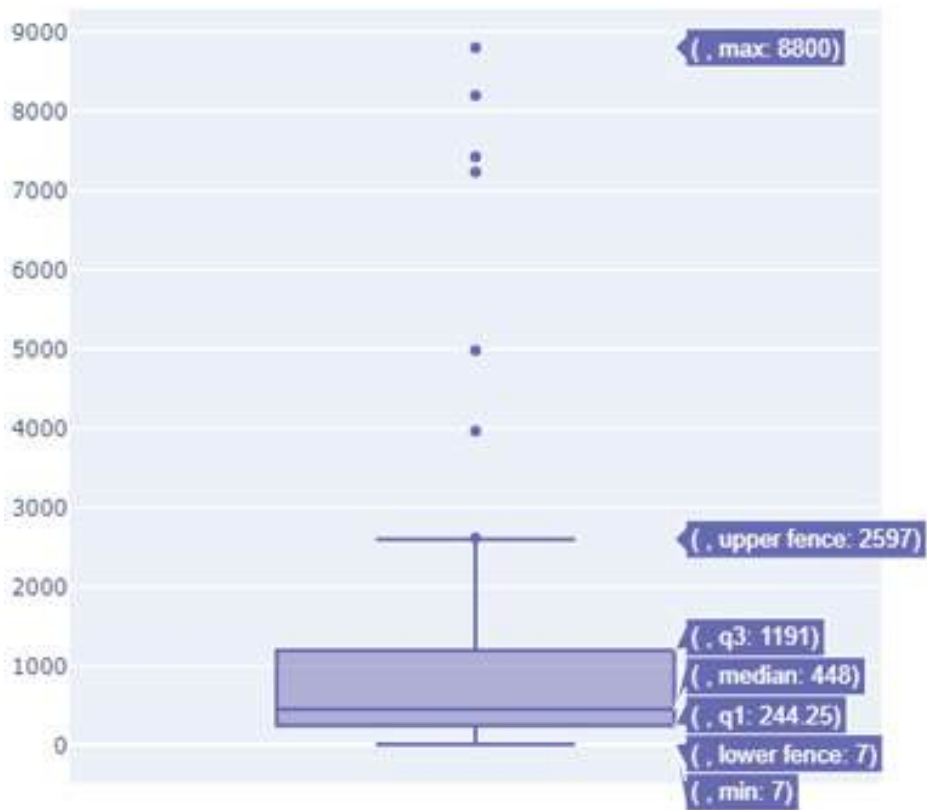
Hình 2. Top 40 đơn vị có tổng số trích dẫn cao nhất trong giai đoạn 2019-2023



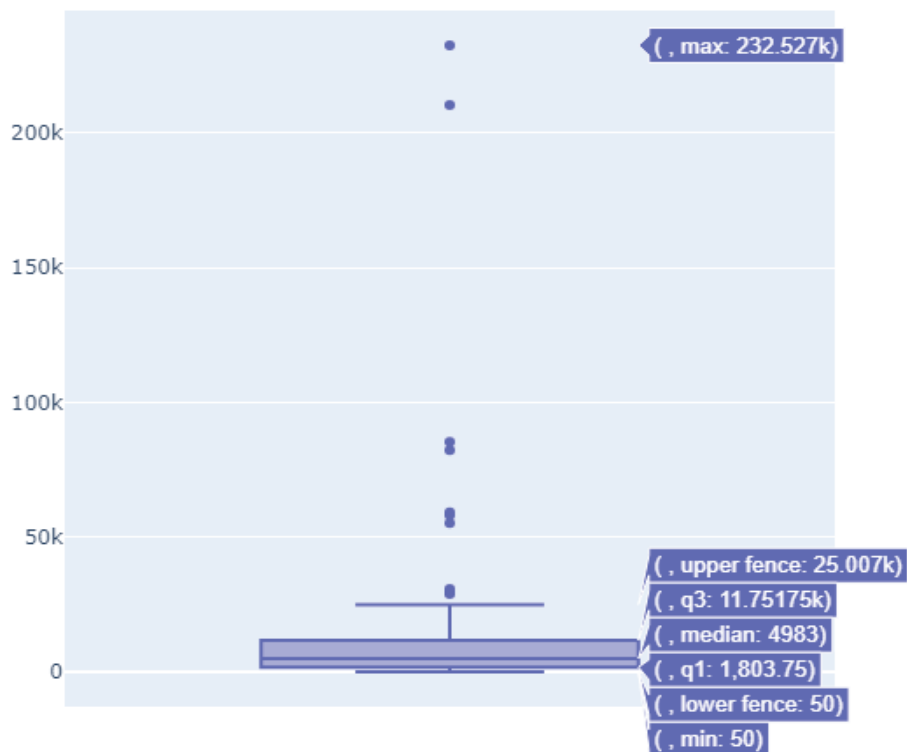
Hình 3. Top 40 đơn vị có chỉ số JNCI cao nhất giai đoạn 2019-2023

Tuy nhiên, kết quả của kiểm định Shapiro-Wilk với giá trị $p\text{-value} < 0,001$ đối với bộ số liệu tổng số bài báo (Hình 4), tổng số trích dẫn (Hình 5), chỉ số JNCI (Hình 6), và tỷ lệ phần trăm số bài báo được trích dẫn (Hình 7) không tuân theo phân phối chuẩn. Điều này cho thấy dữ liệu tồn tại các yếu tố ngoại lệ hoặc sự biến thiên giữa các đơn vị lớn khá lớn. Hơn thế nữa, khi so sánh sự khác biệt giữa các đơn

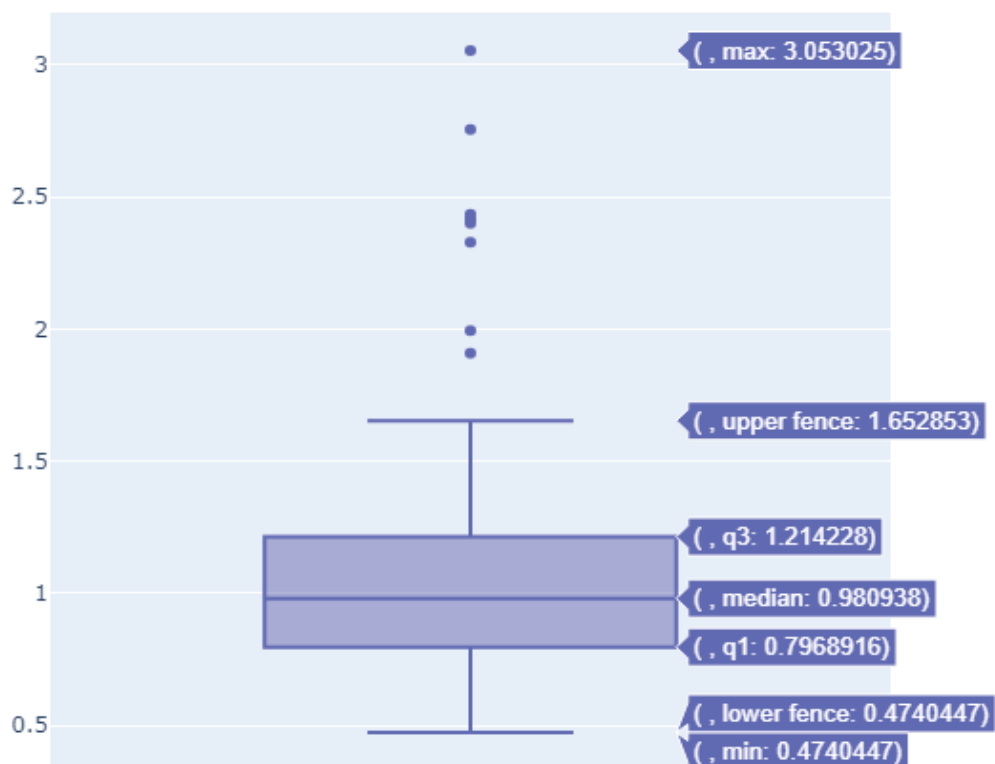
vị nghiên cứu dựa trên yếu tố số lượng bài báo, số lượng trích dẫn, và chỉ số JNCI, kết quả kiểm định Kruskal-Wallis cho thấy không có sự khác biệt đáng kể giữa các nhóm quan sát với mức ý nghĩa 5% ($p\text{-value} > 0.4$). Mặc dù có sự khác nhau về chiến lược nghiên cứu giữa các đơn vị, các yếu tố nghiên cứu khoa học không có sự khác biệt lớn giữa các đơn vị khi được kiểm định trên toàn bộ mẫu.



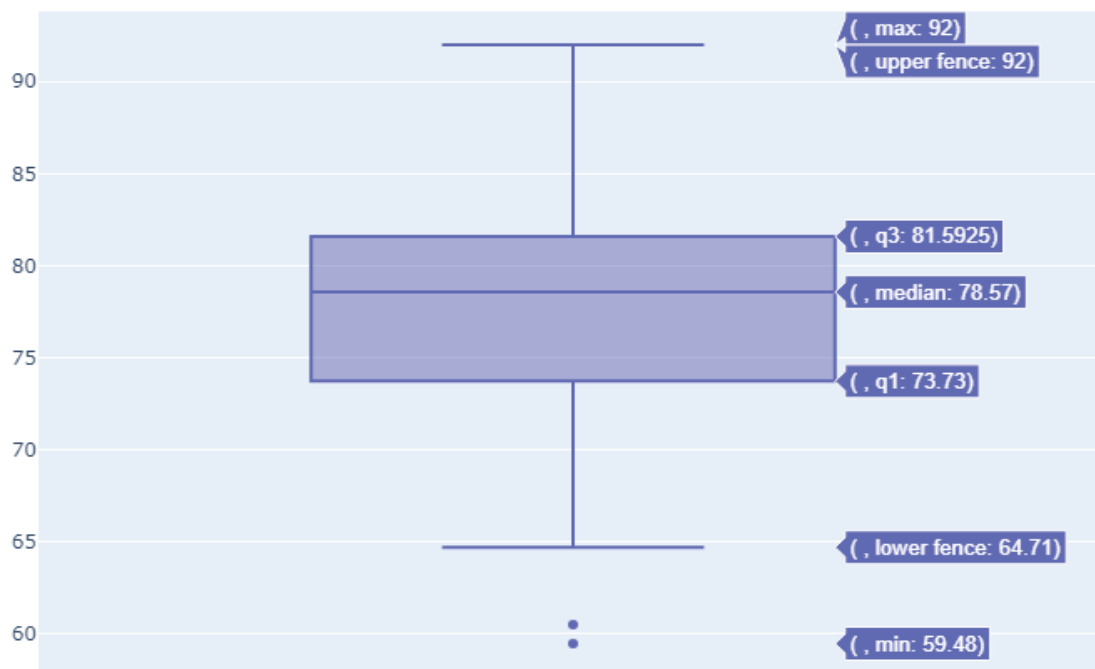
Hình 4. Boxplot (biểu đồ hộp) tổng số lượng bài báo giai đoạn 2019-2023



Hình 5. Boxplot tổng số lượng trích dẫn giai đoạn 2019-2023



Hình 6. Boxplot chỉ số JNCI giai đoạn 2019-2023

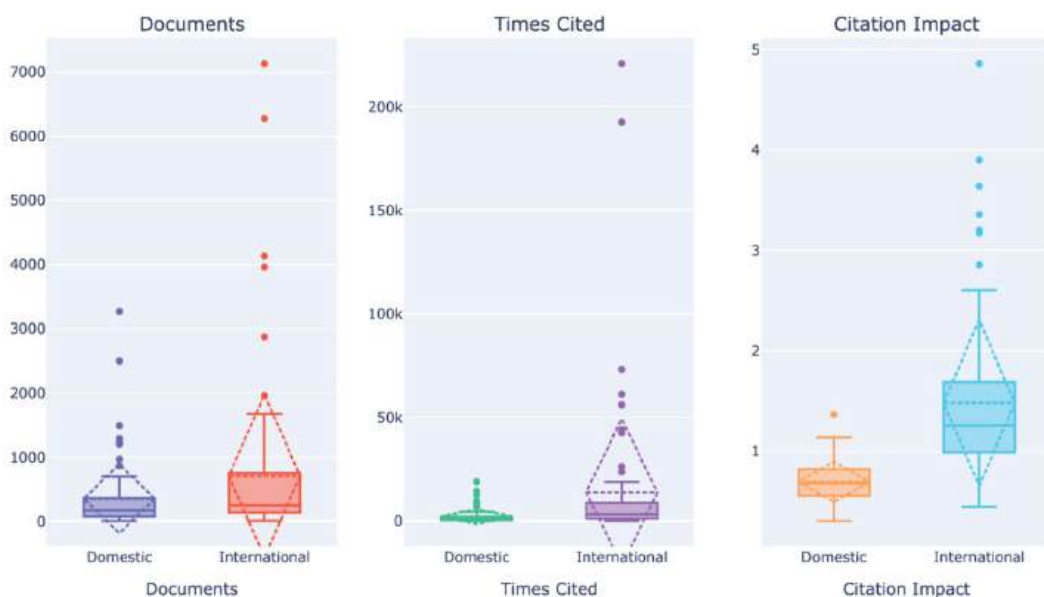


Hình 7. Boxplot tỷ lệ phần trăm số bài báo được trích dẫn giai đoạn 2019-2023

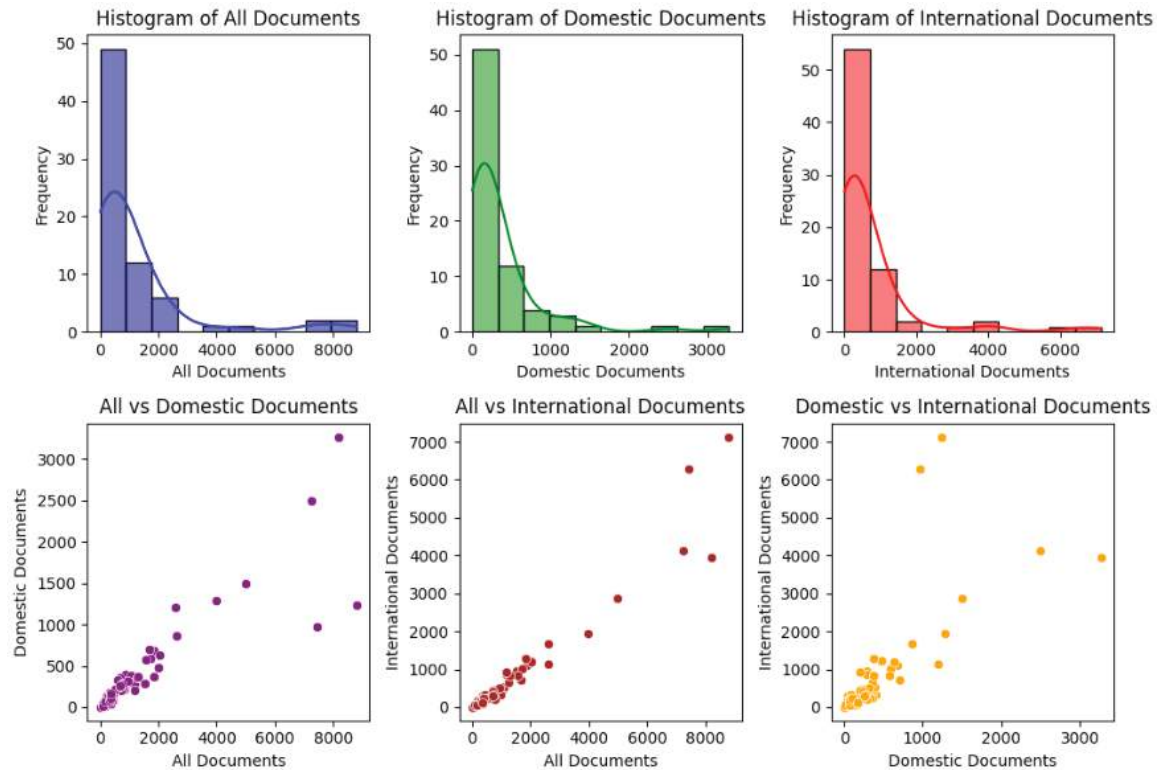
Dữ liệu cho thấy mặc dù Trường Đại học Tôn Đức Thắng và Trường Đại học Duy Tân có tổng số công trình công bố và tổng trích dẫn trong 5 năm là cao nhất trong 75 đơn vị nghiên cứu được khảo sát (Hình 1 và Hình 2) nhưng cơ sở nghiên cứu có chỉ số JNCI - là chỉ số phản ánh tỷ lệ giữa số lượng trích dẫn thực tế với tỷ lệ trích dẫn trung bình của các công trình công bố trong cùng một tạp chí trong cùng một năm và cùng loại tài liệu - lại là Trường Đại học Y Hà Nội (Hình 3). Đáng lưu ý, khi số lượng công bố tại Trường Đại học Y Hà Nội chỉ bằng 21,08% (1855/8800) số lượng công bố của Trường Đại học Tôn Đức Thắng và tổng số trích dẫn trong 05 năm của cơ sở này cũng chỉ bằng 25,0% (58111/232527) so với Trường Đại học Duy Tân. Kết quả này gián tiếp nói lên rằng tuy số lượng công bố và trích dẫn thấp hơn, nhưng các công trình nghiên cứu trong lĩnh vực y khoa tại Đại học Y Hà Nội vẫn đạt mức độ ảnh hưởng khoa học đáng kể, phản ánh sự đóng góp quan trọng của lĩnh vực này trong cộng đồng học thuật.

Phân tích tiếp theo tập trung vào hiệu suất nghiên cứu của các công trình có sự hợp

tác trong nước - là các bài báo có từ hai tác giả trở lên và có địa chỉ/tổ chức (*affiliation*) riêng biệt nhưng cùng thuộc một quốc gia, và hợp tác quốc tế - là các bài báo có ít nhất một đồng tác giả có địa chỉ ở nước ngoài. Dữ liệu thu thập trong giai đoạn 2019 – 2023 cho thấy xu hướng nghiên cứu của Việt Nam nghiêng về các công trình có hợp tác quốc tế. Dù trong nước hay ngoài nước, số lượng bài báo lớn đều tập trung vào một số ít cơ sở, chiếm tỷ lệ khoảng 1/4 cơ sở nghiên cứu. Dữ liệu cho thấy số lượng bài báo có hợp tác quốc tế cao gấp đôi so với bài báo trong nước (51.677 so với 25.793) nhưng tổng số lần trích dẫn từ các công trình quốc tế lại cao gấp 6,27 lần so với các bài trong nước (1.022.810 so với 163.188). Kết quả này chỉ ra các công trình có sự hợp tác quốc tế không chỉ chiếm ưu thế về số lượng, chất lượng mà còn có tác động nghiên cứu cao hơn, điều này thể hiện qua chỉ số trích dẫn (JNCI) vượt trội. Hình 8 và Hình 9 cho thấy được công trình nghiên cứu nếu có yếu tố nước ngoài thì số lượng trích dẫn và chỉ số ảnh hưởng trích dẫn nổi bật hơn so với bài trong nước.



Hình 8. Boxplot số bài báo, số trích dẫn, chỉ số JNCI trong 5 năm theo bài báo trong và ngoài nước

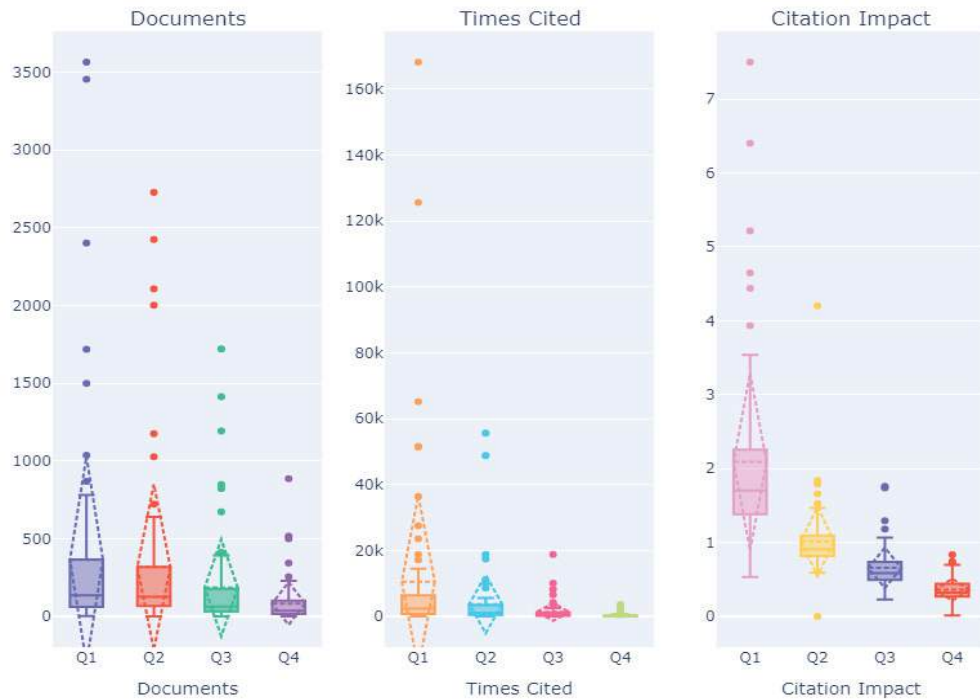


Hình 9. Biểu đồ tần suất (Histogram) số lượng và phân loại loại bài báo trong nước và quốc tế trong 5 năm (2019 – 2023)

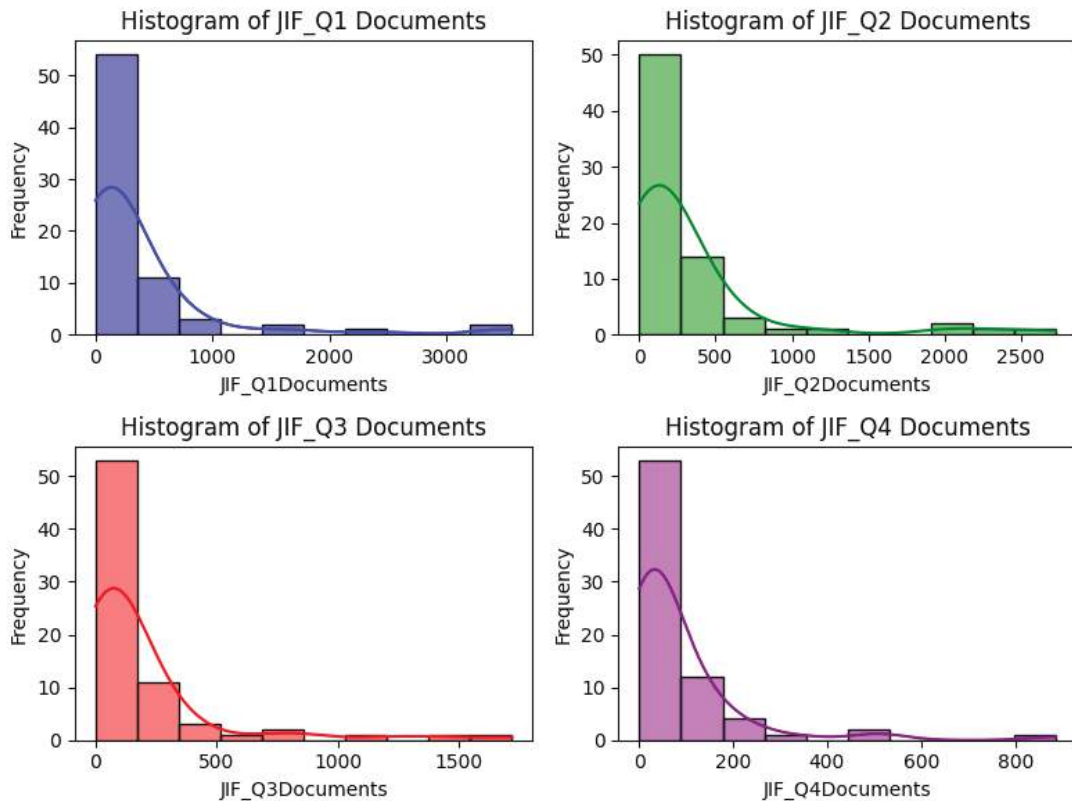
Cả hai chỉ số JIF và JCI đều đo lường mức độ ảnh hưởng của tạp chí dựa trên số lượng trích dẫn từ cơ sở dữ liệu WoS. Tuy nhiên, hai chỉ số này khác nhau về phương pháp tính toán, phạm vi áp dụng, và khả năng chuẩn hóa theo lĩnh vực. Chỉ số JIF tập trung vào số lượng trích dẫn trung bình của các bài báo trong hai năm và không chuẩn hóa theo lĩnh vực nghiên cứu. Ngược lại, JCI là một chỉ số mới hơn, chuẩn hóa theo lĩnh vực dựa trên tác động trích dẫn chuẩn hóa theo phân loại (*Category Normalized Citation Impact - CNCI*), sử dụng khung thời gian ba năm và áp dụng cho tất cả các tạp chí trong Web of Science Core Collection. Điều này giúp JCI cung cấp đánh giá công bằng và toàn diện hơn về tác động trích dẫn của các tạp chí trên nhiều lĩnh vực khoa học khác nhau. Các chỉ số này đều sử dụng tứ phân vị Q (*quartile*) để phân loại tạp chí, trong đó Q1 gồm các tạp chí có mức độ tiếp cận và

trích dẫn cao. Tạp chí thuộc Q2, Q3, và Q4 có mức độ ảnh hưởng giảm dần nhưng vẫn đóng vai trò quan trọng trong việc công bố khoa học.

Nhìn chung, hiệu suất công bố nghiên cứu khoa học của Việt Nam trong 5 năm qua chủ yếu tập trung vào các tạp chí thuộc nhóm Q1 và Q2 trong cơ sở dữ liệu WoS, với tỷ lệ khá cao: 71,77% (49878/69496) theo chỉ số JIF và 69,78% (55281/79224) theo chỉ số JCI. Về mặt trích dẫn, các bài báo thuộc nhóm Q1 cũng ghi nhận số lượng trích dẫn cao nhất theo cả hai chỉ số. Điều này cho thấy những bài báo trong Q1 có mức độ ảnh hưởng đáng kể nhất và điều này cũng thấy sự phân bố theo chỉ số JNCI. Hơn nữa, sự phân bố của các bài báo theo các Q cho thấy xu hướng giảm dần về ảnh hưởng trích dẫn JNCI, với độ lệch tương đối đồng nhất giữa các nhóm Q2, Q3, và Q4 (Hình 10, Hình 11, Hình 12, Hình 13).



Hình 10. Boxplot về số bài báo (documents), số trích dẫn (*times cited*), chỉ số JNCI (*citation impact*) trong 5 năm theo chỉ số JIF

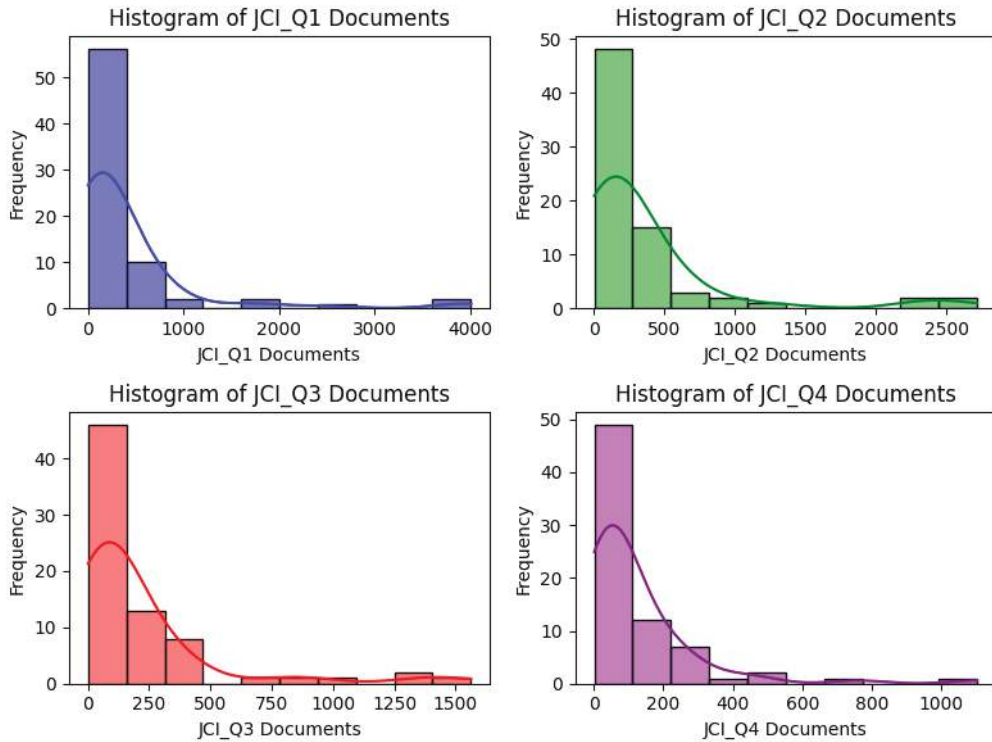


Hình 11. Biểu đồ tần suất (Histogram) chỉ số JIF (Q1, Q2, Q3, Q4) theo số bài báo trong 5 năm

Boxplot of JCI Quartile Papers by Category



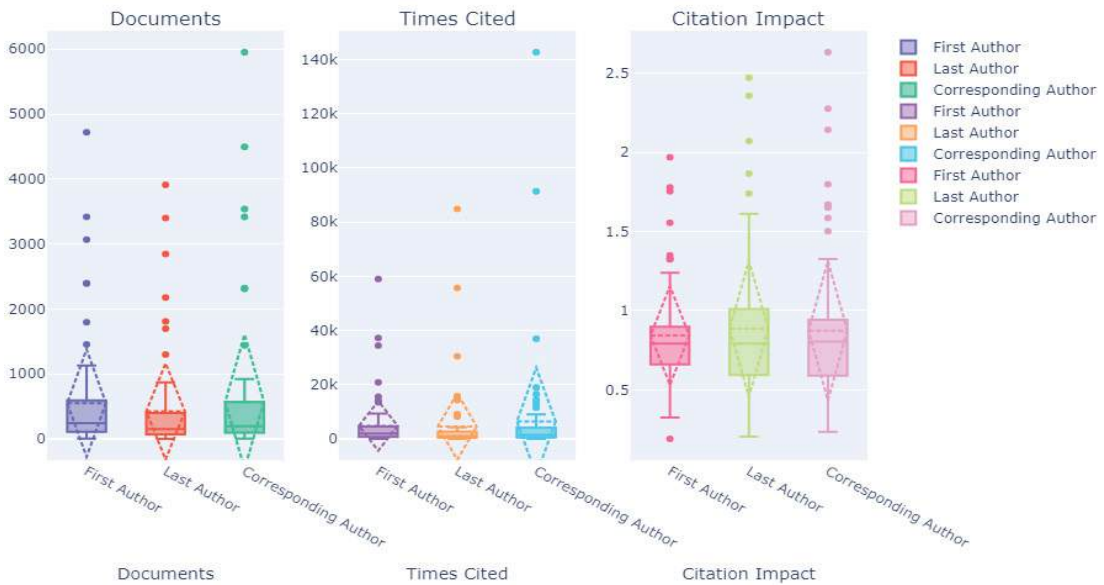
Hình 12. Boxplot về số bài báo (documents), số trích dẫn (*times cited*), chỉ số JNCI (*citation impact*) theo chỉ số JCI



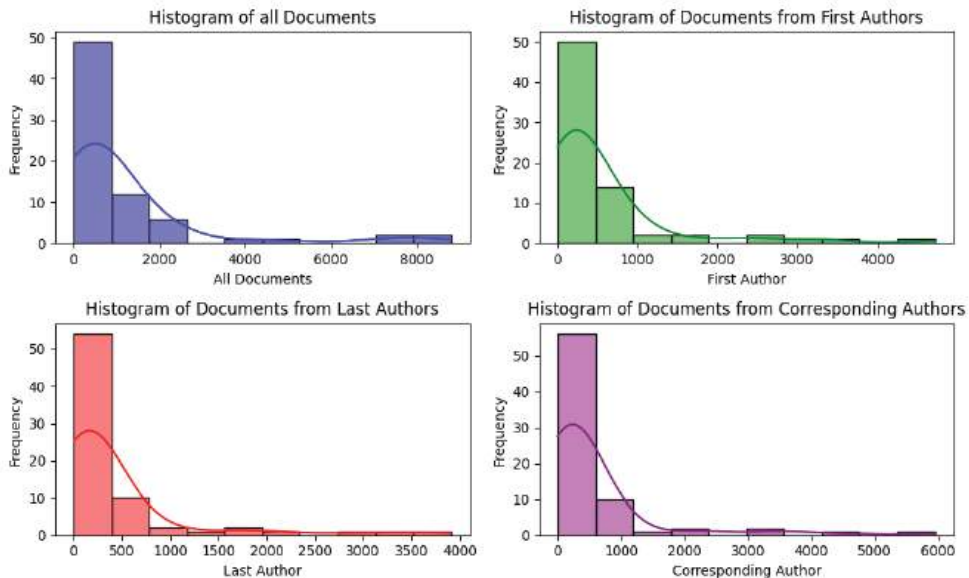
Hình 13. Biểu đồ tần suất (Histogram) chỉ số JCI (Q1, Q2, Q3, Q4) theo số bài báo trong 5 năm

Trong một công trình nghiên cứu, tác giả có thể đảm nhận các vai trò khác nhau như tác giả đầu tiên (*first author*), tác giả liên hệ (*corresponding author*), hoặc tác giả cuối cùng (*last author*). Hình 14 cung cấp cái nhìn toàn diện về hiệu suất nghiên cứu theo từng vị trí tác giả của 75 cơ sở nghiên cứu tại Việt Nam trong giai đoạn 5 năm (2019 - 2023). Kết quả cho thấy, tỷ lệ các công trình có tác giả liên hệ (37,02%; 41802/112915) và tác

giả đầu tiên (35,61%; 40208/112915) chiếm ưu thế. Mặc dù tỷ lệ tác giả tại các vị trí này cao, chỉ số ảnh hưởng trích dẫn (JNCI) cho cả ba vị trí đều cho thấy mức độ tương đồng, với giá trị JNCI trung bình là 0,84 cho tác giả đầu tiên, 0,87 cho tác giả liên hệ, và 0,89 cho tác giả không phải vị trí đầu tiên (Hình 15). Điều này cho thấy sự ảnh hưởng của các tác giả ở tất cả các vị trí trong việc thu hút trích dẫn là tương đối đồng đều.



Hình 14. Boxplot số bài báo, số trích dẫn, chỉ số JNCI trong 5 năm theo vị trí các tác giả

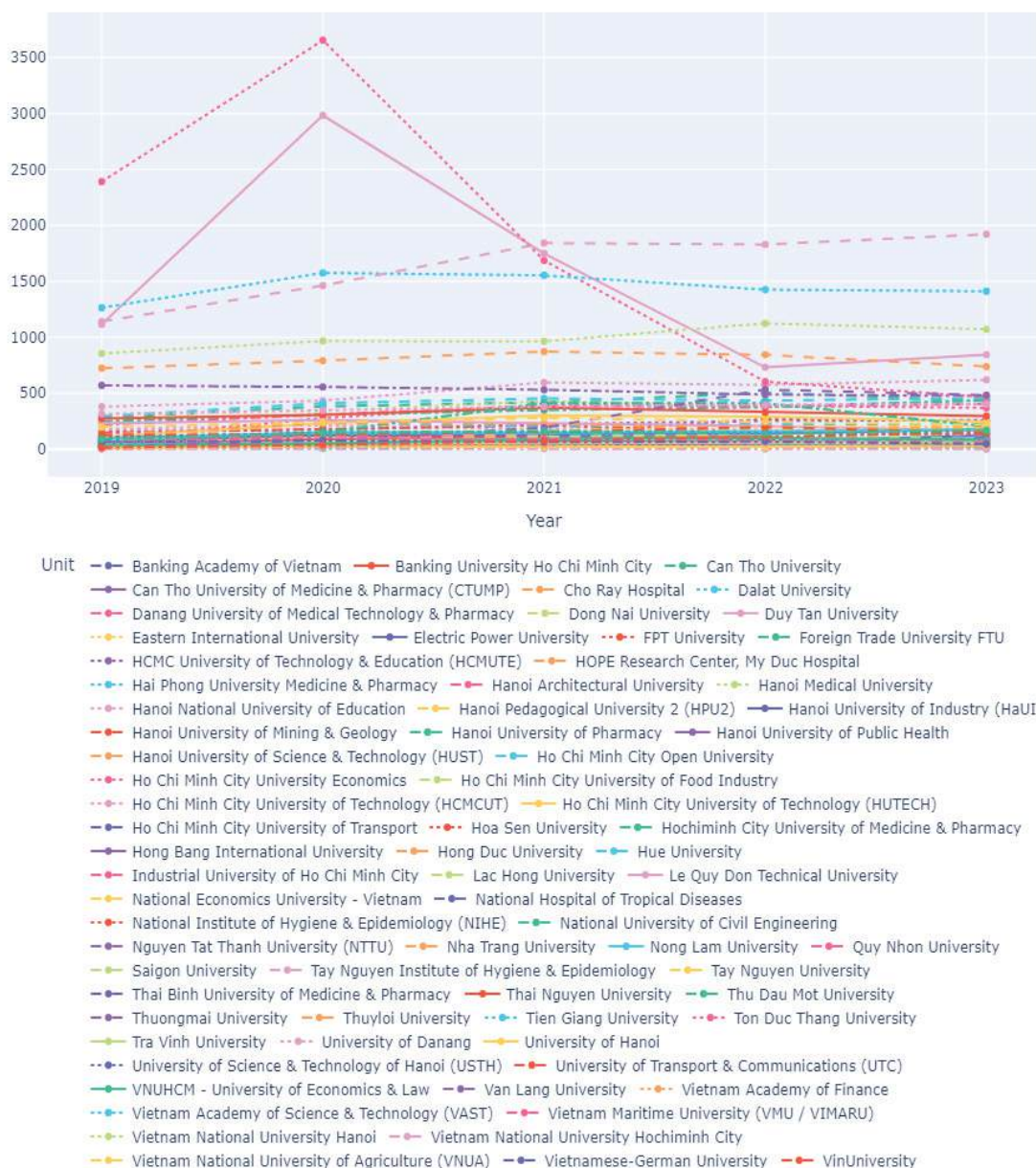


**Hình 15. Biểu đồ tần suất (Histogram) số bài báo theo vị trí tác giả trong 5 năm
Hiệu suất nghiên cứu qua các năm**

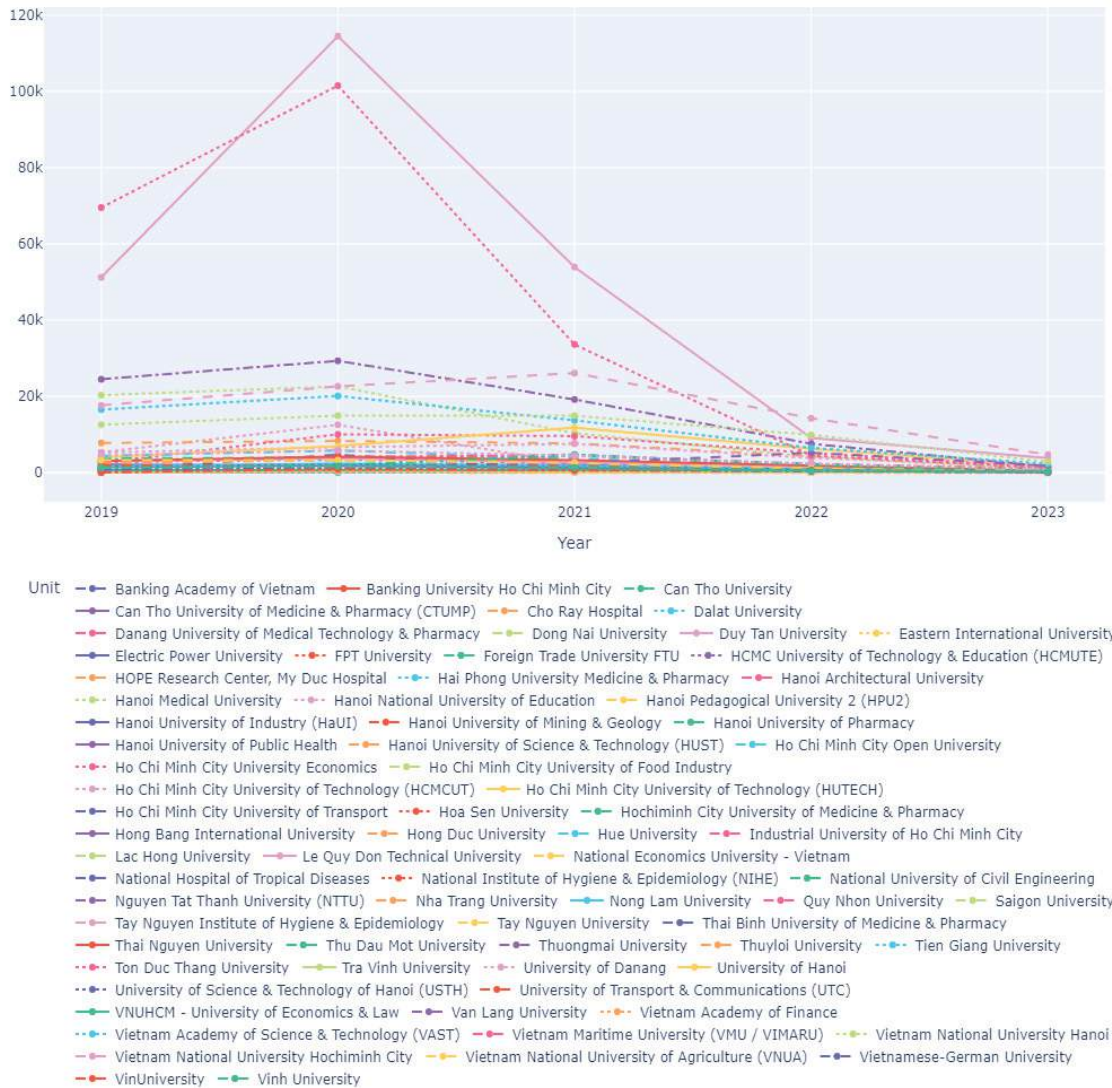
3.1.2. Hiệu suất nghiên cứu qua các năm

Dữ liệu qua các năm trong giai đoạn 2019 – 2023 của 75 cơ sở nghiên cứu được khảo sát cho thấy mỗi trường đại học có các chiến lược nghiên cứu khác nhau. Nhìn chung, số lượng công bố và số trích dẫn của các bài báo có xu hướng tương đồng giữa các trường. Đối với các bài báo trong cùng

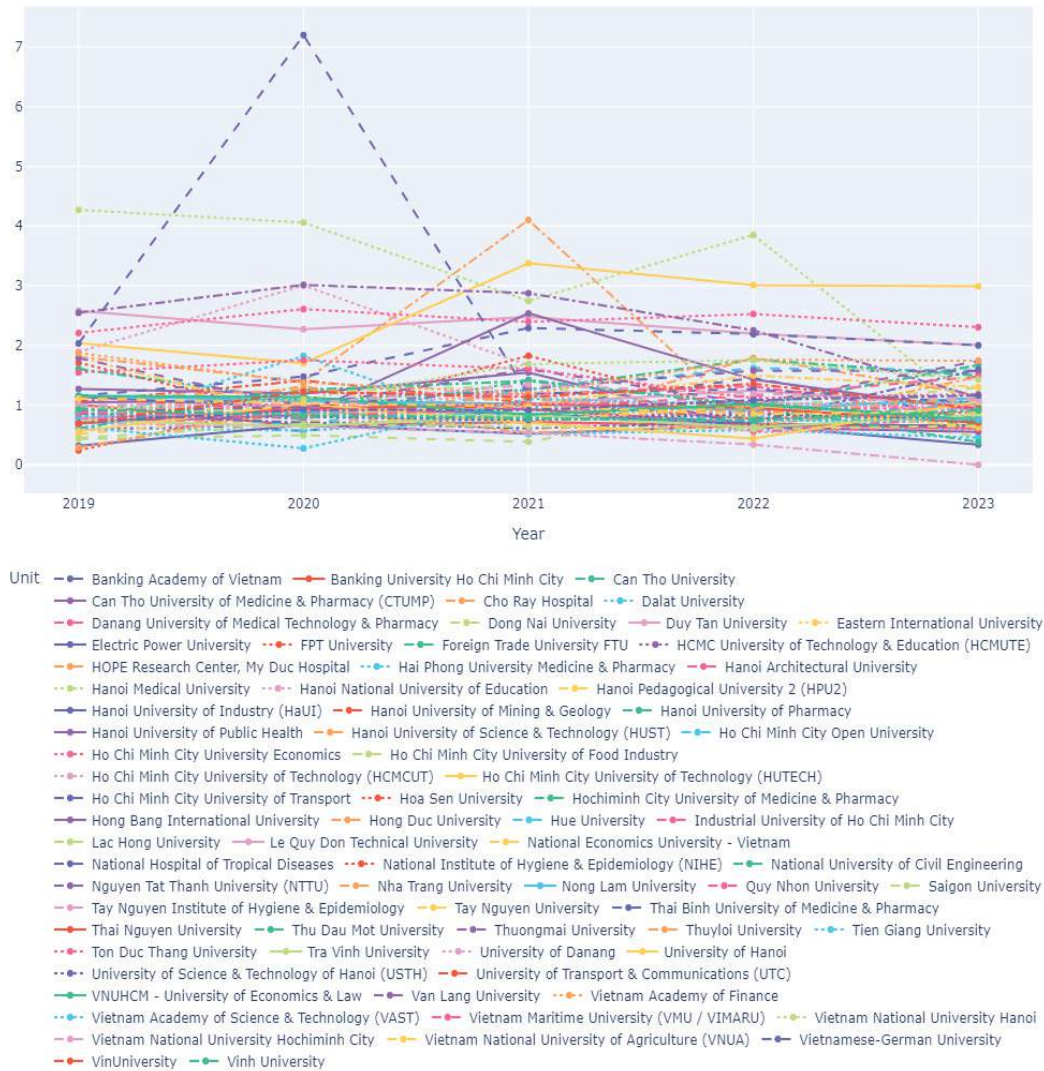
lĩnh vực, được công bố trên cùng tạp chí và tại cùng thời điểm, có chỉ số JNCI phản ánh mức độ ảnh hưởng và chất lượng của chúng so với nhau. Điều này cho thấy chất lượng và tầm ảnh hưởng của các công trình nghiên cứu phụ thuộc vào số lượng trích dẫn cũng như thể hiện mức độ quan tâm của cộng đồng khoa học (Hình 16, Hình 17, Hình 18).



Hình 16. Tổng số công trình công bố theo từng năm 2019-2023



Hình 17. Tổng số trích dẫn theo từng năm 2019-2023



Hình 18. Chỉ số JNCI theo từng năm 2019-2023

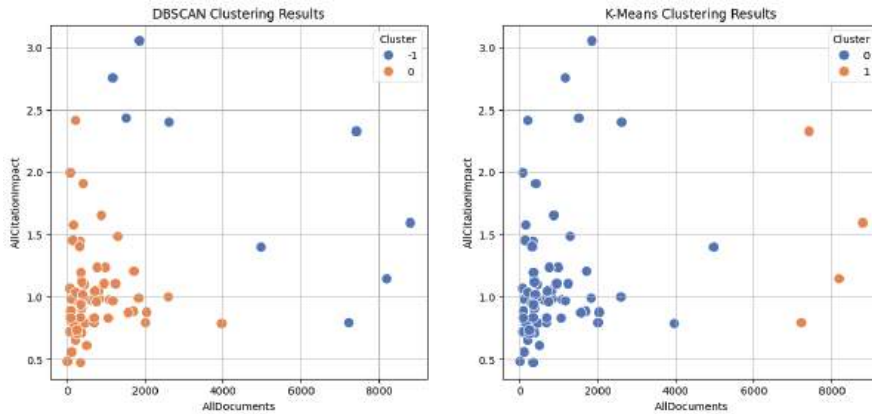
3.1.3. Phân loại cơ sở nghiên cứu dựa trên hiệu suất nghiên cứu

Dựa vào ba yếu tố gồm tổng số bài báo, tổng số trích dẫn và chỉ số JNCI trong giai đoạn 5 năm (2019 – 2023) từ các cơ sở có công trình được Incites ghi nhận, tiến hành áp dụng 02 phương pháp phân nhóm thông dụng là phân cụm không gian dựa trên mật độ

của các ứng dụng có nhiễu (*Density-Based Spatial Clustering of Applications with Noise – DBSCAN*) và K-Means. Đồng thời, thuật toán random search được kết hợp để tối ưu phương pháp tìm giá trị ϵ và min_sample cho DBSCAN và cũng như tìm cụm phù hợp cho K-Means. Kết quả các chỉ số đánh giá tại Bảng 1 như sau:

Bảng 1: Chỉ số đánh giá cho 2 phương pháp DBSCAN và K-Means

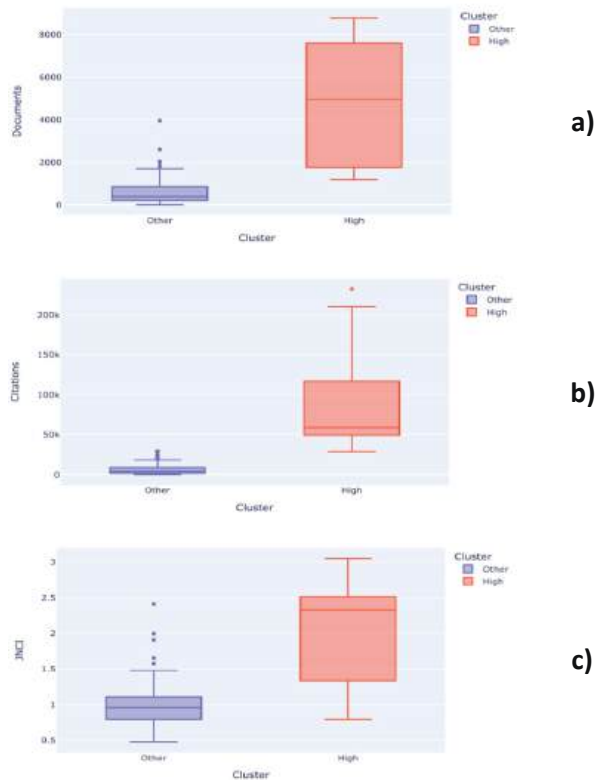
| Chỉ số | DBSCAN | K-Means | Phương án tốt nhất |
|-------------------------|---------|---------|--------------------|
| Silhouette Score | 0.7043 | 0.7512 | K-Means |
| Calinski-Harabasz Score | 70.5972 | 67.0867 | DBSCAN |
| Davies-Bouldin Score | 0.8885 | 0.5803 | K-Means |



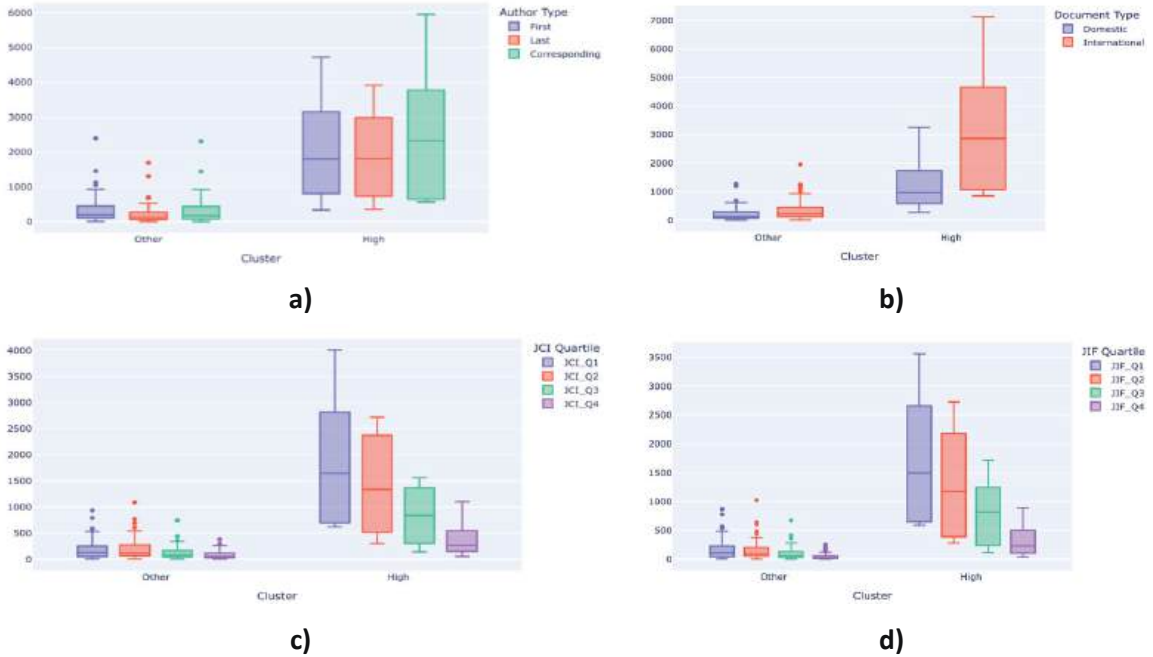
Hình 19. Phân nhóm theo phương pháp DBSCAN và K-Means

Dựa trên kết quả từ Bảng 1 và Hình 19, phương pháp DBSCAN được chọn vì khả năng linh hoạt trong phân cụm, không đồng nhất về hình dạng và dễ phân tách. Do đó trong Cluster -1 của phương pháp DBSCAN các nhóm có số lượng bài báo cao và chỉ số JNCI cao đã được gom vào một cụm để dễ dàng đối sánh với nhóm còn lại. Có 9 cơ sở nghiên cứu được phương pháp DBSCAN gom vào một cụm có giá trị cao bao gồm

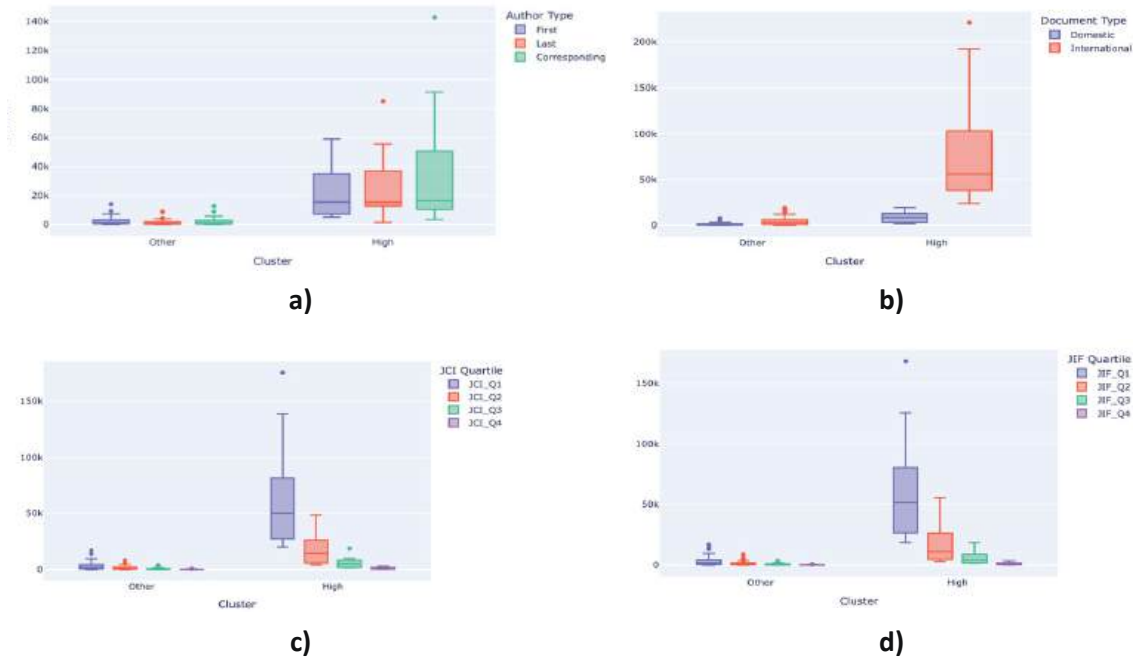
Trường Đại học Duy Tân, Trường Đại học Y Hà Nội, Đại học Kinh tế TP.HCM, Trường Đại học Công nghệ TP.HCM (HUTECH), Trường Đại học Nguyễn Tất Thành, Trường Đại học Tôn Đức Thắng, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, Đại học Quốc gia Hà Nội, và Đại học Quốc gia TP.HCM. Phân bố dữ liệu ở cả 2 nhóm được quan sát theo từng đặc trưng được thể hiện ở Hình 20, Hình 21, Hình 22, và Hình 23.



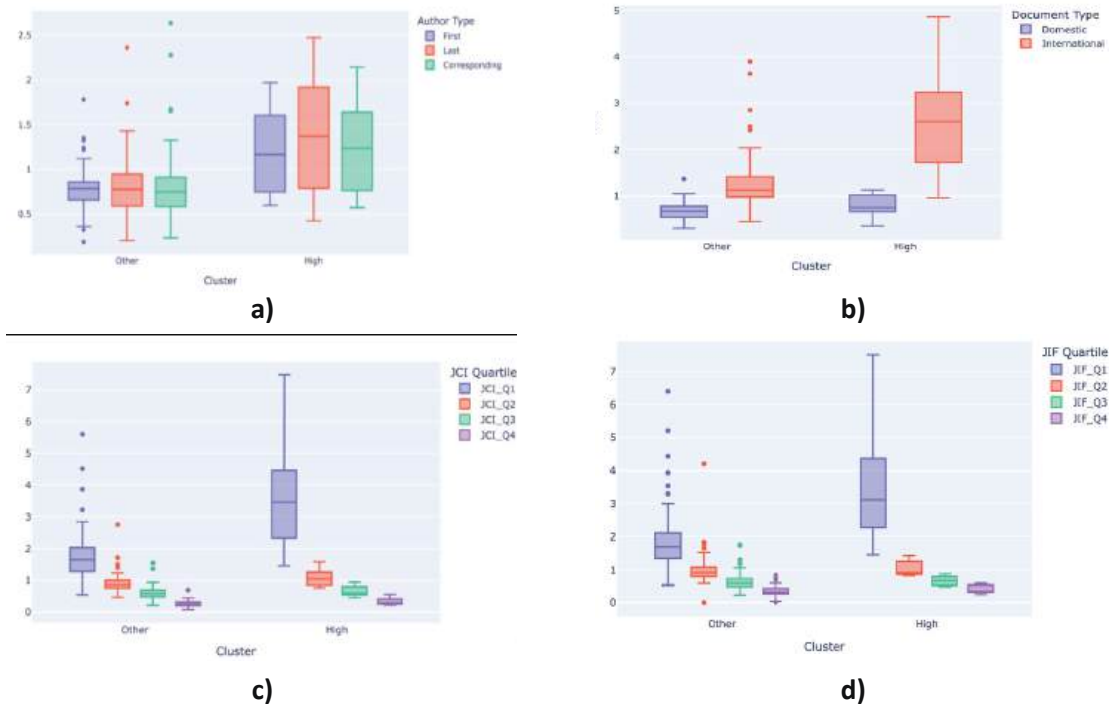
Hình 20. Boxplot tổng số bài báo (a), tổng số trích dẫn (b), chỉ số JNCI (c) giai đoạn 2019-2023 theo 2 nhóm được chia bởi phương pháp DBSCAN



Hình 21. Boxplot tổng số bài báo theo vị trí tác giả (21a) và loại bài báo (21b, c, d) theo 2 nhóm được chia bởi phương pháp DBSCAN



Hình 22. Boxplot tổng số trích dẫn theo vị trí tác giả (a), loại bài báo (b), JCI theo thứ phân vị (c), và JIF theo thứ phân vị (d) theo 2 nhóm được chia bởi phương pháp DBSCAN



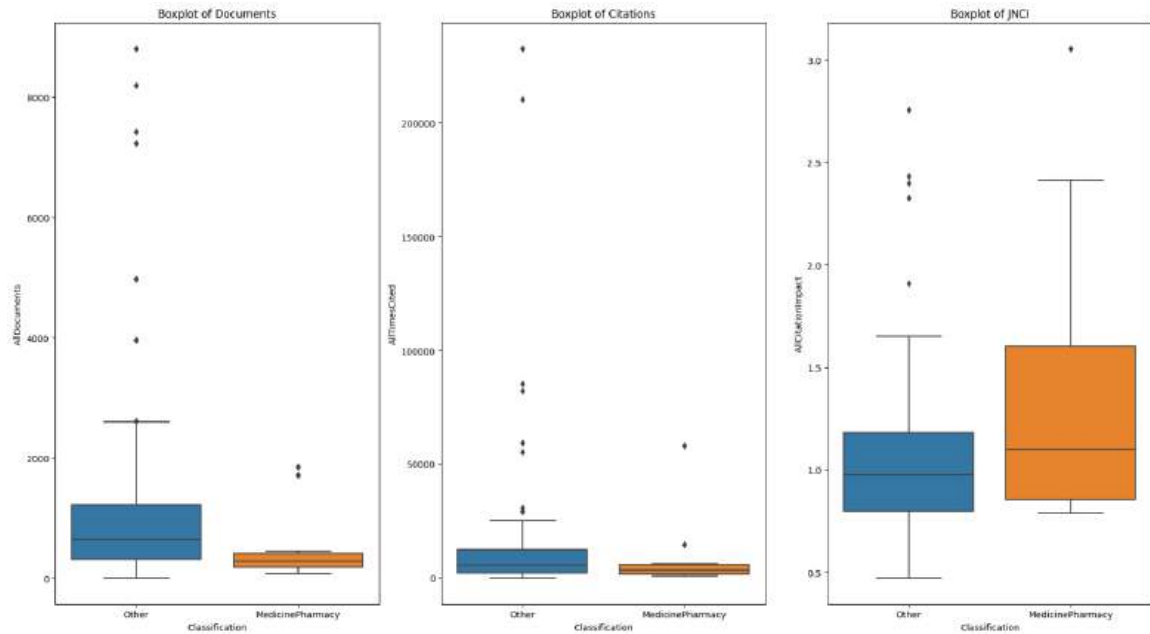
Hình 23. Boxplot chỉ số JNCI theo vị trí tác giả (a), loại bài báo (b), JCI theo thứ phân vị (c), và JIF theo thứ phân vị (d) theo 2 nhóm được chia bởi phương pháp DBSCAN

Mặc dù chỉ có 9/75 cơ sở nghiên cứu (chiếm 12%) thuộc nhóm có nghiên cứu mạnh, và tổng số bài báo của nhóm này đạt 43.780 bài, cao hơn 1,02 lần, nhưng tổng số trích dẫn của nhóm này nhận được chiếm tỷ trọng lớn hơn, cao gấp 2,22 lần so với nhóm còn lại. Về tổng số công trình công bố, nhóm cơ sở nghiên cứu mạnh có xu hướng tập trung vào các bài báo hợp tác quốc tế, được đăng trên các tạp chí uy tín, với tỷ lệ bài đăng ở các tạp chí Q1 và Q2 cao. Vị trí tác giả gửi bài trong nhóm giữ vai trò cao đáng kể, điều này cũng thể hiện vai trò chủ động trong công bố khoa học. Bên cạnh đó, số lượng trích dẫn của nhóm nghiên cứu mạnh cũng vượt trội, đặc biệt là từ các bài báo thuộc tạp chí Q1 và những công trình hợp tác quốc tế, chiếm tỷ trọng lớn trong

tổng số trích dẫn. Chỉ số JNCI của các cơ sở này cũng cao hơn qua đó phản ánh mức độ ảnh hưởng và chất lượng các công trình công bố của nhóm tốt hơn so với các cơ sở nghiên cứu còn lại.

3.2. Tình hình nghiên cứu khoa học trong lĩnh vực sức khỏe

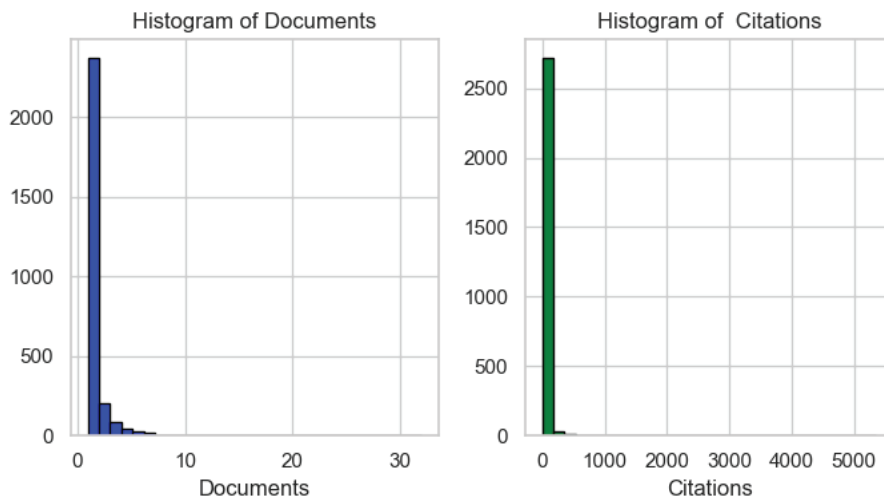
Từ kết quả phân tích 3.1.1, dữ liệu được chia thành hai nhóm là cơ sở chuyên nghiên cứu các vấn đề trong lĩnh vực khoa học sức khỏe và nhóm còn lại. Phân tích tiếp theo cho thấy kết quả thu về có nét tương đồng với kết luận trước đó. Điều này chỉ ra mức độ đóng góp của các nghiên cứu trong lĩnh vực y – dược khá lớn ngay cả khi số lượng công bố và số trích dẫn có thể thấp hơn các lĩnh vực khác (Hình 24).



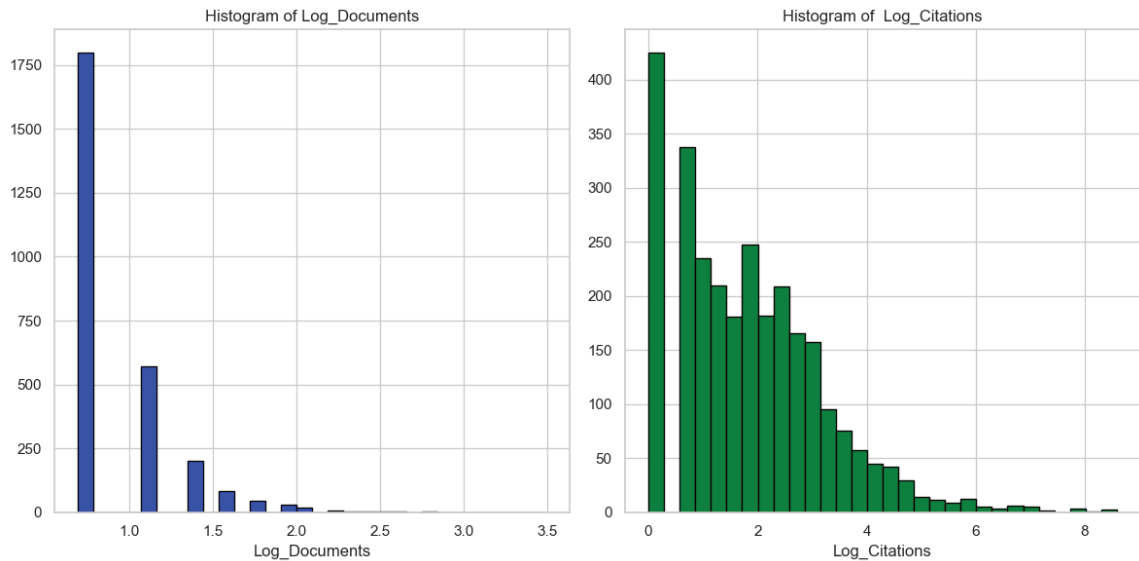
Hình 24. Boxplot số bài báo, số trích dẫn, chỉ số JNCI theo nhóm y-dược và còn lại

Ứng với mỗi đơn vị quan sát sẽ có một tập dữ liệu đặc trưng cho các chủ đề nghiên cứu của từng đơn vị đó, mục đích của điều này là hiểu rõ hướng nghiên cứu chính của từng đơn vị. Mỗi chủ đề được ghi nhận với hai chỉ số: tổng số bài báo và tổng số trích dẫn. Biểu đồ phân phối được sử dụng để thấy sự phân bố các dữ liệu này. Kết quả cho thấy dữ liệu có sự lệch phải rõ rệt, nghĩa

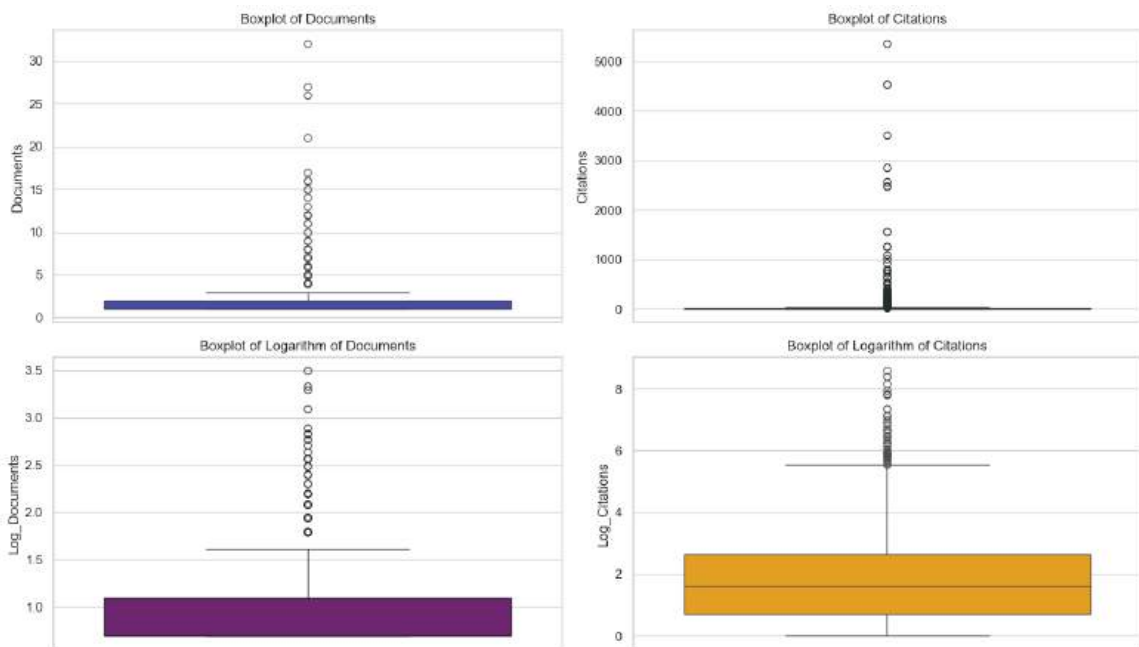
là có những chủ đề mà tổng số bài báo hoặc tổng số trích dẫn cao hơn so với các chủ đề khác, nói cách khác là một số lĩnh vực nghiên cứu có mức độ xuất bản và trích dẫn vượt trội đáng kể so với phần còn lại. Nhằm điều chỉnh dữ liệu giảm bớt lượng bên phải và đồng đều hơn, bài phân tích đã áp dụng phương pháp biến đổi logarit cho các biến quan sát (Hình 25, Hình 26, và Hình 27).



Hình 25. Biểu đồ phân phối số bài báo, số trích dẫn theo nhóm y-dược trong 5 năm



Hình 26. Biểu đồ phân phối logarit của số bài báo, số trích dẫn theo nhóm y-được trong 5 năm



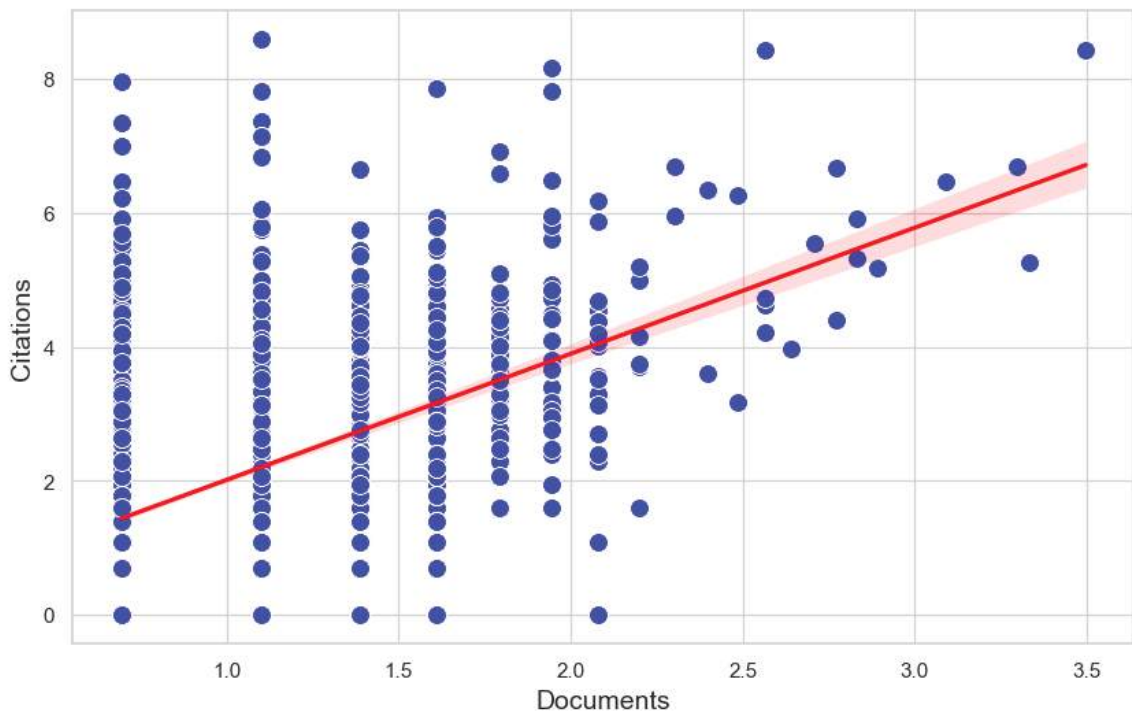
Hình 27. Boxplot không logarit và logarit của số bài báo, số trích dẫn theo nhóm y-được trong 5 năm

Hệ số tương quan Pearson là 0,49 giữa biến logarit của tổng số bài báo và tổng trích dẫn chỉ ra mối quan hệ tuyến tính giữa hai biến quan sát, tuy nhiên mức độ tương quan khá thấp nên sự đồng biến không quá mạnh. Trên lý thuyết, mặc dù số lượng bài báo của một chủ đề nghiên cứu càng nhiều

thường sẽ kèm theo số lượng trích dẫn càng tăng nhưng hệ số phản ánh tồn tại yếu tố nghịch biến giữa hai biến. Bên cạnh đó, hệ số tương quan Spearson là 0,42 khẳng định thêm xu hướng cùng chiều về thứ bậc giữa hai biến, nhưng ý nghĩa kết quả thể hiện ngoài yếu tố tuyến tính còn hàm chứa yếu tố

phi tuyến. Sự dao động của hai hệ số trong khoảng 0.42-0.49 cho thấy ngoài yếu tố chất lượng nghiên cứu, được đánh giá qua số bài

báo và trích dẫn, còn có các yếu tố khác tác động đến sự phân phối của hai biến quan sát (Hình 28).



Hình 28. Biểu đồ mối tương quan giữa logarit tần suất xuất hiện chủ đề trong bài báo và trích dẫn

Hình 29 và Hình 30 sử dụng công cụ word cloud (đám mây từ) để phân tích cụm từ được quan tâm trong số lượng bài báo và số trích dẫn được thu thập trong giai đoạn 2019 – 2023 của 11 đơn vị nghiên cứu nằm trong dữ liệu InCites, bao gồm Bệnh viện Chợ Rẫy, Bệnh viện Đa khoa Mỹ Đức, Bệnh viện Nhiệt đới trung ương, Đại học Y Dược TP.Hồ Chí Minh, Trường Đại học Dược Hà Nội, Trường Đại học Kỹ thuật Y Dược Đà Nẵng, Trường Đại học Y Dược Cần Thơ, Trường Đại học Y Dược Hải

Phòng, Trường Đại học Y Dược Thái Bình, Trường Đại học Y Hà Nội, Trường Đại học Y tế Công cộng. Dữ liệu từ word cloud cho lĩnh vực sức khỏe trong 05 năm cho thấy điểm nổi bật nhất là những công trình nghiên cứu liên quan đến virus Corona (*Coronavirus*) với tổng số 266 bài báo và 14.174 lượt trích dẫn. Điều này thể hiện rõ ràng tính thời sự của vấn đề nghiên cứu, mang tính chất cấp bách và thời sự trong giai đoạn đại dịch với đỉnh điểm nghiên cứu vào năm 2020-2022.

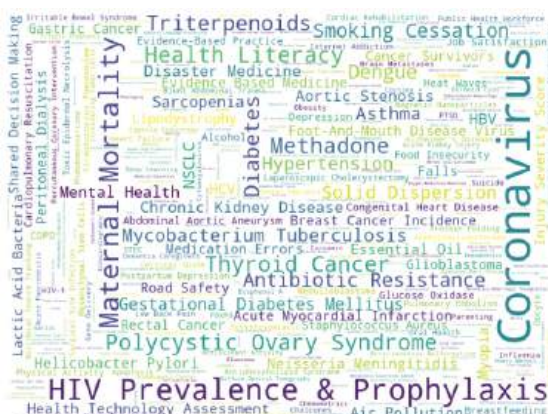
Tổng quan công trình nghiên cứu theo chuỗi thời gian cho thấy xu hướng nghiên cứu trong lĩnh vực sức khỏe của Việt Nam hiện nay (Hình 31) như sau:



a) Word cloud về số bài báo trong năm 2019



b) Word cloud về số trích dẫn trong năm 2019



c) Word cloud về số bài báo trong năm 2020



d) Word cloud về số trích dẫn trong năm 2020



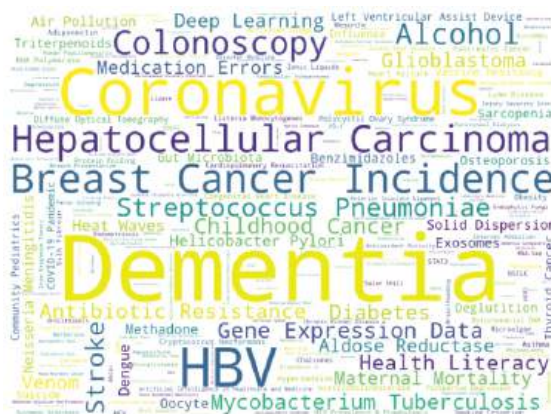
e) Word cloud về số bài báo trong năm 2021



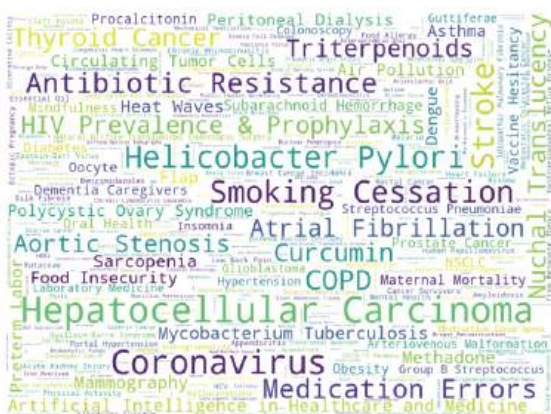
f) Word cloud về số trích dẫn trong năm 2021



e) Word cloud về số bài báo trong năm 2022



f) Word cloud về số trích dẫn trong năm 2022



g) Word cloud về số bài báo trong năm 2023



h) Word cloud về số trích dẫn trong năm 2023

Hình 31. Word cloud về số bài báo và số trích dẫn trong giai đoạn 2019 - 2023

Nhìn vào dữ liệu word cloud qua từng năm ta thấy được toàn cảnh các lĩnh vực nghiên cứu sức khỏe tại Việt Nam. Đồng thời, việc phân tích tần suất trích dẫn của các chủ đề cung cấp thông tin và xu hướng nghiên cứu đối với các bệnh phẩm trong lĩnh vực này. Đơn cử, dựa vào dữ liệu là trích dẫn nhiều nhất trong năm 2023, ta thấy sự quan tâm của cộng đồng tập trung ở nhóm vấn đề sau:

+ **Nhóm bệnh mãn tính và ung thư**, bao gồm tiểu đường (*Diabetes*); bệnh phổi tắc nghẽn mạn tính (*COPD*); ung thư phổi không tế bào nhỏ (*NSCLC*); ung thư gan (*Hepatocellular Carcinoma*); viêm loét đại

tràng (*Ulcerative Colitis*); tế bào ung thư lưu hành (*Circulating Tumor Cells*).

+ **Nhóm bệnh truyền nhiễm và vi khuẩn**, bao gồm virus Corona (*Coronavirus*); *Helicobacter Pylori* – là vi khuẩn gây nhiễm trùng dạ dày có thể dẫn đến loét dạ dày; nghiên cứu về vấn đề vi khuẩn trở nên kháng lại các loại thuốc kháng sinh thông thường (*Antibiotic Resistance*).

+ **Nhóm vấn đề về sức khỏe môi trường và yếu tố tác động**, bao gồm sóng nhiệt khi thời tiết nắng nóng kéo dài gây hại cho sức khỏe (*Heat Waves*); ô nhiễm không khí (*Air Pollution*); thiếu an ninh lương thực (*Food Insecurity*).

+ **Nhóm công nghệ và nghiên cứu**, bao gồm trí tuệ nhân tạo trong chăm sóc sức khỏe và y học (*Artificial Intelligence in Healthcare and Medicine*); Curcumin – là hợp chất trong nghệ có thể có lợi cho sức khỏe và chống viêm; tế bào sinh dục nữ cần thiết cho sự thụ thai (*Oocyte*).

+ **Các vấn đề khác**: Tình trạng dư thừa sắt (*Iron Overload*); Đột quỵ (*Stroke*).

4. KẾT LUẬN

Dữ liệu nghiên cứu khoa học tại Việt Nam giai đoạn 2019-2023 được trích xuất từ InCites cho thấy sự phân hóa rõ rệt về chiến lược của các đơn vị nghiên cứu. Chỉ có khoảng 12% cơ sở có tổng số bài báo khoa học, hoặc tổng số trích dẫn hay chỉ số JNCI cao, phần lớn các đơn vị còn lại có khối lượng nghiên cứu khiêm tốn. Nhóm các cơ sở nghiên cứu nổi bật thể hiện khoảng cách lớn so với các nhóm còn lại ở nhiều khía cạnh quan trọng. Các đơn vị nghiên cứu mạnh thường có chiến lược hợp tác quốc tế, tập trung công bố trên các tạp chí uy tín thuộc phân loại Q1 và Q2.

Đáng chú ý, mối tương quan giữa số lượng bài báo và số lượng trích dẫn có xu

hướng thuận chiều nhưng mức độ tương quan này khá yếu. Điều này cho thấy dù có thể có một số cơ sở công bố càng nhiều thì khả năng nhận được trích dẫn càng cao nhưng không phải mọi trường hợp đều tuân theo quy luật này. Một yếu tố quan trọng khác là các cơ sở có chỉ số thể hiện mức độ ảnh hưởng và chất lượng của các công trình nghiên cứu JNCI cao lại tập trung phần lớn trong lĩnh vực liên quan đến sức khỏe cộng đồng. Mặc dù các đơn vị nghiên cứu trong lĩnh vực khoa học sức khỏe có số lượng công bố hoặc số trích dẫn không quá nổi bật nhưng lại có chỉ số JNCI khá cao, điều này cho thấy tính cấp thiết, tầm quan trọng và sự quan tâm của cộng đồng của các nghiên cứu liên quan đến sức khỏe. Tiêu biểu là xu hướng này thể hiện qua sự gia tăng các công bố về chủ đề dịch bệnh, đặc biệt là COVID-19, trong giai đoạn 5 năm gần đây. Các công trình trong lĩnh vực khoa học sức khỏe không chỉ có khối lượng lớn mà còn thu hút nhiều sự chú ý từ cộng đồng khoa học, qua đó đóng góp đáng kể vào sự gia tăng về số lượng công bố cũng như trích dẫn của các cơ sở nghiên cứu.

Thông tin tác giả:

TS. Nguyễn Minh Tâm (*Tác giả liên hệ), Khoa Cơ bản - Thành viên Hội đồng Biên tập Tạp chí Khoa học Trường Đại học Phan Thiết, tỉnh Bình Thuận, Việt Nam

Email: nmtam@upt.edu.vn

BS. CKII. Nguyễn Văn Nhơn, Khoa Khoa học Sức khỏe, Trường Đại học Phan Thiết, tỉnh Bình Thuận, Việt Nam

Email: nvnhon@upt.edu.vn

BS. CKII. Nguyễn Văn Thành, Khoa Khoa học Sức khỏe, Trường Đại học Phan Thiết, tỉnh Bình Thuận, Việt Nam

Email: nvthanh@upt.edu.vn

Thông tin bài báo:

Ngày nhận bài: 04/11/2024

Ngày hoàn thiện biên tập: 26/11/2024

Ngày duyệt đăng: 30/11/2024

Ghi chú

Các tác giả đã xác nhận không có tranh chấp về lợi ích đối với bài báo này.

5. TÀI LIỆU THAM KHẢO

Bach Xuan Tran, et.al. (2022). Evolution of Interdisciplinary Approaches Among Research-Oriented Universities in Vietnam Toward a Modern Industrial Economy: Exploratory Study. *Interact Journal of Medical Research*, 11((2):e38591).

Binswanger, M. (2014). Excellence by Nonsense: The Competition for Publications in Modern Science. *Opening Science*.

Clarivate (2024). InCites Benchmarking & Analytics - Research Evaluation Tool. *Clarivate*. Truy cập từ <https://clarivate.com/academia-government/scientific-and-academic-research/research-funding-analytics/incites-benchmarking-analytics/>

DePaolo, C. A., & Wilkinson, K. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *Learning & Technology Library (LearnTechLib)*. Truy cập từ <https://www.learntechlib.org/p/167919/>

DePaolo, C. W. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *TECHTRENDS TECH TRENDS*, 58, 38–44.

Ester, M., Kriegl, H.P., Sander, J. & Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD '96)*. AAAI Press, 226–231.

Grabmeier, J. (2021). More than 87,000 scientific papers on coronavirus since pandemic. Ohio State News. Retrieved from <https://news.osu.edu/more-than-87000-scientific-papers-on-coronavirus-since-pandemic/>

InCites Help Center (2021). Journal Citation Reports Data. *InCites Benchmarking & Analytics*. Truy cập từ https://incites.zendesk.com/hc/en-gb/articles/25088350365457-Journal-Citation-Reports-Data#h_01HX-ZKW9C4HQJF4ZQZVWG03F8Q

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. Truy cập từ <https://doi.org/10.1016/j.patrec.2009.09.011>

Lương Đình Hải (2021). Năng suất công bố của các cơ sở giáo dục đại học ở Việt Nam giai đoạn 2016-2020: Dữ liệu từ Scopus Các tác giả. *Tạp chí Giáo dục*. Truy cập từ Truy vấn từ <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/88>

Martin Ester, H.-P. K. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD '96)*. AAAI Press, 226–231.

Teixeira da Silva, J. A. (2020). Publishing volumes in major databases related to Covid-19. *Scientometrics*, 126, 831-842.

McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.)*. O'Reilly Media, Inc.

McKinney, W. (n.d.). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly.

Nguyễn Minh Quân et.al. (2020). Công bố khoa học quốc tế của Việt Nam: Thực trạng và một số khuyến nghị. *Tạp chí Khoa học và Công nghệ Việt Nam*. Retrieved from <https://vjst.vn/vn/tin-tuc/3793/cong-bo-khoa-hoc-quoc-te-cua-viet-nam--thuc-trang-va-mot-so-khuyen-nghi.aspx>

Nguyen V.L., Luong, D., Pham, H. (2022) Publications on COVID-19 from Vietnam during 2020 and 2021: A bibliometric analysis. *Eur Sci Ed*. 48:e83724. Truy cập từ <https://doi.org/10.3897/ese.2022.e83724>

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*, 9(1), 12.