Topic Modeling on the Elderly Studies: Latent Dirichlet Allocation Techniques

A-Phorn Molee[1], Paiboon Manorom[1,2], and Wirapong Chansanam[1,*]

ARTICLEINFO

Article history:
Received: 9 May 2024
Revised: 5 May 2024
Accepted: 26 December 2024
Online: 15 May 2026

Keywords:
Topic modeling
The elderly studies
COVID -19
Latent dirichlet allocation

ABSTRACT

A research study on the elderly is an important tool for development, especially to support the elderly society as well as the challenges of the COVID- 19 pandemic throughout the world. This study aimed to analyze the research topic about the elderly from the Scopus data published from 2020 to 2023 (after COVID -19), in total 43,732 articles. This study employed the Topic Modeling Techniques with Latent Dirichlet Allocation (LDA) to analyze data in this study. The results indicated that the research topics regarding the elderly can be classified into six main topics: 1) Risk factors affecting the elderly, this topic has been researched the most, followed by 2) the rate of increase and death in each age group of the elderly, 3) Situation of COVID -19 infection among the elderly, 4) Diseases in the elderly, 5) Innovation and techniques for caring for the elderly, and 6) Experiments in the elderly, this topic has received slight attention from research in the past. According to the result, the topic model has a correlation value at 0.425, which was an appropriate value for grouping research topics and reflected the concept of determining the previous research topics regarding the elderly.

1. INTRODUCTION

The global landscape has been significantly shaped by the dual challenges of an aging population and the onset of the COVID-19 pandemic, which emerged in December 2019. This unprecedented viral outbreak led to a widespread health crisis, with millions across continents contracting the virus. On March 11, 2020, recognizing the disease's severity and expansive reach, the World Health Organization (WHO) officially declared the outbreak of the novel coronavirus, COVID-19, as a pandemic. This declaration was a call to action for nations worldwide, highlighting the urgent need for coordinated international efforts to combat the spread of the virus, protect vulnerable populations, and address the public health ramifications [1]. The aging population necessitates tailored innovations in products, policies, and health strategies. Jaroenwanit et al. [2] highlighted the nutritional benefits of riceberry and its appeal to elderly consumers through features like digestive relief, skin brightening, and reduced risks of diabetes and hypertension. Similarly, Jensantikul and Aimimtham [3] advocated for adaptable and comprehensive welfare policies for the elderly, addressing local and regional needs. Verma et al. [4] emphasized the utility of the Palmer Drought Severity Index for mitigating environmental impacts, showcasing the need for integrating health, welfare, and environmental strategies for aging communities. Additionally, the elderly's heightened vulnerability to infectious diseases during pandemics due to weaker immune systems and chronic conditions underscores the critical importance of implementing robust protective measures for this demographic. Effective management involves immediate actions to safeguard their health and long-term strategies to enhance the resilience of healthcare systems catering to the elderly. This dual approach requires comprehensive planning and resource allocation to ensure the elderly receive the necessary support and care during and beyond the crisis period, thus highlighting the broader challenge of fostering an age-friendly society [5].

For these reasons, the important tool that every country conducts to drive development is research. The research is the key to discovering new knowledge, analyzing or experimenting systematically with the equipment or methods to find the facts or principles in setting rules, theories, or guidelines for practice for the discovery, solution, and development [6]. Consequently, the development of research on the elderly would play an important role and provide a critical view of the present. Research on the elderly is gaining increasing importance in the context of global aging due to rapid demographic changes, resulting in unprecedented challenges such as the rise of non-communicable diseases and the increasing demand for long-term care. Furthermore, these changes

[1]*Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand.*
[2]*Faculty of Humanities and Social Sciences, Bansomdejchaopraya Rajabhat University, Bangkok, 10600, Thailand.*
**Corresponding author:* Wirapong Chansanam; Phone: +66909562727; Email: wirach@kku.ac.th.

impact income security in old age, challenging the capacity of institutions and societies to cope. Consequently, adapting public health systems is essential to address these issues [7], [8].

Research also emphasizes the necessity of reallocating resources to address the specific needs of the elderly population. The development of policies aligned with the requirements of an aging demographic is crucial in the context of contemporary population dynamics [9]. Furthermore, studies suggest that strengthening public health systems to accommodate healthcare for older adults can mitigate the burden resulting from these demographic shifts [8].

The study would lead to knowledge discovery and build new knowledge from previous studies. Additionally, these could be used for innovation towards the emerging country's problems and needs and were a guideline for allocating budget to support research efficiently, since several previous studies revealed that it seems difficult to access the knowledge in the research for further contributions. Therefore, to maximize the further studies benefits, it might be necessary to use tools to analyze this enormous amount of research data. Given the necessity of analyzing the aforementioned research documents on the elderly, the researcher has identified a widely popular method for document analysis. Each method possesses distinct advantages suitable for specific analytical characteristics and objectives. The method in focus is Latent Dirichlet Allocation (LDA): This technique employs probability distribution to extract latent topics within documents, enabling efficient content categorization [10].

Content Analysis: This qualitative analytical approach focuses on categorizing themes within documents, such as interviews and research articles, to uncover latent content [11]. Thematic Analysis: This technique identifies and analyzes themes within data, emphasizing flexibility in qualitative data analysis [12].Citation Analysis: A quantitative method that analyzes citations to measure research impact and academic trends [13]. Bibliometric Analysis: This analytical approach involves the quantitative examination of bibliographic data, such as article count, to identify academic trends [14]. Network Analysis: This method is employed to study the structure and relationships between data elements, such as citation networks in research [15].

Furthermore, the researcher's literature review on Topic Modeling in elderly-related research reveals the application and benefits of this technique in analyzing large and complex datasets. Latent Dirichlet Allocation (LDA) emerges as a preeminent technique for topic modeling in documents, particularly in research pertaining to the elderly population. Blei et al. [10] propose LDA as an efficient method for topic analysis utilizing probability distribution to extract latent words and themes within documents, thereby enhancing the identification of trends related to the health and welfare of older adults. Zhou et al. [16] employed LDA to analyze research articles pertaining to elderly healthcare, facilitating the identification of principal themes related to health issues and care for the elderly in a corpus of documents. Liu et al. [17] utilized topic modeling to examine policies and approaches in elderly care, aiding in the identification of significant themes associated with management and policy in geriatric care. Blei and Lafferty [18] demonstrate the application of LDA and related topic techniques in analyzing documents and data across various disciplines, showcasing LDA's capacity to manage complex and large-scale datasets. Consequently, Topic Modeling using LDA plays a crucial role in analyzing research documents pertaining to the elderly, enabling efficient organization of data and identification of key topics. This facilitates the development and refinement of policies or guidelines for elderly care. Moreover, it serves as an essential tool in analyzing and identifying primary themes in gerontological research, enhancing our understanding of significant trends and issues in this field.

Topic modeling is the feature of unsupervised machine learning, which has the ability to examine a set of documents, identify words and reveal patterns within the documents and group words and figurative expressions for description with the nature of a corpus. In brief, a topic modeling algorithm has created a collection of phrases and words to help readers understand what those relationships mean based on topic classification [19], [20]. The topic modeling is an innovative text-mining statistics technique for revealing hidden semantic structures in a huge document. The topic modeling with the Latent Dirichlet Allocation (LDA) method is very popular in machine learning and NLP [21], including LDA, which has been applied in modeling topics in Scientific research, Bioinformatics, Social network analysis and software engineering [6], [22]. Additionally, the LDA was often used in research according to the objective of corpus analysis [10], [23]. With these useful contributions for document analysis, the analysis would increase the capacity to explore hidden topics with flexible and effective methods [24]. To reach the research objectives; this study analyzed research articles on the elderly to examine research trends on the elderly after the COVID-19 pandemic, which was a critical period for the elderly group and an interesting period to follow research around the world.

2. MATERIALS AND METHODS

This research study aimed to analyze the data using topic modeling methods by creating a model from research data on the elderly in the Scopus database from 2020 to 2023.

The research methodology was as follows:

1. Data collection: selected research articles for data analysis by searching research articles about the elderly from the Scopus database [https://www.scopus.com] from

2020 to 2023 (June). Selected only articles that have been published in English and downloaded the data on June 5th, 2023. The search results found about 44,227 articles, and then we checked and eliminated duplicate and incomplete articles. The final research included about 43,732 articles and listed data with a vocabulary size of 76,897 words. The data was saved and exported in CSV format for data import in the analysis process.

2. Data preparation and data cleaning: As the data is in English text format, we used Natural Language Processing (NLP) in Python and Text pre-processing by PyCaret, which is an open source and reliable tool [25] for removing marks and punctuation, removing numbers, and removing unnecessary words. This process was conducted as a common data-cleaning process to extract keywords that are representative of research articles using the bigram algorithm to select common organizing and TF-IDF algorithm to extract keywords from the abstract [26], [27].

This study employed specific parameters within the Latent Dirichlet Allocation (LDA) model to ensure optimal performance and interpretability. The model identified six primary topics (k = 6), a choice guided by achieving the highest coherence score of 0.4254, indicating strong semantic similarity within each topic. The dataset was split into training (70%) and test (30%) sets to enhance model robustness. Coherence scores were the main criterion for evaluating topic models, ensuring selecting the most interpretable model. A bigram analysis was utilized to refine further topic identification, which proved more effective than unigram or trigram methods in capturing word relationships. Interactive visualizations using pyLDAvis provided insights into the prominence and interrelationships of topics within the corpus. These methodological choices contribute to the model's transparency, reproducibility, and overall utility in analyzing research trends on the elderly, demonstrating a rigorous approach to topic modeling.

3. Topic modeling: Topic modeling refers to a tool for discovering hidden knowledge structures in documents to make decisions using data and gain insights into complex topics [10]. To conduct the topic modeling for analysis, the data must be divided into a training data set for 70% and a test data set for 30 % and classified using a model to predict. In addition, we have employed the LDA for descriptive topic learning, but it seems to have higher efficiency than semantic representation in the corpus [28]. This research employs a research process based on the conceptual framework of Latent Dirichlet Allocation as proposed by Blei [29], as illustrated in Figure 1.

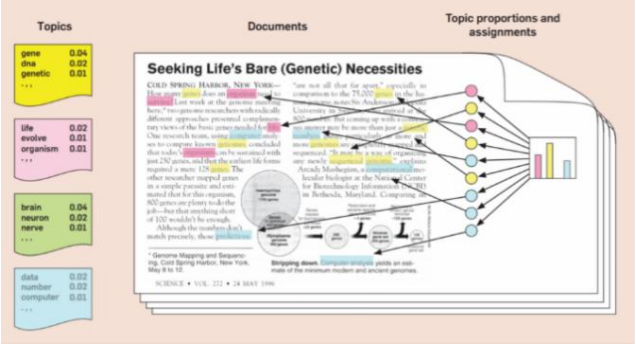

Fig. 1. The conceptual framework of Latent Dirichlet Allocation (LDA) in topic discovery within articles [29].

3. RESULTS

3.1. Text analysis

The study utilized Latent Dirichlet Allocation (LDA) to analyze research abstracts, focusing on data analysis. By applying LDA, the research identified six primary topics within the abstracts, labeled from Topic 0 to 5. The outcomes of the LDA model were quantitatively evaluated, and values were assigned to each topic. These values helped determine the prominence or dominance of each topic. The topic that received the highest value from the model was considered the primary topic of the research abstracts. This main topic is likely the most prevalent or significant theme discussed across the analyzed abstracts. The results, including the ranking and values of the topics, were visually represented in Table 1, which illustrates the distribution and importance of each topic within the dataset. To fully comprehend Table 1, it is essential to carefully examine the columns presenting detailed information on article titles, publication years, research domains, and abstracts. As illustrated in the figure, pay particular attention to the distribution of topics across various research documents. The corresponding topic weights assigned to each document provide valuable insights into their relevance to the identified themes. This analytical approach facilitates the identification of prevalent themes and their relative significance within the broader research context, thereby contributing to a deeper understanding of the patterns and trends underlying the examined body of literature.

The research document dataset underwent a detailed topic distribution analysis to identify underlying patterns and trends. This analysis was pivotal in understanding the prevalence of various topics within the dataset. The study's findings revealed that the topics were unevenly distributed across the documents, with Topic 2 emerging as the most frequently occurring topic. This was followed by Topics 4, 5, 3, and 0 in descending order of frequency. In contrast, Topic 1 was observed to have the lowest frequency among all the topics. The significant variations in topic frequencies could be influenced by several factors related to the dataset's compilation, such as the specific objectives of the research

or the methods used for sample size determination. These factors are crucial as they could skew the representation of topics within the dataset. Figure 2 graphically represents these findings to provide a clearer visualization. This figure illustrates the frequency distribution of each topic, offering a visual interpretation that complements the textual analysis. While the initial results are indicative, the study recognizes the need for further data collection. Additional data is crucial to validate these findings and ensure that the observed distribution is not a result of methodological biases but a true reflection of broader research trends. Therefore, these insights into the dataset's topic distribution are preliminary and require further investigation to be solidified.

Table 1. A vector of topic probability

Authors	Title	Year	Source Title	Cited by	Abstract	Dominant_ Topic	Perc_Dominant_ Topic
Devleesschauwer B.; Willem L.; ...	The direct disease burden of COVID-19 in ...	2023	BMC Public Health	0	Burden disease estimate become important ...	Topic 4	0.85
Mirzaeian R.; Nopour R.; Asghari Varzaneh Z.; ...	Which are best for successful aging prediction...	2023	BioMedical Engineering Online	0	Worldwide society currently faces epidemiologic ...	Topic 5	0.51
Csuka S.I.; Dévényi J.; Konkoly Thege B.; ...	Relationship satisfaction and self-esteem in p...	2023	BMC Women's Health	0	Breast cancer patient partner support...	Topic 5	0.78
Rios S.; García-Gavilán J.F.; Babio N.; ...	Plasma metabolite profiles associated with ...	2023	Cardiovascular Diabetology	0	Healthy lifestyle inversely relates to type 2 diabetes...	Topic 4	0.47
Maieron A.; Duller C.; Püspök A.; ...	SASE, Success and Adverse event Score in Endo...	2023	BMC Gastroenterology	0	Validate accept grade tool preprocedural compli...	Topic 2	0.39
Lefas A.; Bodagh P.N.; Pan J.; Vazir A.; ...	Cardiac tamponade from anticoagulant-related ...	2020	BMJ Case Reports	0	Describe case year-old man background severe ...	Topic 2	0.46
Chari A.; Samur M.K.; Martinez-Lopez J.; Cook ...	Clinical features associated with COVID-19...	2020	Blood	133	Primary cause morbidity mortality patient ...	Topic 4	0.57
Zhuang L.; Ni H.; Wang J.; Liu X.; Lin Y.; Su ...	Aggregation of Vascular Risk Factors Modulates...	2020	Frontiers in Aging Neuroscience	13	Several vascular risk factors include hypertens...	Topic 1	0.5
Hyon R.; Youm Y.; Kim J.; Chey J.; Kwak S.; ...	Similarity in functional brain connectivity at...	2020	Proceedings of the National Academy of Science...	25	People often intuit similar friend evidence...	Topic 5	0.55
Short N.J.; Rafei H.; Daver N.; Hwang N.; Ning ...	Prognostic impact of complete remission with ...	2020	Blood Advances	23	Relapsed_refractory_acute_ myeloid leukemia...	Topic 2	0.48

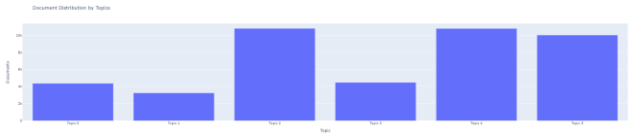

Fig. 2. The visualization of topic distribution.

3.2. Topic modeling

3.2.1. Topic model evaluation

The evaluation of topic models used a method focused on topic coherence. Topic coherence is assessed by measuring the semantic similarity among the words that score highly within a topic model based on the corpus data referenced [30]. This approach is currently regarded as the most effective for comparing different topic models regarding human interpretability, with additional support from references [31] and [32]. In the context of this method, Figure 3 presents coherence scores for models that represent varying numbers of topics. The optimal model is identified as achieving the highest coherence score before any subsequent decline in performance, as noted in [33]. The results specifically highlighted that the model with topics set to six (k = 6) yielded the highest coherence score of 0.4254, indicating it as the most suitable model according to the study. This suggests that setting the number of topics to six leads to the most semantically coherent and interpretable results.

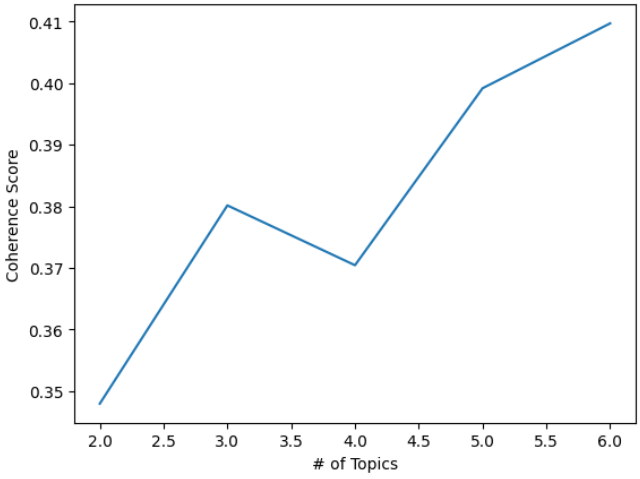

Fig. 3. Topic number and coherence.

3.2.2. Topic naming and topic details

In this process, a bigram analysis method was used to predict the most frequent word clusters associated with the identified six primary topics. A bigram, which considers pairs of consecutive words, was chosen because it provided higher performance metrics in capturing the structure of word relationships within the data than unigram (single word) and trigram (sequences of three words) approaches. This analysis revealed the 100 most common word clusters, which were systematically categorized into six main topics, labeled as Topic 0 through Topic 5. These topics are depicted in Figures 4 through 9. To understand the bigrams, examine the frequency distribution of bigrams (pairs of consecutive words) across the dataset. The taller bars on the left represent the most common bigrams, indicating key phrases or themes. The less frequent bigrams provide insights into less dominant topics as you move right. Each topic was assigned a representative name based on the dominant thematic elements within the clusters, with these names and corresponding explanations presented in Table 2. This structured approach enabled a detailed and organized presentation of the data, facilitating a clearer understanding of the prevalent themes within the bibliometric analysis.

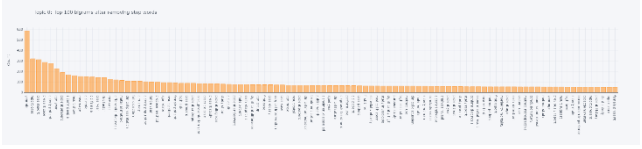

Fig. 4. Topic 0 -Top 100 bigrams.

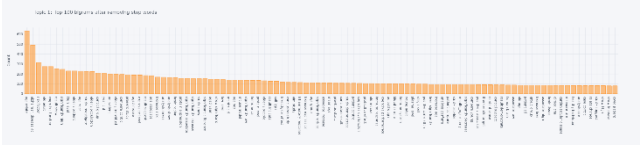

Fig. 5. Topic 1 -Top 100 bigrams.

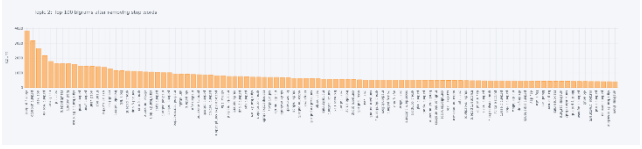

Fig. 6. Topic 2 -Top 100 bigrams.

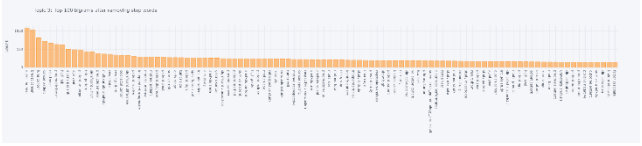

Fig. 7. Topic 3 -Top 100 bigrams.

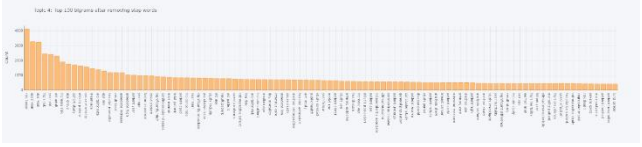

Fig. 8. Topic 4 -Top 100 bigrams.

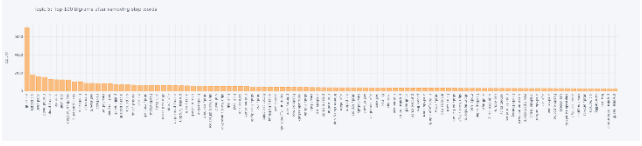

Fig. 9. Topic 5 -Top 100 bigrams.

Table 2. Setting a topic name

Topic no.	Topic name	Representative bigrams
0	Techniques and innovations for caring for the elderly	fall detection, material method, rare case, case year, old male, result show, patient present, elderly patient, case report, year old
1	Experimental study on the elderly	aged mouse, compare control, ad patient, control group, study, elderly patient, significantly high, function, risk factor, age relate, cognitive
2	Risk factors affecting the elderly	control group, high risk, result total, mean age, patient year, significant difference, result patient, risk factor, age year, elderly patient
3	Diseases in the elderly	patient receive, colorectal, cancer, result total, prediction, model, total patient, age year, high risk, risk factor, lung cancer, overall survival
4	Increasing / death rates, causes and risks for each age group of the elderly.	patient year, associate increase, hospital mortality, mortality rate, cause mortality, cohort study, increase, risk, year old, age year, risk factor
5	COVID -19 infection among the elderly	control group, elderly patient, social support, risk factor, mean age, COVID pandemic, age year, elderly people, aged year, old adult

3.2.3. Topic visualization

The interactive visual diagrams were considered a highly effective way to present topic model results. In this study, the pyLDAVis [34] was adapted to create an interactive diagram to illustrate the topics and the representative words. Each circle in the diagram indicated how related the topic is in the corpus, and topics that were close together would represent the similarity. The pyLDAvis visualization allowed users to adjust the relevance of words in a topic through the program slider [34], [35]. Therefore, the pyLDAVis package was considered one of the creative and powerful ways to visualize and interpret topic models to help the researchers gain insight into the structure and content. This tool presented a clear and easy-to-use picture of the relationships and strengths of each topic, which illustrated the words in each topic with a circle and horizontal bar chart. The left -circle contains a panel for showing an overall view of the user model-friendly use and easily understanding the relationships in the presentation. The horizontal bar chart shows the words in each topic for detail and understanding of the topic and allows the user to receive a quick view of summary topics through the most important keywords, which can be categorized into six topics, as in Table 2.

In Figure 10, we illustrate the methodology used for naming and detailing topics in our study. Specifically, we utilized a bigram approach, which involves the use of two consecutive words, to predict and identify the most common word clusters within each topic. This bigram method proved to be more effective than both the unigram (single word) and trigram (three consecutive words) approaches, particularly because it captured overlapping structural relationships more efficiently in our bibliometric analysis. As a result of this methodology, we were able to distinguish the 100 most prevalent word clusters, which were organized into six primary topics, labeled from Topic 0 to Topic 5. Each of these topics was further explored and detailed in Figures 4 through 9. Additionally, we assigned representative names to these topics based on their predominant thematic elements, which are listed and detailed in Table 1. This structured approach allowed for a clear and organized presentation of the data, facilitating a better understanding of the key themes and patterns that emerged from our analysis.

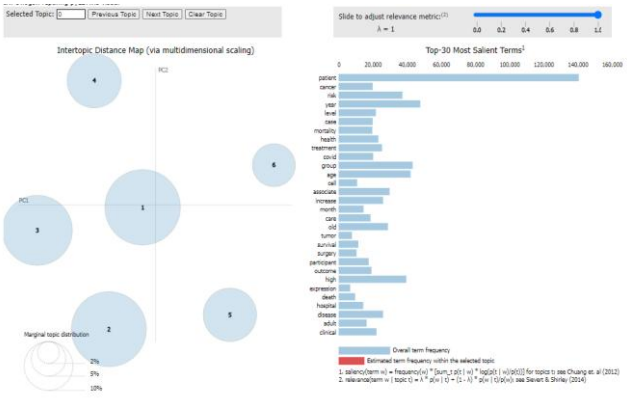

Fig. 10. The Interactive visualization of LDA model.

This section discusses the application of the t-distributed Stochastic Neighbor Embedding (t-SNE) technique for visualizing document classifications in research. The t-SNE method is utilized here to represent each category of topics as a probability distribution. This approach is particularly effective for handling and representing data in a high-dimensional space. By transforming these high-dimensional vectors into three-dimensional projections, t-SNE facilitates a visual inspection of how similar documents cluster together based on their content. The relevance of the topics and categories to contemporary research is underscored by the keywords linked with each topic, suggesting that these categories align well with prevalent research interests. Using t-SNE for 3D projections allows researchers to visually explore and analyze the relationships and distances between these categories. Such visualizations help understand how different documents are related or grouped, enhancing the clarity and insights into the dataset. Figure 11, presumably in a related publication or report, showcases these clusters,

providing a graphical representation of the similarity between multidimensional document vectors. This visual aid is crucial for researchers to grasp complex relationships and patterns in document classification intuitively.

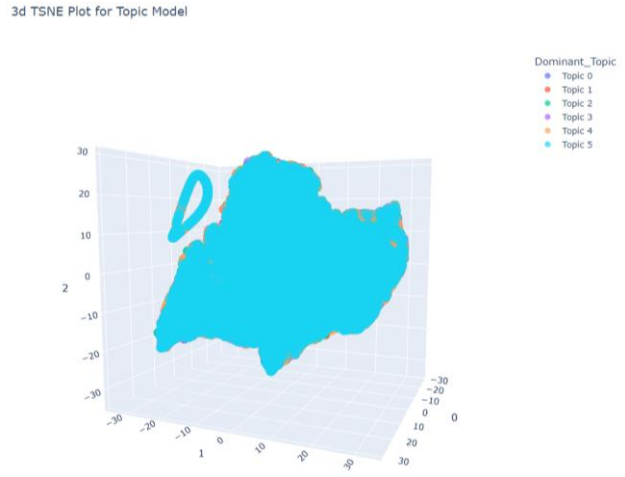

Fig. 11. The visualization of the t-SNE.

4. DISCUSSION

The data analysis of the elderly studies through abstracts with the PyCaret library in Python was used to create topic models and evaluate the performance of the model, which represents a set of words relevant to research related to the study. Employing the LDA technique on extant research documents pertaining to the elderly population, this study aims to elucidate the current body of knowledge and identify critical areas for future gerontological research. Furthermore, the findings of this research may prove beneficial in informing the development of significant research frameworks, thus promoting and supporting the advancement of national research initiatives in the field of gerontology for relevant institutions.

Text mining was considered a quantitative method that uses statistical methods to analyze published works, which related to big data analysis and was widely popular [36], [37]. Apart from this, the text analysis was one of the important tools used for the topic modeling to helps researchers identify hidden patterns and relationships between texts [38]. A topic model commonly used to analyze research documents with the Latent Dirichlet allocation (LDA) technique [10], [29], [39]. As such, this study aimed to explore the trends in research on the elderly after the COVID-19 pandemic, which is a critical period for the elderly group and an interesting period to follow the research on the elderly around the world by analyzing research topic models from Scopus from 2020 – 2023. Therefore, a total of 43,732 issues were analyzed through the topic modeling based on the Latent Dirichlet Allocation (LDA) technique. The results indicated that the six main topics were identified and able to be applied in the development of research on the elderly studies, particularly these topics: a) Risk factors affecting the elderly as the most research studies, followed by b) Increase/Death rate causes and risks for the elderly, c) Situation regarding COVID-19 infection in the elderly, d) Diseases in the elderly, and f) techniques and g) Innovations in caring for the elderly : This study employed Latent Dirichlet Allocation (LDA) to analyze topic distribution in research documents pertaining to the elderly population. The analysis revealed that Topic 2 exhibited the highest frequency, followed by Topics 4, 5, 3, and 0. These findings demonstrate the efficacy of LDA in categorizing topics within large-scale document corpora, corroborating the results of Blei et al. [10]. Similarly, the study by Griffiths and Steyvers [40] affirmed the effectiveness of the LDA technique in topic clustering within extensive document sets. Furthermore, the research of Saeed and Fatemeh [30] substantiated LDA's capacity to efficiently uncover latent trends in big data.

The topic coherence analysis in this study indicated that setting the number of topics to six (k = 6) yielded optimal results, achieving the highest coherence score of 0.4254. This finding aligns with the studies of Ray, Ahmad, and Kumar [31], and Qiao and Williams [32], which confirmed the reliability of LDA's topic coherence analysis methodology.

Furthermore, this study demonstrates that LDA is an invaluable research tool for tracking trends in gerontological research. It differs from previous research methodologies that may not have comprehensively addressed or deeply explored rapidly changing trends and emerging societal needs in the post-pandemic era [41], [10]. While traditional research often employed qualitative or conventional statistical analysis methods, LDA offers a more quantitative and efficient approach to identifying and uncovering latent "topics" within extensive research documents. This method enables research to more clearly identify risk factors and challenges associated with the elderly population, such as mortality causes, COVID-19 infection scenarios, and innovations in elderly care. These results can be applied to develop crucial policies and research frameworks for relevant institutions [10], [40]. Thus, LDA not only strengthens research but also serves as a tool that leads to the development of efficient research frameworks that concretely align with national development needs in the field of gerontological research [41]. Nonetheless, this study also has some limitations with the data analysis from the research abstract. As a result, it may be difficult to identify urgent issues in the context of each country. Future studies should explore research in various fields more clearly in order to get in-depth and comprehensive results in the context of development. In addition, the further benefit from this study might be used for other information sources apart from Scopus for all-inclusive results.

5. CONCLUSIONS

The analysis using the PyCaret library in Python to create topic models from research abstracts on the elderly has provided significant insights into the current state and direction of research within this demographic group. The study utilized the Latent Dirichlet Allocation (LDA) technique to identify six main research topics from a substantial corpus of 43,732 documents indexed in Scopus from 2020 to 2023. These topics, which include risk factors affecting the elderly, mortality rates, COVID-19 impacts, prevalent diseases, and innovations in elderly care, reflect the breadth of issues currently being explored in the field.

The results underscore the critical need for focused research on the elderly, especially in the context of the post-COVID-19 era, which has been particularly challenging for this group. However, the study also acknowledges certain limitations, primarily the restrictive scope of data derived solely from abstracts, which may not capture the urgent, country-specific issues affecting the elderly. This limitation suggests a need for future research to encompass a broader range of data sources and to dive deeper into localized contexts to enhance the comprehensiveness and applicability of the findings.

In conclusion, this study not only highlights the primary areas of current research on the elderly but also sets the stage for future investigations that could provide more detailed insights into the needs and challenges faced by the elderly population. By expanding the sources of information and increasing the granularity of the data analyzed, subsequent studies could greatly enhance our understanding and ability to effectively support the aging population, aligning with national development goals. This study emphasizes the need for increased promotion of gerontological research to comprehensively address the development and accommodation of an aging society in all aspects.

The findings of this study should serve as a guideline for establishing national research policy frameworks, focusing on research topics in the following order. Topic 2: Risk factors affecting the elderly (control group, high risk, result total, mean age, patient year, significant difference, result patient, risk factor, age year, elderly patient), Topics 4: Increasing / death rates, causes and risks for each age group of the elderly. (patient year, associate increase, hospital mortality, mortality rate, cause mortality, cohort study, increase, risk, year old, age year, risk factor), Topic 5: COVID -19 infection among the elderly (control group, elderly patient, social support, risk factor, mean age, COVID pandemic, age year, elderly people, aged year, old adult), Topic 3: Diseases in the elderly (patient receive, colorectal, cancer, result total, prediction, model, total patient, age year, high risk, risk factor, lung cancer, overall survival), Topic 0: Techniques and innovations for caring for the elderly (fall detection, material method, rare case, case year, old male, result show, patient present, elderly patient, case report, year old), and Topic 1: Experimental study on the elderly (aged mouse, compare control, ad patient, control group, study, elderly patient, significantly high, function, risk factor, age relate, cognitive). Furthermore, the findings of this study can be utilized to identify urgent research topics and areas of research deficiency crucial for national development in preparation for an aging society. When research comprehensively covers all relevant issues, it provides accurate and precise tools for national development, leading to efficient progress, economic stability, and improved well-being of the population.

ACKNOWLEDGEMENTS

This work has received a scholarship under the Research Assistant Program from Khon Kaen University, Thailand (Grant No. RA2566-12). The authors gratefully acknowledge the financial cooperation of the Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand.

REFERENCES

[1] T. H. Farzad and A. Hassan, "The powerful immune system against powerful COVID-19: A hypothesis," Medical Hypotheses, vol. 140, 109762, 2020.

[2] P. Jaroenwanit, S. Deeboonmee, and W. Thongthip, "The outstanding features of innovative riceberry product influence on purchasing intention of elderly consumers in Thailand," GMSARN International Journal, vol. 15, no. 1, pp. 21–26, 2021.

[3] N. Jensantikul and S. Aimimtham, "Governmental practice guidelines for providing appropriate social welfare for the elderly in Thailand," GMSARN International Journal, vol. 17, no. 2, pp. 163–169, 2023.

[4] A. Verma, H. Verma, and R. K. Saket, "Elderly fall detection for smart home caring using YOLOv8," GMSARN International Journal, vol. 19, no. 2, pp. 368–377, 2025.

[5] W. Chansanam, K. Poonpon, T. Manakul, and U. Detthamrong, "Success and challenges in MOOCs: A literature systematic review technique," TEM Journal, vol. 10, no. 4, pp. 1728–1732, 2021, doi: 10.18421/TEM104-32.

[6] K. G. Rahul et al., "Prediction of research trends using LDA based topic modeling," vol. 3, no. 1, pp. 298-304, 2022.

[7] D. E. Bloom, D. Canning, and A. Lubet, "Global population aging: Facts, challenges, solutions and perspectives," Daedalus, vol. 144, no. 2, pp. 80-92, 2020.

[8] J. L. Dieleman, E. Maddison, and R. Sadana, "An aging world requires more support for health systems," Think Global Health, 2021.

[9] United Nations, "World population prospects 2019: Highlights," United Nations Department of Economic and Social Affairs, 2019.

[10] D. Blei, AY. NG, and MI. Jordan, "Latent Dirichlet Allocation," The Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[11] K. Krippendorff, Content analysis: An introduction to its methodology, 2nd ed. Thousand Oaks, CA: Sage, 2004.

[12] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101, 2006.

[13] E. Garfield, "Citation analysis as a tool in journal evaluation," Science, vol. 178, no. 4060, pp. 471-479, 1972.
[14] A. Pritchard, "Statistical bibliography or bibliometrics," Journal of Documentation, vol. 25, no. 4, pp. 348-349, 1969.
[15] S. Wasserman and K. Faust, Social network analysis: Methods and applications. Cambridge, U.K.: Cambridge University Press, 1994.
[16] H. Zhou, S. Wang, and Q. Liu, "Topic modeling and its applications in the health care domain," Journal of Health Informatics, vol. 24, no. 3, pp. 335-347, 2018.
[17] J. Liu, Y. Zhang, and X. Liu, "Analyzing health care policies for the elderly using topic modeling techniques," Health Policy and Management Review, vol. 12, no. 4, pp. 52-63, 2020.
[18] D. M. Blei and J. D. Lafferty, "A correlated topic model of Science," The Annals of Applied Statistics, vol. 1, no. 1, pp. 17-35, 2007.
[19] M. B. P. Maja, J. D. Jasminka, and K. L. Luka, "Topic modeling in social sciences-case study of web of science," Central European Conference on Information and Intelligent Systems, Faculty of Organization and Informatics Varazdin, 2020.
[20] L. E. George and L. Birla, "A Study of Topic Modeling Methods," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 109-113, 2018.
[21] K. D. Kuhn, "Using structural topic modeling to identify latent topics and trends in aviation incident reports," Transportation Research Part C: Emerging Technologies, vol. 87, pp. 105-122, 2018.
[22] H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," Multimedia Tools and Applications, vol. 78, pp. 15169-15211, 2019.
[23] B. Yin and C. H. Yuan, "Detecting latent topics and trends in blended learning using LDA topic modeling," Education and Information Technologies, vol. 27, no. 9, pp. 12689-12712, 2022.
[24] M. W. Nielsen and L. Börjeson, "Gender diversity in the management field: Does it matter for research outcomes?" Research Policy, vol. 48, no. 7, pp. 1617-1632, 2019.
[25] A. Ali, "Data preparation and data cleaning: As the data is in English text format, we used Natural Language Processing (NLP) in Python and Text pre-processing by PyCaret, which is an open source and reliable tool," Journal of Data Science, vol. 15, no. 3, pp. 245-259, 2020.
[26] B. G. Bettina and H. K. Kurt, "Topicmodels: an R package for fitting topic models," Journal of Statistical Software, vol. 40, no. 13, pp. 1-30, 2011.
[27] C. R. Chowdhury and P. Bhuyan, "Information retrieval using fuzzy c-means clustering and modified vector space model," 2010 3rd International Conference on Computer Science and Information Technology, vol. 1, pp. 696-700, July 2010, IEEE.
[28] K. Stevens et al., "Exploring topic coherence over many models and many topics," Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pp. 952-961, July 2012.
[29] D. Blei, "Probabilistic Topic Models," Communications of the ACM, vol. 55, pp. 77-84, 2012.
[30] R. Saeed and M. Fatemeh, "Sentiment analysis researches story narrated by topic modeling approach," Social Sciences & Humanities Open, 100309, 2022.
[31] S. Ray, A. Ahmad, and C. A. Kumar, "Review and Implementation of Topic Modeling in Hindi," Appl Artif Intell, vol. 33, no. 11, pp. 979–1007, 2019.
[32] F. Qiao and J. Williams, "Topic modeling and sentiment analysis of global warming tweets: Evidence from big data analysis," Journal of Organizational and End User Computing, vol. 34, no. 3, pp. 1–18, 2022.
[33] A. Singh and A. Glińska-Neweś, "Modeling the public attitude towards organic foods: a big data and text mining approach," J Big Data, vol. 9, no. 1, p. 2, 2022.
[34] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63-70, June 2014.
[35] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," Proceedings of the international working conference on advanced visual interfaces, pp. 74-77, May 2012.
[36] N. Donthu et al., "How to conduct a bibliometric analysis: An overview and guidelines," Journal of Business Research, vol. 133, pp. 285-296, 2021.
[37] W. Chansanam and K. Tuamsuk, "Thai Twitter sentiment analysis: Performance monitoring of politics in Thailand using text mining techniques," International Journal of Innovation, Creativity and Change, vol. 11, no. 12, pp. 436–452, 2020.
[38] X. Li and L. Lei, "A bibliometric analysis of topic modeling studies (2000–2017)," Journal of Information Science, vol. 47, no. 2, pp. 161-175, 2021.
[39] S. B. Cho, S. Shin, and D. S. Kang, "A study on the research trends on open innovation using topic modeling," Informatization Policy, vol. 25, no. 3, pp. 52-74, 2018.
[40] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101, Suppl 1, pp. 5228-5235, 2004.
[41] W. Zhao, J. Chen, and L. Zou, "Topic modeling for focused analysis of big data: A case study on COVID-19," *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 345-357, 2020.