Ciência & Saúde Coletiva
cienciaesaudecoletiva.com.br
ISSN 1413-8123. v.31, n.5
DOI: 10.1590/1413-81232026315.01352026EN

Analysis of patterns of violence against women in Brazil: an unsupervised machine learning approach

Andre Massahiro Shimaoka https://orcid.org/0000-0002-9400-8083 [1]; Antonio Carlos da Silva Junior https://orcid.org/0000-0002-6316-8711 [1]; José Marcio Duarte https://orcid.org/0000-0002-6159-0206 [1]; Kelsy Catherina Nema Areco https://orcid.org/0000-0002-7801-757X [2]; Marcelo Batista Ribeiro https://orcid.org/0009-0007-1045-0688 [2]; Leonardo Martins Araujo https://orcid.org/0009-0006-2531-7700 [2]; Débora Dupas Gonçalves do Nascimento https://orcid.org/0000-0003-2291-2302 [3]; Maria Elisabete Salvador https://orcid.org/0000-0001-8692-5968 [1]; Paulo Bandiera-Paiva https://orcid.org/0000-0001-9409-3970 [1]

Abstract *Analyze patterns of association between diagnoses related to domestic violence in hospital admissions and identify clinical-demographic profiles using unsupervised machine learning. Data from the SUS Hospital Information System between 2008 and 2023 were used, covering 90,798 hospitalizations of women aged 20 to 59 years with ICD-10 codes related to violence. Frameworks for data preparation were applied, as well as algorithms for identifying association rules between diagnoses and topic modeling. The hospitalization rate stabilized after 2012, with an average between 2.6 and 3.2 per 100,000 women. States such as São Paulo, Bahia, and Minas Gerais accounted for 46% of absolute cases; Rio Grande do Norte and Pará presented the highest proportional rates. The algorithm identified significant rules between types of injury and mechanisms of aggression. Latent Dirichlet Allocation modeling revealed nine distinct profiles, highlighting young women undergoing emergency surgeries due to polytrauma. The use of machine learning identified relevant clinical-epidemiological patterns to support prediction and surveillance strategies in Primary Care, contributing to addressing gender-based violence.*

Key words *Violence Against Women, Hospital Information Systems, Machine Learning, Artificial Intelligence*

[1] Departamento de Informática em Saúde, Escola Paulista de Medicina, Universidade Federal de São Paulo. R. Botucatu 740, Vila Clementino. 04023-062 São Paulo SP Brasil. andre.shimaoka@unifesp.br
[2] Hospital São Paulo, Associação Paulista para o Desenvolvimento da Medicina. São Paulo SP Brasil.
[3] Fiocruz Mato Grosso do Sul. Campo Grande MS Brasil.

Introduction

Violence against women is a serious public health problem in Brazil, leading to high rates of morbidity, mortality, and psychological distress among women[1]. Beyond its physical and emotional consequences, gender-based violence places a substantial burden on the health system. Primary Health Care (PHC) is the main gateway to the Brazilian Unified Health System (SUS). It plays a strategic role in early detection, qualified reception, and coordination with social protection networks[2].

In Brazil, the steady increase in reports of violence against women over recent decades underscores the magnitude and urgency of this public health issue[3]. Despite expanded policies and protection networks, regional inequalities, structural barriers, and limited access to health and social services persist, compromising prevention and early detection[4].

Analyzing the most severe cases, such as those resulting in hospitalizations, is especially relevant for understanding complex clinical patterns. Previous Brazilian studies have examined individual or sociodemographic characteristics[1,5], yet integrated analyses of diagnoses and clinical features of hospitalizations are limited, hindering identification of severity profiles and care needs.

International experiences[6,7] highlight the role of Digital Health, including Data Science, in strengthening surveillance, integrating information, and identifying vulnerable populations. Guidelines such as the Fit for the Future plan[6] and the World Health Organization (WHO) Digital Transformation Handbook for Primary Health Care[7] point to the adoption of predictive, responsive, and population-centered care models. Unsupervised machine learning techniques have emerged as promising tools for revealing hidden patterns in large datasets[8,9].

Accordingly, this study aimed to analyze, using unsupervised machine learning techniques, patterns of association among diagnoses related to violence against women in hospitalizations recorded in the SUS Hospital Information System (SIH-SUS), and to identify the clinical-demographic profiles of these admissions, thereby providing evidence to inform future surveillance, prevention, and comprehensive care strategies.

Methods

We conducted a retrospective, analytical, and observational study using secondary SIH-SUS data from 2008 to 2023, covering the entire Brazilian territory. The study population comprised women aged 20-59 years admitted to SUS hospitals with domestic violence-related diagnoses. To explore data patterns, we applied unsupervised machine learning techniques, including topic modeling and association rules. The Unifesp Research Ethics Committee approved the study under Opinion No. 00263/2023.

The study was conducted at the Department of Health Informatics of the São Paulo School of Medicine, Federal University of São Paulo, in partnership with the São Paulo Association for the Development of Medicine, from April to July 2025. All data used were public, anonymized, and handled in accordance with current ethical and regulatory principles, including the Brazilian General Personal Data Protection Law (Law No. 13,709/2018) and National Health Council Resolution No. 510/2016.

We employed public secondary databases, including SIH-SUS[10] for information on SUS hospitalizations, ICD-10[11] for diagnostic classification, SIGTAP[12] for descriptions of the procedures performed, and population data from the Brazilian Institute of Geography and Statistics (IBGE)[13]. Based on these sources, the hospitalization rate per 100,000 inhabitants was calculated and stratified by federative unit (state) and year, using the number of hospitalizations as the numerator and the female population aged 20-59 years as the denominator, multiplied by 100,000.

The inclusion criteria comprised hospitalizations of adult women aged 20-59 years, excluding children, adolescents, and older women, as per the categorization proposed in the Ministry of Health's Health Surveillance Guide[14]. Records with diagnoses related to domestic violence were selected, including ICD-10 codes for assaults (X85-X99, Y00-Y09), maltreatment syndromes (T74), sequelae of assault (Y87.1), and examination and observation following alleged rape and seduction (Z04.4). To ensure comprehensive capture of diagnoses, we considered the primary diagnosis, secondary diagnosis, and additional diagnosis fields available in the SIH-SUS database. The final sample comprised 90,798 hospitalizations.

To guide planning and the data preparation process in health, we applied the HRSP-AI[15] and HealthDataPrep[9] frameworks. The steps included: 1) De-identification: removal of identifying fields such as the Hospitalization Authorization (AIH), taxpayer identification numbers for individuals and companies; 2) Cleaning: exclusion of variables with only one value or variance below 1%, such as schooling level, occupation, social security status, among others; 3) Integration: combining SIH-SUS, population (IBGE), procedure table (SIGTAP), and ICD data; 4) Transformation: conversion to categorical attribute-value format (state, gender, ethnicity/skin color, bed type, progress, death, and ICU) and discretization of the continuous age variable into intervals (age groups); and 5) Attribute selection: state, age group, ethnicity/skin color, bed type, procedure, progress, death, ICU, hospitalization type, and diagnoses (Figure 1).

For association rule analysis, we employed the Apriori algorithm to identify patterns among diagnoses[16]. The following criteria were established: Support≥1%, Confidence≥30%, and Lift≥1.5. These values were adjusted by size of the database and were based on previous studies that applied the Apriori algorithm to health datasets[17,18]. The metrics used were *Support*, defined as the frequency of an item combination vis-à-vis the total dataset; *Confidence*, defined as the probability of occurrence of the consequent diagnosis given the antecedent; and *Lift*, defined as the strength of dependency between diagnoses (>1 indicates positive dependence)[19,20].

For topic modeling, we applied the LDA algorithm[8,21], and the data were structured in attribute-value format. In addition, we used the Perplexity metric[8,22] to determine the optimal number of topics, testing models with 2 to 15 topics. This approach was grounded in previous studies using health data[23,24]. The final number of nine topics was selected based on stabilization of perplexity. The source code used for all steps of data preparation and application of the Apriori and LDA algorithms was made publicly available at https://doi.org/10.5281/zenodo.17872231.

Results

The results obtained from the analysis of SIH-SUS data using unsupervised machine learning techniques revealed relevant patterns of hospitalizations for violence against women in Brazil from 2008 to 2023. This section is organized into three parts: (1) General characteristics of hospitalizations; (2) Identification of association rules among diagnoses using the Apriori algorithm; and (3) Topic modeling with the LDA algorithm to describe latent profiles of hospitalizations.

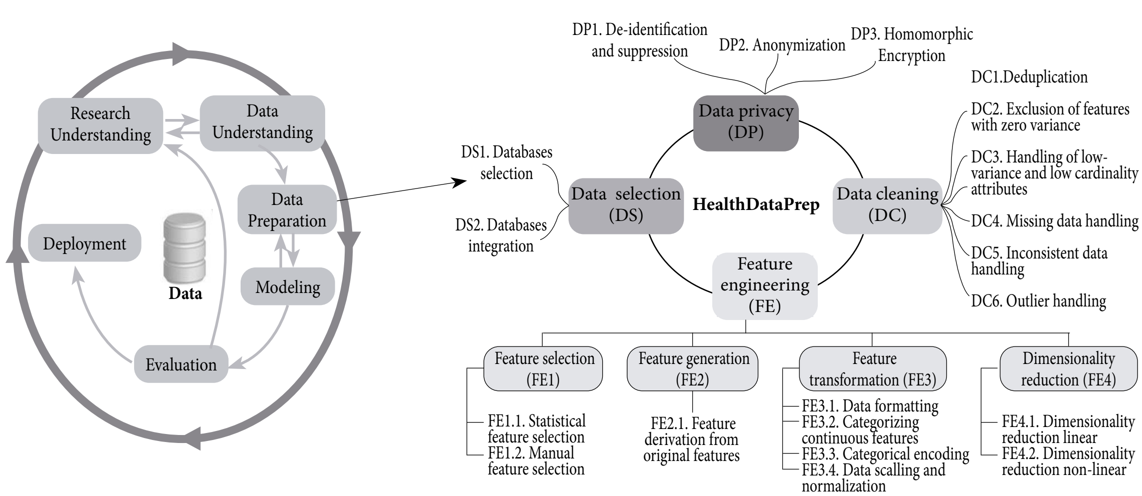

Figure 1. Use of the HRSP-AI[15] framework to guide the AI project process and the HealthDataPrep framework for preparing unsupervised data.

Source: Shimaoka *et al.*[9].

Each approach highlighted different dimensions of gender-based violence in the hospital context, with implications for health surveillance, public policy formulation, and clinical practices sensitive to the complexity of these cases.

General characteristics of hospitalizations

A total of 828,264 (0.43%) of the 193,455,105 hospitalizations recorded in SUS from 2008 to 2023 corresponded to violence-related codes, of which 144,140 (17.4%) involved women. After applying the age criteria, 90,798 cases of women aged 20-59 years were retained.

The hospitalization rate for violence against women stabilized after 2012, ranging from 2.6 to 3.2 per 100,000 inhabitants. In absolute numbers, São Paulo (17,963 cases), Bahia (13,893), and Minas Gerais (10,184) accounted for 46% of the national total. When population rates were considered, Rio Grande do Norte had the highest mean rate (17.49/100,000 inhabitants), followed by Para (6.49/100,000).

Age distribution showed greater concentration in the 20-30-year age group, with a progressive decline with age. Regarding ethnicity/skin color, 35.57% self-identified as brown, 20.49% as white, and 38.29% had no information recorded. Most hospitalizations occurred as urgent care (81.87%) and in surgical beds (67.17%) (Table 1).

Association rules among diagnoses

The Apriori algorithm identified nine significant association rules according to the established criteria. The rule with the highest confidence linked injury of muscle and tendon at wrist and hand level (S66) with assault by sharp object (X99), with a support of 1.73%, Confidence of 69.5%, and Lift of 3.14 (Chart 1). Other relevant associations included:

• Fracture of skull and facial bones (S02) – Assault by physical force (Y04): Confidence 61.05%, Lift 2.99.

• Open wound of breast (S21) – Assault by sharp object (X99): Confidence 59.01%, Lift 2.67.

• Injury of other and unspecified intrathoracic organs (S27) – Assault by sharp object (X99): Confidence 57.28%, Lift 2.59.

The results revealed consistent patterns between specific types of injuries and assault mechanisms, suggesting correlations between the instruments used and the resulting injuries.

Profiles of hospitalizations based on topic modeling

Perplexity analysis indicated that nine was the optimal number of topics, with substantial change up to that point and variations below 1% in subsequent values, suggesting relative stabilization (Figure 2). Each identified topic corresponded to a recurrent profile of hospital care related to violence against women, characterized by specific combinations of the most frequent attribute-value pairs. The nine identified profiles showed distinct distributions (Chart 2):

Topic 7 (17.2% of cases): Young women (20-29 years) receiving urgent surgical care for multiple trauma or tube thoracostomy with pleural drainage, discharged with improvement. Brown women predominated, with diagnoses of assault by sharp object (X99) or injury of other and unspecified intrathoracic organs (S27).

Topic 6 (13.5%): Women hospitalized in Sao Paulo and Minas Gerais, in clinical beds, admitted as urgent care, undergoing procedures related to injuries of unspecified location, with diagnoses of assault by physical force (Y04) or open wounds involving multiple body regions (T01).

Topic 3 (10.8%): Older women (40-59 years), predominantly white, residing in São Paulo and Santa Catarina, admitted to surgical beds after assault by bodily force (Y04) or blunt object (Y00).

Topic 5 (5.5%): Mixed-race women from São Paulo and Ceará, victims of assault by rifle, shotgun and larger firearm discharge (X94) or exposure to the toxic effect of other and unspecified substances (T65), undergoing treatment for poisoning or emergency surgical procedures, with ICU admission and death as the outcome.

The remaining topics represented other profiles with diverse geographic, age-related, and clinical characteristics, evidencing heterogeneous violence against women cases within the hospital system.

Discussion

This study conducted a comprehensive analysis of hospitalizations for violence against women in Brazil from 2008 to 2023 using unsupervised machine learning techniques. The findings reveal patterns across three main dimensions: the general characteristics of hospitalizations, including age distribution, ethnicity/skin color, bed type, and regional patterns; patterns of as-

Table 1. Sociodemographic and clinical characteristics of hospitalizations due to violence against women (20-59 years) in Brazil, 2008-2023 (n=90,798).

Variable	Category	n	%
Age group	20-29 years	31,258	34.42%
	30-39 years	27,637	30.44%
	40-49 years	19,256	21.21%
	50-59 years	12,647	13.93%
Ethnicity/Skin color	Brown	32,297	35.57%
	White	18,609	20.49%
	Black	3,965	4.37%
	Yellow	897	0.99%
	Indigenous	260	0.29%
	Unknown/Not informed	34,770	38.29%
Hospitalization type	Urgent care	74,326	81.87%
	Other	8,316	9.15%
	Elective	8,156	8.98%
Bed type	Surgical	60,990	67.17%
	Clinical	28,031	30.87%
	Other	1,777	1.96%
ICU	Yes	84,239	92.78%
	No	6,559	7.22%
5 UFs with the highest number	SP	17,965	19.77%
	BA	13,893	15.30%
	MG	10,184	11.22%
	PA	7,266	8.00%
	RN	5,538	6.10%

Source: Brazilian Unified Health System Hospital Information System (SIH-SUS).

Chart 1. Results of the rules for patterns of association between diagnoses related to violence in SUS hospitalizations.

Antecedent	Consequent	Support	Confidence	Lift
S66 - Injury of muscle and tendon at wrist and hand level	X99 - Assault by sharp object	1.73%	69.50%	3.14
S02 - Fracture of skull and facial bones	Y04 - Assault by physical force	4.74%	61.05%	2.99
S21 - Open wound of breast	X99 - Assault by sharp object	1.42%	59.01%	2.67
S27 - Injury of other and unspecified intrathoracic organs	X99 - Assault by sharp object	2.88%	57.28%	2.59
S31 - Open wound of abdomen, lower back and pelvis	X99 - Assault by sharp object	1.50%	52.27%	2.36
S06 - Concussion	Y04 - Assault by physical force	4.08%	45.25%	2.21
S36 - Injury of intra-abdominal organs	X99 - Assault by sharp object	2.27%	44.18%	2.00
T01 - Open wounds involving multiple body regions	X99 - Assault by sharp object	1.62%	42.54%	1.92
S52 - Fracture of forearm	Y04 - Assault by physical force	1.81%	31.30%	1.53

Source: Authors, 2025.

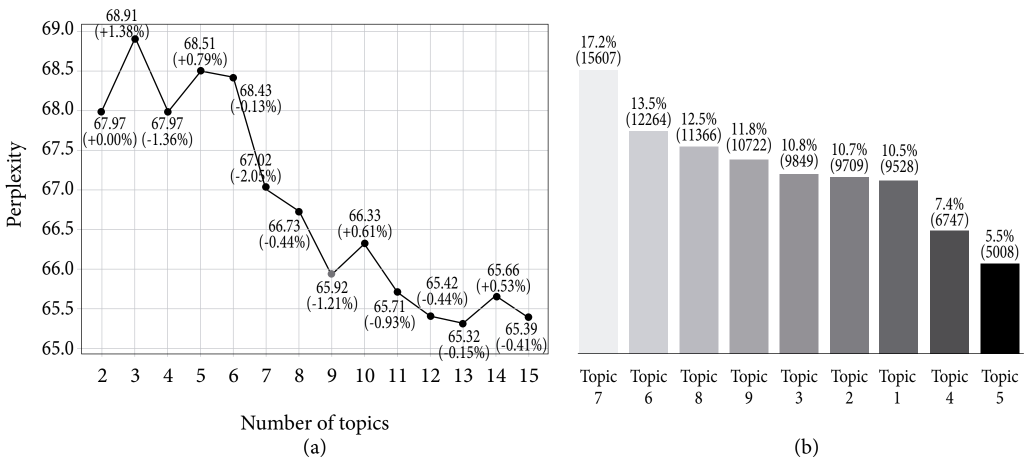

Figure 2. (a) Analysis of the Perplexity metric for Different Numbers of Topics; (b) Distribution of topics for hospitalizations due to violence against women.

Source: Authors, 2025.

sociation among diagnoses; and hospitalization profiles. This approach lead us to understand the heterogeneous cases, identify risk groups, and inform public policies, health service planning, and prevention strategies, with particular emphasis on PHC.

The stabilization of hospitalization rates after 2012 may be interpreted as reflecting regulatory advances, such as the Maria da Penha Law (Law 11.340/2006), and the expanded services within the Network of Care for Women in Situations of Violence, especially referral centers and hospital services for sexual violence victims. Nevertheless, studies indicate that a substantial proportion of violence against women remains invisible within the health system because of underreporting, inadequate labeling of SIH-SUS diagnostic codes, or the redirection of victims that do not require hospitalization to emergency and urgent care services[25,26].

The regional disparities observed, especially in Rio Grande do Norte and Pará, suggest a complex interaction among the prevalence of violence, service coverage, reporting culture, and access to care. Recent research on the geographic distribution of interpersonal violence in Brazil reinforces the existence of 'high-incidence pockets' associated with social vulnerability, weak protection networks, and impunity of perpetrators[27-29].

The association rules identified by the Apriori algorithm are particularly relevant from a clinical-forensic perspective. The strong association between wrist/hand trauma and assault by sharp object (Confidence 69.5%; Lift 3.14) may reflect instinctive defensive movements during the attack, as also described in forensic medicine studies[30]. Similarly, the correlation between craniofacial fractures and assault by physical force (Lift 2.99) reinforces recognized patterns of direct and repeated assault, often associated with severe domestic violence situations[31].

LDA modeling identified nine latent hospitalization profiles, revealing case heterogeneity and highlighting the need for differentiated care approaches. The most prevalent profile, involving young women with multiple trauma resulting from assaults with sharp objects, raises concerns about the severity of violent episodes and their physical and emotional repercussions. Qualitative studies have shown that young women tend to be more reluctant to seek formal help, which may delay preventive interventions and deteriorate clinical outcomes[32,33].

Based on analysis of these profiles, it becomes possible to propose stratification of the risk of recurrence, clinical deterioration, and social vulnerability. Initiatives such as the implementation of hospital-based centers for comprehensive care for women in situations of violence

Chart 2. Topics with the most relevant terms that make up the profile of hospitalizations due to violence against women.

	Age group	Ethnicity/ Skin color	UF	Bed	Type	ICD(s)	Procedure	Progress	ICU
Topic 1	20-49	Brown	SP	Clinical	Urgent care	S06	Conservative treatment of mild/moderate traumatic brain injury	Discharge improved	No
Topic 2	20-49	Brown and Not informed	PA	Surgical	Urgent care	S52, X99, Y04, S66	-	Discharge improved	No
Topic 3	40-59	White	SP and SC	Surgical	Urgent care, Elective, and Other	Y04 and Y00	-	Discharge improved and discharged with a return visit scheduled for follow-up	No
Topic 4	20-59	Not informed	RN	Surgical	Urgent care and Elective	Y08, S82	Treatment involving multiple surgeries	Discharge healed	No
Topic 5	20-39	Brown	SP and CE	Clinical and Surgical	Urgent care	X94, T65	Treatment for Intoxication or Poisoning Surgical urgent care	Death certificate issued by the Forensic Medical Institute	Yes
Topic 6	20-39	Brown and Not informed	SP and MG	Clinical	Urgent care	X99, Y04, T01	Treatment of injuries of unspecified location	Discharge improved and Transfer to another establishment	No
Topic 7	20-29	Brown and Not informed	-	Surgical	Urgent care	X99, S27	Surgical treatment in polytrauma patients Thoracostomy with pleural drainage	Discharge improved	No
Topic 8	20-39	Brown, White, and Not informed	SP	Surgical	Urgent care	S02, Y04, X99, S31	-	Discharge improved and discharged with a return visit scheduled for follow-up	No
Topic 9	30-59	Brown	BA	Clinical and Surgical	Urgent care	Y09, S82 and Y00	-	Discharge improved	No

Source: Authors, 2025.

and active notification and sentinel surveillance protocols may benefit from these findings, improving coordination between epidemiological surveillance, hospital care, and social protection networks[34,35].

International studies[36,37] have showed the potential of machine learning techniques to identify hidden patterns of gender-based violence and predict risk, as evidenced in analyses using large hospital or public health datasets combined with association and clustering algorithms[36,37]. Such experiences reinforce the importance of investments in data infrastructure and interoperability within health systems, elements that remain fragile in the Brazilian context[38,39].

In this regard, WHO recommendations[7] for digital transformation in Health, especially in the field of Data Science, emphasize that this change transcend the mere digitization of records and involve systemic reorganization of information flows with a user-centered focus. One of the proposals in the document prioritizes creating person-centered point-of-service systems and incorporating functionalities for clinical decision support, prediction, and automated surveillance. The application of algorithms to identify patterns of hospital violence, as performed in this study, is aligned with the WHO proposal[7] to use technologies to recognize vulnerable populations, identify risks, and promote personalized, intersectoral responses in Health.

In addition, the WHO[7] document recommends that, to ensure the effectiveness of these tools, it is essential to map work processes and align functional requirements with local realities, thereby ensuring that digital systems support both health professionals and managers in decision-making. The present study provides knowledge on clinical-epidemiological profiles, which may inform future discussions on improving information systems and decision-support tools.

Despite the robust database and the methodology applied, this study has limitations. Temporal trends and regional differences may reflect changes in recordkeeping, access, and service availability rather than actual variations in violence. Underreporting and inadequate recording of diagnostic codes may compromise internal validity. Moreover, the study includes only hospitalizations and does not encompass visits to emergency care, primary care, or cases that never reach the health system, which limits external validity and prevents a broader assessment of the magnitude of this event.

In conclusion, this study characterized hospitalizations for violence against women in Brazil from 2008 to 2023 and identified patterns of association among diagnoses and clinical-demographic profiles using unsupervised techniques. The results reveal that hospitalizations are concentrated mainly among young, brown women treated on an urgent care basis and often undergoing surgical procedures, with particular emphasis on the strong association between sharp object injuries and multiple trauma. The heterogeneous identified profiles demonstrate that violence produces distinct clinical outcomes, suggesting differentiated care needs and groups with greater vulnerability.

The findings of this study align with national[40,41] and international[6,7,42] frameworks that recognize violence against women as a complex and multifactorial public health problem. By identifying and organizing little-explored clinical-epidemiological patterns, the present study fills gaps in the literature and provides relevant conceptual support for future discussions on improving surveillance, improving databases, and developing analytical tools in Health.

Collaborations

AM Shimaoka, AC Silva Junior, JM Duarte, KCN Areco, MB Ribeiro, LM Araujo, DDG Nascimento, ME Salvador and P Bandiera-Paiva contributed to the conception of the study, analysis and interpretation of the data, writing of the article, relevant critical review of the intellectual content and approval of the final version to be published.

Data availability statement

The data sources adopted in the research are indicated in the article's body.

References

1. Silva BBA, Pereira RM, Rodrigues CMO, Fagundes RAA. Analyzing Patterns of Violence Against Women in Pernambuco Using Self-Organizing Maps and Apriori Algorithm. In: *2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. Bogota: IEEE; 2024. p. 1-6.
2. Mendonça CS, Machado DF, Almeida MASD, Castanheira ERL. Violência na Atenção Primária em Saúde no Brasil: uma revisão integrativa da literatura. *Cien Saude Colet* 2020; 25(6):2247-2257.
3. Lima GCC, Passos CMD, Pinheiro ALS, Ribeiro ÍJS, Maia EG. Temporal trend and epidemiological profile of notifications of violence against women in Brazil: 2014-2023. *Epidemiol Serv Saude* 2025; 34:e20240475.
4. Signorelli MC, Taft A, Pereira PPG. Domestic violence against women, public policies and community health workers in Brazilian Primary Health Care. *Cien Saude Colet* 2018; 23(1):93-102.
5. Vasconcelos NMD, Gomes CS, Souza JBD, Andrade FMDD, Bernal RTI, Machado EL, et al. Quem são as mulheres adultas expostas à violência no Brasil? *Rev Saude Publica* 2025; 59:e236103.
6. National Health Service (NHS). *Department of Health and Social Care. Fit for the Future: 10 Year Health Plan for England* [Internet]. 2025 [cited 2025 dez 10]. Available from: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future.
7. World Health Organization (WHO). *Digital transformation handbook for primary health care: optimizing person-centred point of service systems* [Internet]. 2024 [cited 2025 dez 9]. Available from: https://www.who.int/publications/i/item/9789240093362.
8. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003; 993-1022.
9. Shimaoka AM, Lopes LR, Junior ACDS, Bandiera-Paiva P. Health data preparation framework for unsupervised models (HealthDataPrep): A study with Alzheimer's disease mortality data from Brazil. *Int J Health Manag Rev* 2025; 11(1):e0407.
10. Brasil. DATASUS. *Bases de Dados do DATASUS* [Internet]. 2025 [acessado 2025 dez 9]. Disponível em: https://datasus.saude.gov.br/transferencia-de-arquivos/.
11. Brasil. DATASUS. *CID-10* [Internet]. 2008 [acessado 2025 dez 9]. Disponível em: http://www2.datasus.gov.br/cid10/V2008/cid10.htm.
12. Brasil. DATASUS. *SIGTAP* [Internet]. 2025 [acessado 2025 dez 9]. Disponível em: http://tabela-unificada.datasus.gov.br/tabela-unificada/app/download.jsp.
13. Instituto Brasileiro de Geografia e Estatística (IBGE). *Projeção da População* [Internet]. 2024 [acessado 2025 dez 9]. Disponível em: https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html.
14. Brasil. Ministério da Saúde (MS). *Guia de Vigilância em Saúde* [Internet]. 2024 [acessado 2025 dez 9]. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/vigilancia/guia-de-vigilancia-em-saude-volume-1-6a-edicao.
15. Bandiera-Paiva P, Shimaoka AM, Areco KCN, Ribeiro MB, Araujo LM, Junior ACS, Nascimento DD, Salvador ME, Mansur NS, Barbosa DA. Uso de Inteligência Artificial para análise da violência contra mulheres em São Paulo. *Rev Saude Publica*. No prelo 2025.
16. Agrawal R, Srikant R. *Fast Algorithms for Mining Association Rules in Large Databases* [Internet]. San Francisco: Morgan Kaufmann Publishers Inc; 1994 [acessado 2025 dez 9]. p. 487-99. Disponível em: https://dl.acm.org/doi/10.5555/645920.672836.
17. Cha S, Kim SS. Discovery of Association Rules Patterns and Prevalence of Comorbidities in Adult Patients Hospitalized with Mental and Behavioral Disorders. *Healthcare* 2021; 9(6):636.
18. Ma H, Ding J, Liu M, Liu Y. Connections between Various Disorders: Combination Pattern Mining Using Apriori Algorithm Based on Diagnosis Information from Electronic Medical Records. *BioMed Res Int* 2022; 2022(1):2199317.

19. Hipp J, Güntzer U, Nakhaeizadeh G. Algorithms for association rule mining – a general survey and comparison. ACM SIGKDD *Explor Newsl* 2000; 2(1):58-64.
20. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. Washington D.C.: ACM; 1993. p. 207-216.
21. Mustakim M, Wardoyo R, Mustofa K, Rahayu GR, Rosyidah I. Latent Dirichlet Allocation for Medical Records Topic Modeling: Systematic Literature Review. In: *2021 Sixth International Conference on Informatics and Computing (ICIC)*. Jakarta: IEEE; 2021. p. 1-7.
22. Gan J, Qi Y. Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example. *Entropy* 2021; 23(10):1301.
23. Li W, Min X, Ye P, Xie W, Zhao D. Temporal topic model for clinical pathway mining from electronic medical records. *BMC Med Inform Decis Mak* 2024; 24(1):20.
24. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, Amin S, Limper AH, Khosla S, Liu H. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. *J Biomed Inform* 2020; 102:103364.
25. Vasconcelos NMD, Bernal RTI, Souza JBD, Bordoni PHC, Stein C, Coll CDVN, Murray J, Malta DC. Subnotificação de violência contra as mulheres: uma análise de duas fontes de dados. *Cien Saude Colet* 2024; 29(10):e07732023.
26. Kind L, Orsini MDLP, Nepomuceno V, Gonçalves L, Souza GAD, Ferreira MFF. Subnotificação e (in)visibilidade da violência contra mulheres na atenção primária à saúde. *Cad Saude Publica* 2013; 29(9):1805-1815.
27. Soares MLM, Guimarães NGM, Bonfada D. Tendência, espacialização e circunstâncias associadas às violências contra populações vulneráveis no Brasil, entre 2009 e 2017. *Cien Saude Colet* 2021; 26(11):5751-5763.
28. Instituto de Pesquisa Econômica Aplicada (Ipea). *Atlas da Violência 2025* [Internet]. 2025 [acessado 2025 jul 30]. Disponível em: https://www.ipea.gov.br/atlasviolencia/arquivos/artigos/5999-atlasdaviolencia2025.pdf.
29. Cassini MRDOL, Ribeiro AP, Oliveira GL. Lacunas na abordagem da violência sexual contra a mulher: a quem de fato estamos protegendo? *Cad Saude Publica* 2024; 40(7):e00058424.
30. Das NG, Das N, Sil A. Pattern and Forensic Significance of Defense Injuries in Homicide Cases: A Cross-Sectional Study. *Cureus* 2025; 17(3):e79959.
31. Dourado SDM, Noronha CV. Marcas visíveis e invisíveis: danos ao rosto feminino em episódios de violência conjugal. *Cien Saúde Colet* 2015; 20(9):2911-2920.
32. Prosman G, Lo Fo Wong SH, Lagro-Janssen ALM. Why abused women do not seek professional help: a qualitative study. *Scand J Caring Sci* 2014; 28(1):3-11.
33. Kiss L, d'Oliveira AFL, Zimmerman C, Heise L, Schraiber LB, Watts C. Brazilian policy responses to violence against women: government strategy and the help-seeking behaviors of women who experience violence. *Health Hum Rights* 2012; 14(1):E64-E77.
34. Minayo MCDS, Souza ERD, Silva MMAD, Assis SGD. Institucionalização do tema da violência no SUS: avanços e desafios. *Cien Saude Colet* 2018; 23(6):2007-2016.
35. Avanci JQ, Oliveira QBM, Assis SGD. Ações de vigilância das violências em serviços da atenção primária, hospitalar e de reabilitação no Brasil. *Cien Saude Colet* 2025; 30(3):e17372024.
36. Bernal-Monroy ER, Castañeda-Monroy ED, Rentería-Ramos RR, Campaña-Bastidas SE, Barrera J, Palacios-Yampuezan TM, et al. Detection of Victimization Patterns and Risk of Gender Violence Through Machine Learning Algorithms. *Informatics* 2025; 12(1):21.
37. González-Prieto Á, Brú A, Nuño JC, González-Álvarez JL. Hybrid machine learning methods for risk assessment in gender-based crime. *Knowl-Based Syst* 2023; 260:110130.
38. Coelho Neto GC, Andreazza R, Chioro A. Integração entre os sistemas nacionais de informação em saúde: o caso do e-SUS Atenção Básica. *Rev Saude Publica* 2021; 55:93.
39. Chiavegatto Filho ADP. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiol Serv Saude* 2015; 24(2):325-332.
40. Brasil. Ministério da Saúde (MS). *Plano Nacional de Saúde 2024-2027 – CIDES* [Internet]. 2024 [acessado 2025 jul 31]. Disponível em: https://cides.se.gov.br/wp-content/uploads/2025/03/PLANO-NACIONAL-DE-SAUDE-2024-2027.pdf.
41. Brasil. Ministério da Saúde (MS). Departamento de Informática do SUS. *Estratégia de Saúde Digital para o Brasil 2020-2028* [Internet]. 2020 [acessado 2025 jul 30]. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf.
42. Organização Pan-Americana da Saúde (OPAS). *Plano Estratégico da Organização Pan-Americana da Saúde 2020-2025* [Internet]. 2020 [acessado 2025 jul 30]. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52968/9789275722756_por.pdf.

Article submitted 02/08/2025
Approved 26/01/2026
Final version submitted 28/01/2026