

Análise de padrões da violência contra mulheres no Brasil: uma abordagem com aprendizado de máquina não supervisionado

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

Análisis de los patrones de violencia contra las mujeres en Brasil: un enfoque con aprendizaje automático no supervisado

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

Resumo Analisar padrões de associação entre diagnósticos relacionados à violência doméstica em internações hospitalares e identificar perfis clínico-demográficos por aprendizado de máquina não supervisionado. Utilizaram-se dados do Sistema de Informações Hospitalares do SUS entre 2008 e 2023, abrangendo 90.798 internações de mulheres de 20 a 59 anos com códigos CID-10 relacionados à violência. Aplicaram-se frameworks para preparação dos dados, e algoritmos para identificação de regras de associação entre diagnósticos e modelagem de tópicos. A taxa de internações estabilizou após 2012, com média entre 2,6 e 3,2 por 100.000 mulheres. Estados como São Paulo, Bahia e Minas Gerais concentraram 46% dos casos absolutos; Rio Grande do Norte e Pará apresentaram as maiores taxas proporcionais. O algoritmo identificou regras significativas entre tipos de lesão e mecanismos de agressão. A modelagem Latent Dirichlet Allocation revelou nove perfis distintos, com destaque para mulheres jovens submetidas a cirurgias de urgência por politraumatismos. O uso de aprendizado de máquina identificou padrões clínico-epidemiológicos relevantes para subsidiar estratégias de predição e vigilância na Atenção Primária, contribuindo para o enfrentamento da violência de gênero.

Palavras-chave Violência contra a Mulher, Sistemas de Informação Hospitalar, Aprendizado de Máquina, Inteligência Artificial

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

Resumen Analizar patrones de asociación entre diagnósticos relacionados con la violencia doméstica en hospitalizaciones e identificar perfiles clínico-demográficos mediante aprendizaje automático no supervisado. Se utilizaron datos del Sistema de Información Hospitalaria del SUS entre 2008 y 2023, abarcando 90.798 hospitalizaciones de mujeres de 20 a 59 años con códigos CIE-10 relacionados con la violencia. Se aplicaron marcos para la preparación de los datos, así como algoritmos para la identificación de reglas de asociación entre diagnósticos y el modelado de temas. La tasa de hospitalizaciones se estabilizó después de 2012, con un promedio entre 2,6 y 3,2 por 100.000 mujeres. Estados como São Paulo, Bahia y Minas Gerais concentraron el 46% de los casos absolutos; Río Grande do Norte y Pará presentaron las tasas proporcionales más altas. El algoritmo identificó reglas significativas entre tipos de lesión y mecanismos de agresión. El modelado Latent Dirichlet Allocation reveló nueve perfiles distintos, destacándose mujeres jóvenes sometidas a cirurgías de urgencia por politraumatismos. El uso de aprendizaje automático identificó patrones clínico-epidemiológicos relevantes para sustentar estrategias de predicción y vigilancia en la Atención Primaria, contribuyendo al enfrentamiento de la violencia de género.

Palabras clave Violencia contra la Mujer, Sistemas de Información Hospitalaria, Aprendizaje Automático, Inteligencia Artificial

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.

Introdução

A violência contra a mulher constitui um grave problema de saúde pública no Brasil, representando uma das principais causas de morbidade, mortalidade e sofrimento psíquico entre as mulheres¹. Além das consequências físicas e emocionais, a violência de gênero impõe sobrecarga ao sistema de saúde. Entretanto, a Atenção Primária à Saúde (APS), como porta de entrada do Sistema Único de Saúde (SUS), tem papel estratégico na detecção precoce, no acolhimento qualificado e na articulação com redes de proteção social².

No contexto brasileiro, o crescimento consistente das notificações de violência contra mulheres nas últimas décadas reforça a magnitude e a urgência desse problema de saúde pública³. Apesar da expansão de políticas e redes de proteção, persistem desigualdades regionais, barreiras estruturais e limitações no acesso aos serviços de saúde e assistência, comprometendo a efetividade das estratégias de prevenção e detecção precoce⁴.

Nesse cenário, a análise dos casos de maior gravidade, como aqueles que evoluem para internação hospitalar, torna-se particularmente relevante, pois permite compreender padrões clínicos complexos. Estudos anteriores no Brasil exploraram características individuais ou sociodemográficas^{1,5}, mas a análise integrada dos diagnósticos e das características clínicas das internações ainda é limitada, dificultando a identificação de perfis de gravidade, riscos e necessidades assistenciais.

Experiências internacionais^{6,7} destacam o papel da Saúde Digital, como a Ciência de Dados, no fortalecimento da vigilância, na integração das informações e na identificação de populações vulneráveis. Diretrizes como o plano *Fit for the future*⁶ e *Digital Transformation Handbook for Primary Health Care da World Health Organization (WHO)*⁷ apontam para a adoção de modelos assistenciais preditivos, responsivos e centrados nas necessidades reais da população. As técnicas de aprendizado de máquina não supervisionado surgem como ferramentas promissoras para revelar padrões ocultos em grandes bases de dados^{8,9}.

Diante disso, este estudo tem como objetivos analisar, por meio de técnicas de aprendizado de máquina não supervisionado, padrões de associação entre diagnósticos relacionados à violência contra mulheres em internações hospitalares registradas no Sistema de Informações Hospitalares do SUS (SIH-SUS), e identificar

perfis clínico-demográficos dessas internações, oferecendo subsídios para futuras estratégias de vigilância, prevenção e cuidado integral.

Métodos

Realizou-se estudo observacional retrospectivo, desenho analítico, utilizando dados secundários do SIH-SUS no período de 2008 a 2023, abrangendo todo o território brasileiro. A população estudada compreendeu mulheres de 20 a 59 anos internadas no SUS com diagnósticos relacionados à violência doméstica. Para explorar padrões nos dados, foram aplicadas técnicas de aprendizado de máquina não supervisionado, incluindo modelagem de tópicos e regras de associação. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Unifesp, sob o parecer nº 00263/2023.

O estudo foi realizado no Departamento de Informática em Saúde da Escola Paulista de Medicina da Universidade Federal de São Paulo, em parceria com a Associação Paulista para o Desenvolvimento da Medicina, no período entre abril e julho de 2025. Todos os dados utilizados são públicos, anonimizados e foram tratados de acordo com os princípios éticos e regulatórios vigentes, conforme a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018) e a Resolução nº 510/2016 do Conselho Nacional de Saúde.

Para este estudo, foram utilizadas bases de dados secundárias públicas, incluindo SIH-SUS¹⁰ com informações sobre internações hospitalares no SUS, CID-10¹¹ para classificação de diagnósticos, SIGTAP¹² com descrições dos procedimentos realizados, e dados populacionais do IBGE¹³. Com essas informações, foi calculada a taxa de internações por 100.000 habitantes, estratificada por unidade federativa (UF) e ano, considerando o número de internações como numerador e a população feminina de 20 a 59 anos como denominador, multiplicando o resultado por 100.000.

Os critérios de inclusão contemplaram internações de mulheres adultas com idade entre 20 e 59 anos, excluindo crianças, adolescentes e idosas, em conformidade com a categorização proposta no Guia de Vigilância em Saúde do Ministério da Saúde¹⁴. Foram selecionados registros com diagnósticos relacionados à violência doméstica, abrangendo os códigos CID-10 de agressões (X85-X99, Y00-Y09), síndromes de maus-tratos (T74), sequelas de agressão (Y871) e exame/observação após alegação de

estupro (Z044). Para garantir a abrangência dos diagnósticos, consideraram-se os campos de diagnóstico principal, diagnóstico secundário e diagnósticos adicionais disponíveis na base do SIH-SUS. A amostra final compreendeu 90.798 internações.

Para orientar o planejamento e o processo de preparação dos dados na saúde, aplicaram-se os frameworks HRSP-AI¹⁵ e framework HealthDataPrep⁹. As etapas incluíram 1) Desidentificação: remoção de campos identificadores, como Autorização de Internação Hospitalar (AIH), CPF e CNPJ; 2) Limpeza: exclusão de variáveis com valores únicos ou variância inferior a 1%, como nível de instrução, ocupação, vínculo previdenciário, entre outros; 3) Integração: combinação de dados do SIH-SUS, população (IBGE), tabela de procedimentos (SIGTAP) e CID; 4) Transformação: conversão para formato atributo-valor categórico (UF, sexo, raça/cor, leito, evolução, óbito e UTI) e discretização da variável contínua idade em intervalos (faixas etárias); e 5) Seleção de atributos: UF, faixa etária, raça/cor, leito, procedimento, evolução, óbito, UTI, caráter de internação e diagnósticos (Figura 1).

Para a análise de regras de associação foi utilizado o algoritmo Apriori, para identificar padrões entre diagnósticos¹⁶. Estabeleceram-se os seguintes critérios: *Support* $\geq 1\%$, *Confidence* $\geq 30\%$ e *Lift* $\geq 1,5$. Esses valores foram ajustados considerando o porte da base de dados e funda-

mentados em estudos anteriores que aplicaram o algoritmo Apriori em bases de saúde^{17,18}. As métricas utilizadas foram: *Support*: frequência da combinação de itens em relação ao conjunto total; *Confidence*: probabilidade de ocorrência do diagnóstico consequente dado o antecedente e, *Lift*: força da dependência entre os diagnósticos (>1 indica dependência positiva)^{19,20}.

Sobre a modelagem de tópicos, aplicou-se o algoritmo LDA^{8,21} e os dados foram estruturados no formato atributo-valor. Além disso, utilizou-se a métrica *Perplexity*^{8,22}, para determinar o número ótimo de tópicos, testando-se de 2 a 15 tópicos. Essa abordagem foi fundamentada em estudos anteriores com dados de saúde^{23,24}. O número final de 9 tópicos foi selecionado com base na estabilização da perplexidade. O código fonte utilizado para todas as etapas de preparação dos dados e aplicação dos algoritmos de Apriori e LDA foi disponibilizado em <https://doi.org/10.5281/zenodo.17872231>.

Resultados

Os resultados obtidos a partir da análise dos dados do SIH-SUS, com aplicação de técnicas de aprendizado de máquina não supervisionado, revelam padrões relevantes de internações hospitalares por violência contra mulheres no Brasil entre 2008 e 2023. A seção está estrutu-

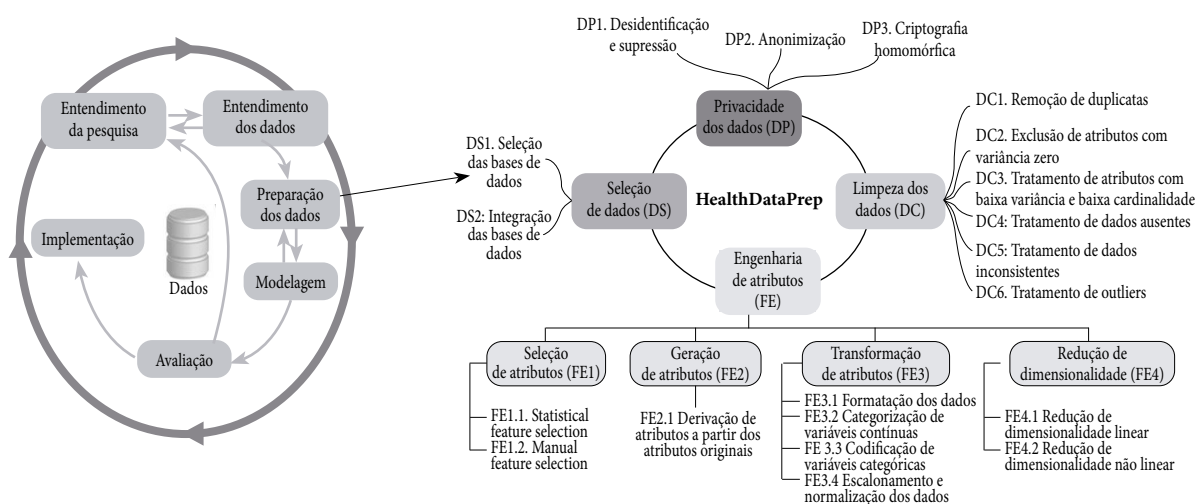


Figura 1. Utilização do framework HRSP-AI¹⁵ para nortear o processo do projeto de IA e do framework HealthDataPrep para preparação de dados não supervisionados.

rada em três partes: 1. Características gerais das internações; 2. Identificação de regras de associação entre diagnósticos por meio do algoritmo Apriori; e 3. Modelagem de tópicos com o algoritmo LDA, a fim de descrever perfis latentes das internações. Cada abordagem permitiu evidenciar diferentes dimensões do fenômeno da violência de gênero no contexto hospitalar, com implicações para vigilância em saúde, formulação de políticas públicas e práticas clínicas sensíveis à complexidade dos casos.

Características gerais das internações

De 193.455.105 internações no SUS entre 2008-2023, 828.264 (0,43%) corresponderam a códigos de violência, sendo 144.140 (17,4%) em mulheres. Após aplicação dos critérios de idade, obteve-se 90.798 casos de mulheres de 20-59 anos.

A taxa de internações por violência contra mulheres estabilizou após 2012, oscilando entre 2,6 e 3,2 por 100.000 habitantes. Em valores absolutos, São Paulo (17.963 casos), Bahia (13.893) e Minas Gerais (10.184) concentraram 46% do total nacional. Considerando-se taxas populacionais, Rio Grande do Norte apresentou a maior taxa média (17,49/100.000 habitantes), seguido pelo Pará (6,49/100.000).

A distribuição etária mostrou maior concentração na faixa de 20-30 anos, com declínio progressivo com o avanço da idade. Quanto à raça/cor, 35,57% declararam-se pardas, 20,49% brancas e 38,29% não informaram. A maioria das internações ocorreu em caráter de urgência (81,87%) e em leitos cirúrgicos (67,17%) (Tabela 1).

Regras de associação entre diagnósticos

O algoritmo Apriori identificou nove regras de associação significativas com os critérios estabelecidos. A regra com maior confiança associou traumatismo de músculo e tendão no punho/mão (S66) com agressão por objeto cortante/penetrante (X99), apresentando *Support* de 1,73%, *Confidence* de 69,5% e *Lift* de 3,14 (Quadro 1).

Outras associações relevantes incluíram:

- Fratura do crânio/face (S02) – Agressão por força corporal (Y04): *Confidence* 61,05%, *Lift* 2,99.
- Ferimento do tórax (S21) – Agressão por objeto cortante (X99): *Confidence* 59,01%, *Lift* 2,67.

- Traumatismo intratorácico (S27) – Agressão por objeto cortante (X99): *Confidence* 57,28%, *Lift* 2,59.

Os resultados evidenciaram padrões consistentes entre tipos específicos de lesões e mecanismos de agressão, sugerindo correlações entre instrumentos utilizados e lesões resultantes.

Perfis de internações por modelagem de tópicos

A análise da perplexidade indicou nove como o número ótimo de tópicos, com alteração significativa até esse ponto e variações inferiores a 1% nos valores subsequentes, sugerindo uma estabilização relativa (Figura 2). Cada tópico identificado corresponde a um perfil recorrente de atendimento hospitalar relacionado à violência contra mulheres, caracterizado por combinações específicas dos pares atributo-valor mais frequentes. Os nove perfis identificados apresentaram distribuições distintas (Quadro 2).

Tópico 7 (17,2% dos casos): Mulheres jovens (20-29 anos) em atendimento cirúrgico urgente por politraumatismos ou toracostomia com drenagem pleural, evoluindo para alta melhorada. Predominância de raça/cor parda, com diagnósticos de agressão por objeto cortante (X99) ou traumatismo intratorácico (S27).

Tópico 6 (13,5%): Mulheres internadas em São Paulo e Minas Gerais, em leito clínico, com caráter de urgência, submetidas a procedimentos relacionados a traumatismos de localização não especificada com diagnósticos de agressão por força corporal (Y04) ou ferimentos múltiplos (T01).

Tópico 3 (10,8%): Mulheres com faixa etária mais elevada (40-59 anos), majoritariamente brancas, residentes em São Paulo e Santa Catarina, internadas em leitos cirúrgicos após agressão por força corporal (Y04) ou objeto contundente (Y00).

Tópico 5 (5,5%): Mulheres pardas, de São Paulo e Ceará, vítimas de agressão por arma de fogo (X94) ou exposição a substâncias tóxicas (T65), submetidas a tratamento por envenenamento ou a procedimentos cirúrgicos de emergência, com necessidade de UTI e evolução para óbito.

Os demais tópicos representaram outros perfis com características geográficas, etárias e clínicas diversas, demonstrando a heterogeneidade dos casos de violência contra mulheres no sistema hospitalar.

Tabela 1. Características sociodemográficas e clínicas das internações por violência contra mulheres (20-59 anos) no Brasil, 2008-2023 (n=90.798).

Variável	Categoria	n	%
Faixa etária	20-29 anos	31.258	34,42%
	30-39 anos	27.637	30,44%
	40-49 anos	19.256	21,21%
	50-59 anos	12.647	13,93%
Raça/Cor	Parda	32.297	35,57%
	Branca	18.609	20,49%
	Preta	3.965	4,37%
	Amarela	897	0,99%
	Indígena	260	0,29%
	Ignorado/Não informado	34.770	38,29%
Caráter da internação	Urgência	74326	81,87%
	Outros	8316	9,15%
	Eletiva	8156	8,98%
Tipo de leito	Cirúrgico	60.990	67,17%
	Clínico	28.031	30,87%
	Outros	1.777	1,96%
Uso de UTI	Sim	84239	92,78%
	Não	6559	7,22%
5 UFs com maior número	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%

Fonte: Sistema de Informações Hospitalares do SUS (SIH-SUS).

Quadro 1. Resultados das regras de padrões de associação entre diagnósticos relacionados à violência em internações no SUS.

Antecedentes	Consequentes	Support*	Confidence	Lift
S66 - Traumatismo de músculo e tendão ao nível do punho e mão	X99 - Agressão por meio de objeto cortante ou penetrante	1,73%	69,50%	3,14
S02 - Fratura do crânio e dos ossos da face	Y04 - Agressão por meio de força corporal	4,74%	61,05%	2,99
S21 - Ferimento do tórax	X99 - Agressão por meio de objeto cortante ou penetrante	1,42%	59,01%	2,67
S27 - Traumatismo de outros órgãos intratorácicos e dos não especificados	X99 - Agressão por meio de objeto cortante ou penetrante	2,88%	57,28%	2,59
S31 - Ferimento do abdome, do dorso e da pelve	X99 - Agressão por meio de objeto cortante ou penetrante	1,50%	52,27%	2,36
S06 - Traumatismo intracraniano	Y04 - Agressão por meio de força corporal	4,08%	45,25%	2,21
S36 - Traumatismo de órgãos intra-abdominais	X99 - Agressão por meio de objeto cortante ou penetrante	2,27%	44,18%	2,00
T01 - Ferimentos envolvendo múltiplas regiões do corpo	X99 - Agressão por meio de objeto cortante ou penetrante	1,62%	42,54%	1,92
S52 - Fratura do antebraço	Y04 - Agressão por meio de força corporal	1,81%	31,30%	1,53

Fonte: Autores, 2025.

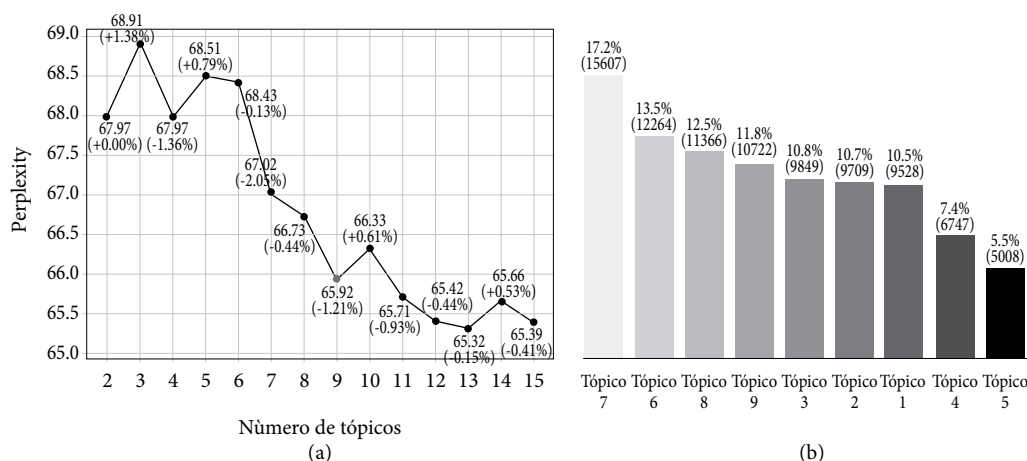


Figura 2. (a) Análise da métrica *Perplexity* para Diferentes Números de Tópicos; (b) Distribuição dos tópicos das interações por violência contra mulheres.

Fonte: Autores, 2025.

Discussão

Este estudo realizou uma análise abrangente das interações por violência contra mulheres no Brasil entre 2008 e 2023 utilizando técnicas de aprendizado de máquina não supervisionado. Os achados revelam padrões em três dimensões principais: as características gerais das interações, incluindo distribuição etária, raça/cor, tipo de leito e padrões regionais, os padrões de associação entre diagnósticos e os perfis de interações. Essa abordagem permite compreender a heterogeneidade dos casos, identificar grupos de risco e subsidiar políticas públicas, planejamento de serviços de saúde e estratégias de prevenção com destaque na atenção primária à saúde (APS).

A estabilização das taxas de internação a partir de 2012 pode ser interpretada como reflexo de avanços normativos, como a Lei Maria da Penha (Lei 11.340/2006), e da expansão dos serviços da Rede de Atenção à Mulher em Situação de Violência, especialmente os Centros de Referência e os serviços hospitalares de referência para vítimas de violência sexual. No entanto, estudos sinalizam que parte expressiva da violência contra mulheres permanece invisível no sistema de saúde, seja por subnotificação, seja por subnotificação, seja por rotulação inadequada dos códigos diagnósticos no SIH-SUS ou pelo redirecionamento das vítimas a serviços de

urgência e pronto-atendimento que não exigem internação^{25,26}.

As disparidades regionais observadas, com destaque para o Rio Grande do Norte e o Pará, sugerem uma complexa interação entre prevalência da violência, cobertura assistencial, cultura de notificação e acesso a serviços. Pesquisa recente sobre a distribuição geográfica da violência interpessoal no Brasil reforça a existência de “bolsões de alta incidência” relacionados à vulnerabilidade social, à fragilidade das redes de proteção e à impunidade dos agressores²⁷⁻²⁹.

As regras de associação identificadas por meio do algoritmo Apriori são particularmente relevantes do ponto de vista clínico-forense. A forte associação entre traumatismo de punho/mão e agressão por objeto cortante (*Confidence* 69,5%; *Lift* 3,14) pode refletir movimentos de defesa instintivos durante o ataque, como também já descrito em estudos de medicina legal³⁰. Similarmente, a correlação entre fraturas craniofaciais e agressão por força corporal (*Lift* 2,99) reforça padrões reconhecidos de agressão direta e repetida, frequentemente associados a situações de violência doméstica grave³¹.

A modelagem LDA permitiu identificar nove perfis latentes de interações, revelando a heterogeneidade dos casos e evidenciando para a necessidade de abordagens assistenciais diferenciadas. O perfil mais prevalente, de mulheres jovens com politraumatismos decorren-

Quadro 2. Tópicos com os termos mais relevantes que compõem o perfil de internações por violência contra mulheres.

	Faixa Etária	Raça/Cor	UF	Leito	Caráter	CIDs	Procedimento	Evolução	UTI
Tópico 1	20-49	Parda	SP	Clínico	Urgência	S06	Tratamento conservador de traumatismo cranioencefálico grau leve/médio	Alta melhorado	Não
Tópico 2	20-49	Parda e Não informado	PA	Cirúrgico	Urgência	S52, X99, Y04, S66	-	Alta melhorado	Não
Tópico 3	40-59	Branca	SP e SC	Cirúrgico	Urgência, Eletivo e Outros	Y04, Y00	-	Alta melhorado e Alta com previsão de retorno para acompanhamento	Não
Tópico 4	20-59	Não informado	RN	Cirúrgico	Urgência e Eletivo	Y08, S82	Tratamento com cirurgias múltiplas	Alta curado	Não
Tópico 5	20-39	Parda	SP e CE	Clínico e Cirúrgico	Urgência	X94, T65	Tratamento por intoxicação ou Envenenamento Atendimento cirúrgico de urgência	Óbito com Declaração de Óbito fornecida pelo IML	Sim
Tópico 6	20-39	Parda e Não informado	SP e MG	Clínico	Urgência	X99, Y04, T01	Tratamento de traumatismos de localização não especificada	Alta melhorado e Transferência para outro estabelecimento	Não
Tópico 7	20-29	Parda e Não informado	-	Cirúrgico	Urgência	X99, S27	Tratamento cirúrgico em politraumatizado Toracostomia com drenagem pleural	Alta melhorado	Não
Tópico 8	20-39	Parda, Branca e Não informado	SP	Cirúrgico	Urgência	S02, Y04, X99, S31	-	Alta melhorado e Alta com previsão de retorno para acompanhamento	Não
Tópico 9	30-59	Parda	BA	Clínico e Cirúrgico	Urgência	Y09, S82, Y00	-	Alta melhorado	Não

Fonte: Autores, 2025.

tes de agressão por objetos cortantes, levanta preocupações sobre a gravidade dos episódios de violência e suas repercussões físicas e emocionais. Estudos qualitativos têm demonstrado que mulheres jovens tendem a apresentar maior resistência em procurar ajuda formal, o que pode atrasar intervenções preventivas e agravar os desfechos clínicos^{32,33}.

A partir da análise dos perfis, torna-se possível propor a estratificação do risco de reinci-

dência, de agravamento clínico e de vulnerabilidade social. Iniciativas como a implantação de núcleos hospitalares de atendimento integral à mulher em situação de violência, e protocolos de notificação ativa e vigilância sentinela podem se beneficiar desses achados, aprimorando a articulação entre vigilância epidemiológica, assistência hospitalar e redes de proteção social^{34,35}.

Internacionalmente, estudos^{36,37} têm demonstrado o potencial de técnicas de aprendi-

zado de máquina na identificação de padrões ocultos de violência de gênero e na predição de risco, como evidenciado nas análises com uso de big data hospitalar ou de saúde pública, combinados com algoritmos de associação e clusterização^{36,37}. Tais experiências reforçam a importância de investimentos em infraestrutura de dados e interoperabilidade nos sistemas de saúde, elementos ainda frágeis no contexto brasileiro^{38,39}.

Nesse sentido, as recomendações da WHO⁷ para a transformação digital na saúde, especialmente no campo da Ciência de Dados, ressaltam que essa mudança deve ir além da mera digitalização de registros, envolvendo a reorganização sistêmica dos fluxos de informação com foco no usuário. Entre as propostas do documento, destaca-se a criação dos *person-centred point of service systems* (sistemas de ponto de serviço centrados na pessoa) e a incorporação de funcionalidades de apoio à decisão clínica, predição e vigilância automatizada são aspectos prioritários. A aplicação de algoritmos para identificação de padrões de violência hospitalar, como neste estudo, está alinhada à proposta da WHO⁷ de uso de tecnologias para reconhecer populações vulneráveis, identificar riscos e promover respostas personalizadas e intersetoriais em saúde.

Adicionalmente, o documento da WHO⁷ recomenda que, para garantir a efetividade dessas ferramentas, é essencial mapear os processos de trabalho e alinhar os requisitos funcionais às realidades locais, garantindo que os sistemas digitais apoiem tanto os profissionais de saúde quanto os gestores na tomada de decisão. O presente estudo contribui ao fornecer conhecimento sobre perfis clínico-epidemiológicos, podendo subsidiar em discussões futuras sobre aprimoramento de sistemas de informação e de apoio à decisão.

Apesar da robustez da base de dados e da metodologia aplicada, este estudo apresenta limitações. Tendências temporais e diferenças regionais podem refletir mudanças nos registros, no acesso e na oferta de serviços, e não necessariamente variações reais da violência. A subnotificação e o registro inadequado de códigos diagnósticos podem comprometer a validade interna. Além disso, o estudo abrange apenas internações hospitalares, não incluindo atendimentos em pronto-atendimento, na atenção primária ou casos que não chegam ao sistema de saúde, o que limita a validade externa e impede captar a magnitude mais ampla do fenômeno.

Em conclusão, este estudo caracterizou internações por violência contra mulheres no Brasil entre 2008 e 2023 e identificou padrões de associação entre diagnósticos e perfis clínico-demográficos por meio de técnicas não supervisionadas. Os resultados evidenciam que as internações se concentram sobretudo em mulheres jovens, pardas, atendidas em caráter de urgência e frequentemente submetidas a procedimentos cirúrgicos, com destaque para a forte associação entre lesões cortantes e politraumatismos. A heterogeneidade dos perfis identificados demonstra que a violência gera desfechos clínicos distintos, sugerindo necessidades assistenciais diferenciadas e grupos com maior vulnerabilidade.

Os achados deste estudo estão alinhados a marcos nacionais^{40,41} e internacionais^{6,7,42} que reconhecem a violência contra mulheres como um problema de saúde pública complexo e multifatorial. Ao identificar e organizar padrões clínico-epidemiológicos pouco explorados, o estudo contribui para preencher lacunas na literatura e oferece subsídios conceituais relevantes para discussões futuras sobre aprimoramento da vigilância, qualificação das bases de dados e desenvolvimento de ferramentas analíticas em saúde.

Colaboradores

AM Shimaoka, AC Silva Junior, JM Duarte, KCN Areco, MB Ribeiro, LM Araujo, DDG Nascimento, ME Salvador e P Bandiera-Paiva contribuíram com a concepção do estudo, análise e interpretação dos dados, redação do artigo, revisão crítica relevante do conteúdo intelectual e aprovação da versão final a ser publicada.

Declaração de disponibilidade de dados

As fontes de dados utilizados na pesquisa estão indicadas no corpo do artigo.

Referências

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-CI)*. 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 JJS, 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/transferecia-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 Saude 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 de 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.

Artigo apresentado em 02/08/2025

Aprovado em 26/01/2026

Versão final apresentada em 28/01/2026

Editores-chefes: Maria Cecília de Souza Minayo, Romeu Gomes, Antônio Augusto Moura da Silva, Vania de Matos Fonseca