ReGeo ISSN: 2177-3246

MACHINE LEARNING APLICADO À SAÚDE COLETIVA: ANÁLISE DE PADRÕES DE VIOLÊNCIA CONTRA A MULHER NO BRASIL POR APRENDIZAGEM DE MÁQUINA NÃO SUPERVISIONADA - POTENCIAL ANALÍTICO E LIMITES ÉTICOS DO USO DE DADOS

MACHINE LEARNING APPLIED TO PUBLIC HEALTH: ANALYSIS OF PATTERNS OF VIOLENCE AGAINST WOMEN IN BRAZIL THROUGH UNSUPERVISED MACHINE LEARNING - ANALYTICAL POTENTIAL AND ETHICAL LIMITS OF DATA USE

APRENDIZAJE AUTOMÁTICO APLICADO A LA SALUD PÚBLICA: ANÁLISIS DE PATRONES DE VIOLENCIA CONTRA LA MUJER EN BRASIL MEDIANTE APRENDIZAJE AUTOMÁTICO NO SUPERVISADO - POTENCIAL ANALÍTICO Y LÍMITES ÉTICOS DEL USO DE DATOS

https://doi.org/10.56238/revgeov17n8-005

Data de submissão: 19/06/2026 **Data de publicação:** 03/08/2026

Isabela Molinero de Paula
Ensino superior incompleto
Lattes: http://lattes.cnpq.br/9053521715251778

Rycherd Dheurymy da Silva
Mestrando em Computação Aplicada
Instituição: IFMA
Lattes: http://lattes.cnpq.br/9062012817558431

Otto Guilherme Gerstenberger Junior
Pós-Doutor em Direito, Instituição e Negócios
Instituição: UFF
Lattes: http://lattes.cnpq.br/6920563188820015

Bruno Viana Gonçalves
Mestre em Ciências
Instituição: UNICAMP
Lattes: http://lattes.cnpq.br/1336507248212992

Jônatas da Silva Bezerra
Especialista
Lattes: http://lattes.cnpq.br/1247484269673001

RESUMO

A violência contra a mulher constitui grave problema de saúde coletiva no Brasil, cuja compreensão depende de vastos conjuntos de dados de notificação e mortalidade cujo potencial analítico permanece subaproveitado. Este artigo, elaborado como revisão integrativa da literatura, examina as possibilidades e os limites da aplicação da aprendizagem de máquina não supervisionada à análise de padrões de violência contra a mulher, confrontando seu potencial analítico com os riscos éticos do uso de dados. Foram analisados dados oficiais sobre a magnitude da violência, estudos que caracterizam as notificações no país e revisões sobre a aplicação da aprendizagem de máquina ao enfrentamento da violência doméstica e de gênero, além de literatura sobre viés algorítmico, proteção de dados e ética

da inteligência artificial. Argumenta-se que as técnicas não supervisionadas, ao agrupar registros sem categorias predefinidas, permitem descobrir perfis de vulnerabilidade, padrões de vitimização e agrupamentos de risco que escapam à análise convencional, oferecendo subsídios à vigilância e à priorização de ações. Esse potencial, contudo, é indissociável de riscos substanciais: a reprodução de vieses presentes nos dados, a opacidade dos modelos, a possibilidade de reidentificação de vítimas e as exigências da Lei Geral de Proteção de Dados quanto ao tratamento de dados sensíveis. Conclui-se que a aprendizagem não supervisionada é ferramenta promissora para a saúde coletiva, desde que sua adoção seja acompanhada de transparência metodológica, governança de dados e salvaguardas éticas que protejam as populações que se pretende beneficiar.

Palavras-chave: Violência Contra a Mulher. Aprendizado de Máquina. Saúde Pública.

ABSTRACT
Violence against women is a serious public health problem in Brazil, whose understanding depends on vast sets of notification and mortality data whose analytical potential remains underutilized. This article, prepared as an integrative literature review, examines the possibilities and limits of applying unsupervised machine learning to the analysis of patterns of violence against women, confronting its analytical potential with the ethical risks of data use. Official data on the magnitude of violence, studies characterizing notifications in the country, and reviews of the application of machine learning to combat domestic and gender violence were analyzed, as well as literature on algorithmic bias, data protection, and artificial intelligence ethics. It is argued that unsupervised techniques, by grouping records without predefined categories, allow for the discovery of vulnerability profiles, victimization patterns, and risk clusters that escape conventional analysis, providing support for surveillance and action prioritization. This potential, however, is inseparable from substantial risks: the reproduction of biases present in the data, the opacity of models, the possibility of victim reidentification, and the requirements of the General Data Protection Law regarding the treatment of sensitive data. It is concluded that unsupervised learning is a promising tool for public health, provided that its adoption is accompanied by methodological transparency, data governance, and ethical safeguards that protect the populations intended to benefit.

Keywords: Violence Against Women. Machine Learning. Public Health.

RESUMEN
La violencia contra la mujer constituye un grave problema de salud pública en Brasil, cuya comprensión depende de vastos conjuntos de datos de notificación y mortalidad cuyo potencial analítico sigue sin aprovecharse. Este artículo, elaborado como revisión integrativa de la literatura, examina las posibilidades y límites de la aplicación del aprendizaje automático no supervisado al análisis de patrones de violencia contra la mujer, confrontando su potencial analítico con los riesgos éticos del uso de datos. Se analizaron datos oficiales sobre la magnitud de la violencia, estudios que caracterizan las notificaciones en el país y revisiones sobre la aplicación del aprendizaje automático para combatir la violencia doméstica y de género, así como literatura sobre sesgo algorítmico, protección de datos y ética de la inteligencia artificial. Se argumenta que las técnicas no supervisadas, al agrupar registros sin categorías predefinidas, permiten descubrir perfiles de vulnerabilidad, patrones de victimización y grupos de riesgo que escapan al análisis convencional, proporcionando apoyo para la vigilancia y la priorización de acciones. Este potencial, sin embargo, es inseparable de riesgos sustanciales: la reproducción de sesgos presentes en los datos, la opacidad de los modelos, la posibilidad de reidentificación de víctimas y los requisitos de la Ley General de Protección de Datos en cuanto al tratamiento de datos sensibles. Se concluye que el aprendizaje no supervisado es una herramienta prometedora para la salud pública, siempre que su adopción vaya acompañada de

transparencia metodológica, gobernanza de datos y salvaguardas éticas que protejan a las poblaciones destinadas a beneficiarse.

Palabras clave: Violencia Contra la Mujer. Aprendizaje Automático. Salud Pública.

1 INTRODUÇÃO

A violência contra a mulher é reconhecida, há décadas, como grave problema de saúde pública e como violação de direitos humanos, com repercussões que se estendem da mortalidade por agressão ao adoecimento físico e mental das vítimas. No Brasil, a dimensão do fenômeno é expressiva: o Atlas da Violência registrou 3.642 homicídios de mulheres em 2024, o equivalente a uma taxa de 3,4 mortes por 100 mil mulheres, com marcada desigualdade racial — as mulheres negras responderam por 67,5% das vítimas (IPEA; FBSP, 2024). Esses números, ainda que indiquem redução na última década, convivem com a persistência da violência praticada no ambiente doméstico e com significativa subnotificação, o que sinaliza que a face registrada do problema é apenas parte de sua extensão real.

O enfrentamento desse problema apoia-se, em larga medida, em sistemas de informação que produzem grandes volumes de dados. As notificações de violência interpessoal registradas no Sistema de Informação de Agravos de Notificação e os registros de mortalidade constituem bases extensas cujo aproveitamento analítico é ainda limitado. Estudo que caracterizou as notificações de violência contra mulheres em contextos rurais entre 2011 e 2020 analisou 79.229 registros e evidenciou o predomínio da violência física, seguida da psicológica e da sexual, praticadas majoritariamente na residência e por parceiros íntimos (STOCHERO; PINTO, 2024). A riqueza desses dados contrasta com a predominância de análises descritivas, que nem sempre revelam os padrões latentes capazes de orientar políticas mais precisas.

É nesse ponto que a aprendizagem de máquina se apresenta como recurso analítico de grande potencial. Ramo da inteligência artificial dedicado a extrair conhecimento a partir de dados, a aprendizagem de máquina permite identificar regularidades, agrupar registros semelhantes e antecipar desfechos em conjuntos de informação demasiado extensos ou complexos para a análise manual. Entre suas modalidades, a aprendizagem não supervisionada distingue-se por não depender de categorias previamente definidas: em vez de aprender a reproduzir rótulos fornecidos por especialistas, ela descobre estruturas ocultas nos próprios dados, agrupando casos por semelhança e revelando perfis e padrões que não haviam sido antecipados.

Aplicada à violência contra a mulher, essa capacidade abre possibilidades relevantes para a saúde coletiva — a identificação de perfis de vulnerabilidade, o reconhecimento de padrões de vitimização e a priorização de grupos de maior risco. Ao mesmo tempo, o uso de dados sensíveis sobre um fenômeno que atinge pessoas em situação de vulnerabilidade suscita questões éticas de primeira ordem, relativas ao viés, à privacidade e à possibilidade de que a própria tecnologia reproduza ou aprofunde as desigualdades que se pretende combater. O problema que orienta este trabalho reside precisamente nessa tensão entre o potencial analítico e os limites éticos do uso de dados.

O objetivo geral deste artigo é analisar as possibilidades e os limites da aplicação da aprendizagem de máquina não supervisionada à análise de padrões de violência contra a mulher no Brasil, sob a perspectiva da saúde coletiva. Como objetivos específicos, busca-se dimensionar o problema e a produção de dados que o documenta; descrever os fundamentos da aprendizagem não supervisionada; sintetizar suas aplicações ao enfrentamento da violência; e discutir os limites éticos de seu uso. A relevância do estudo está em oferecer uma reflexão que articule a promessa metodológica das novas técnicas às exigências éticas e jurídicas que devem governar o tratamento de dados em saúde.

Cabe situar, ainda na introdução, a dupla natureza do objeto deste estudo, que se coloca na interseção de dois campos em rápida transformação. De um lado, a saúde coletiva, que há muito incorporou a vigilância baseada em dados como instrumento de compreensão e de resposta aos agravos, e que dispõe, no caso da violência, de sistemas nacionais de notificação e de mortalidade que acumulam registros ao longo de décadas. De outro, a ciência de dados, cujo avanço recente tornou acessíveis técnicas antes restritas a especialistas, ampliando as possibilidades de análise de grandes volumes de informação. O encontro entre esses campos define um território novo, cujas promessas e cujos riscos ainda estão sendo delineados, e que exige uma reflexão que não se limite ao entusiasmo técnico nem à recusa apriorística.

Esse território, ademais, não é axiologicamente neutro. Trata-se de aplicar ferramentas poderosas a dados sobre um fenômeno que atinge pessoas em situação de vulnerabilidade e que expressa desigualdades estruturais de gênero e de raça. A escolha de quais perguntas fazer aos dados, de quais grupos priorizar e de quais usos autorizar não é uma decisão puramente técnica, mas uma decisão política e ética, que envolve valores e que produz consequências concretas sobre a vida de pessoas reais. Reconhecer essa dimensão é condição para uma reflexão honesta, que não se refugie na aparência de objetividade que os números e os algoritmos frequentemente projetam.

A pertinência desse debate é reforçada pelo momento histórico. A difusão acelerada de aplicações de inteligência artificial em diversos setores, inclusive na saúde e na segurança pública, torna previsível que também o enfrentamento da violência contra a mulher venha a incorporar tais ferramentas, seja em iniciativas governamentais, seja em pesquisas acadêmicas, seja em soluções oferecidas por empresas. Antecipar-se a esse movimento, refletindo sobre suas possibilidades e seus limites antes que ele se consolide, é oportunidade de orientar a incorporação da tecnologia em bases eticamente sólidas, em vez de reagir tardiamente a usos já estabelecidos. É a essa antecipação reflexiva que este trabalho pretende contribuir.

2 METODOLOGIA

Adotou-se o método da revisão integrativa da literatura, que permite reunir, analisar e sintetizar de forma sistematizada os resultados de estudos sobre determinado tema, contribuindo para a construção do conhecimento e para a fundamentação da prática (SOUZA; SILVA; CARVALHO, 2010). A opção justifica-se pela natureza interdisciplinar do objeto, que articula dados epidemiológicos, métodos computacionais e questões éticas e jurídicas, exigindo um procedimento capaz de integrar fontes de naturezas distintas.

A busca foi conduzida nas bases SciELO, PubMed, MDPI e Google Acadêmico, complementada pela consulta a documentos oficiais e à legislação pertinente. Empregaram-se, isoladamente e em combinação, os descritores e termos livres: *"violência contra a mulher"*, *"aprendizagem de máquina"*, *"aprendizagem não supervisionada"*, *"agrupamento"*, *"saúde coletiva"*, *"viés algorítmico"* e *"proteção de dados"*, nos idiomas português, inglês e espanhol.

Adotaram-se como critérios de inclusão: documentos oficiais que dimensionassem a violência contra a mulher no Brasil; estudos que caracterizassem suas notificações; revisões e estudos sobre a aplicação da aprendizagem de máquina ao enfrentamento da violência; e trabalhos que discutissem os aspectos éticos e jurídicos do tratamento de dados. Priorizaram-se publicações recentes, admitindo-se referências de valor metodológico ou normativo consolidado. Foram excluídos materiais sem revisão editorial. Os dados quantitativos reproduzidos foram extraídos diretamente das fontes primárias, sem estimativas próprias.

3 RESULTADOS E DISCUSSÃO

3.1 A VIOLÊNCIA CONTRA A MULHER COMO PROBLEMA DE SAÚDE COLETIVA E A PRODUÇÃO DE DADOS

Compreender o potencial analítico da aprendizagem de máquina exige, primeiro, dimensionar o problema e a base de dados que o documenta. Os indicadores oficiais revelam um fenômeno de grande magnitude e marcada iniquidade: além do elevado número absoluto de homicídios de mulheres, chama atenção a sobre-representação das mulheres negras entre as vítimas e a persistência da violência no ambiente doméstico, que não acompanhou a redução observada em outros contextos (IPEA; FBSP, 2024). A violência letal, contudo, é apenas o extremo de um continuum que inclui agressões físicas, psicológicas e sexuais de frequência muito superior, cuja expressão mais sistemática encontra-se nos registros de notificação.

Esses registros constituem uma fonte de dados de grande valor. A análise das notificações de violência contra mulheres rurais no país documentou dezenas de milhares de registros e permitiu

caracterizar o perfil das vítimas, os tipos de violência, os locais de ocorrência e os agressores, revelando o predomínio da violência física, a centralidade da residência como local das agressões e o papel dos parceiros íntimos (STOCHERO; PINTO, 2024). A própria consistência desses achados descritivos sugere a existência de padrões estruturados nos dados — padrões que técnicas analíticas mais sofisticadas poderiam explorar de forma mais fina.

Há, contudo, uma limitação incontornável: a subnotificação. Uma parcela expressiva dos casos de violência não chega a ser registrada, seja pelo medo e pela dependência das vítimas, seja pela dificuldade de identificação por parte dos profissionais. Esse fato tem implicação direta para qualquer análise baseada em dados: os registros disponíveis não são uma amostra neutra do fenômeno, mas um recorte enviesado, que sub-representa determinados grupos e contextos. Qualquer técnica analítica — inclusive a aprendizagem de máquina — herda esse viés de origem, o que impõe cautela na interpretação de seus resultados.

A natureza enviesada dos registros merece um exame mais detido, pois dela dependem a validade e a interpretação de qualquer análise subsequente. A decisão de notificar um caso de violência resulta de uma cadeia de fatores — a disposição da vítima em relatar, o acesso a um serviço de saúde ou de segurança, a capacidade e a sensibilidade do profissional em identificar a agressão e registrá-la — de modo que o dado produzido é, ele mesmo, um artefato social. Grupos com menor acesso aos serviços, ou submetidos a maior estigma, tendem a estar sub-representados, ao passo que a maior presença institucional em determinados territórios pode inflar artificialmente a contagem de casos. O registro, portanto, não espelha o fenômeno: espelha o encontro entre o fenômeno e o sistema que o documenta.

Essa constatação tem consequência metodológica direta. Uma análise que tome os dados de notificação como retrato fiel da violência corre o risco de confundir padrões do fenômeno com padrões do próprio registro, atribuindo à realidade aquilo que é efeito da desigualdade no acesso e na identificação. Para a saúde coletiva, cujo compromisso é justamente com a equidade, esse risco é particularmente grave, pois pode conduzir a decisões que reforcem as desigualdades que se pretendia corrigir. Qualquer técnica analítica sofisticada, ao ser aplicada a esses dados, precisa incorporar essa ressalva como premissa, e não como nota de rodapé.

A dimensão da iniquidade merece um comentário adicional, por ser central à leitura do fenômeno em chave de saúde coletiva. A sobre-representação das mulheres negras entre as vítimas de homicídio, com taxa expressivamente superior à das mulheres não negras, evidencia que a violência não se distribui de forma homogênea, mas incide de modo mais intenso sobre grupos historicamente submetidos a múltiplas formas de opressão (IPEA; FBSP, 2024). Essa interseção entre gênero e raça é

justamente o tipo de padrão que uma análise atenta à estrutura dos dados deveria captar e explicitar, em vez de diluir na média geral — o que ilustra tanto o potencial das técnicas de agrupamento quanto a responsabilidade de utilizá-las para iluminar, e não para obscurecer, as desigualdades.

Essa mesma iniquidade, todavia, adverte contra usos ingênuos da análise de dados. Se os grupos mais vulneráveis são também aqueles com menor acesso aos serviços e, portanto, mais sujeitos à subnotificação, existe o risco de que a análise, ao basear-se nos registros disponíveis, subestime precisamente a violência que atinge quem está mais desprotegido. O paradoxo é revelador: os grupos que mais precisariam da proteção derivada da análise podem ser aqueles que menos aparecem nos dados que a alimentam. Enfrentar esse paradoxo exige combinar a análise dos registros com estratégias ativas de identificação e com a leitura crítica das ausências, tão informativas quanto as presenças.

3.2 APRENDIZAGEM DE MÁQUINA SUPERVISIONADA E NÃO SUPERVISIONADA

A aprendizagem de máquina divide-se, classicamente, em modalidades distintas conforme a natureza da tarefa e a disponibilidade de rótulos nos dados. Na aprendizagem supervisionada, o algoritmo é treinado com exemplos previamente classificados e aprende a reproduzir essa classificação em novos casos; na não supervisionada, não há rótulos, e o algoritmo busca descobrir estruturas latentes nos próprios dados. Revisões sobre a aplicação dessas técnicas ao enfrentamento da violência doméstica abrangem ambas as modalidades e evidenciam o crescimento do campo (HUI; CONSTANTINO; LEE, 2023). A distinção é relevante para a saúde coletiva porque cada modalidade responde a perguntas diferentes: a supervisionada, a prever um desfecho conhecido; a não supervisionada, a explorar o desconhecido.

A aprendizagem não supervisionada, foco deste trabalho, opera fundamentalmente por meio de técnicas de agrupamento. Algoritmos como os de particionamento organizam os registros em grupos internamente homogêneos e distintos entre si, de modo que casos semelhantes fiquem reunidos. O resultado é a emergência de perfis — conjuntos de vítimas, situações ou territórios que compartilham características e que, uma vez identificados, podem ser objeto de análise e de intervenção específicas. O valor dessa abordagem está em não exigir hipóteses prévias sobre quais grupos existem: são os próprios dados que os revelam.

Mapeamentos sistemáticos da produção sobre aprendizagem de máquina aplicada à violência de gênero confirmam a diversidade de técnicas empregadas e a expansão do interesse acadêmico pelo tema, ao mesmo tempo em que apontam a predominância de determinados métodos e a necessidade de maior rigor e padronização (PINTO-MUÑOZ; ZUÑIGA-SAMBONI; ORDOÑEZ-ERAZO, 2023).

Esse panorama indica um campo em consolidação, no qual as técnicas não supervisionadas ocupam espaço crescente ao lado das abordagens preditivas.

Convém detalhar o funcionamento das técnicas de agrupamento, pois nele reside a especificidade da abordagem não supervisionada. Um algoritmo de agrupamento parte de uma medida de semelhança entre os registros — construída a partir das variáveis disponíveis — e organiza os casos de modo que a semelhança interna a cada grupo seja máxima e a semelhança entre grupos seja mínima. O número de grupos pode ser definido previamente ou inferido dos próprios dados, e a qualidade do agrupamento é avaliada por critérios estatísticos que medem sua coesão e sua separação. O resultado não é uma resposta a uma pergunta fechada, mas um mapa da estrutura latente do conjunto, que o analista então interpreta.

Essa lógica distingue-se radicalmente da abordagem preditiva. Enquanto um modelo supervisionado responde a uma pergunta do tipo "dado este perfil, qual a probabilidade de determinado desfecho?", a análise não supervisionada responde a uma pergunta anterior e mais aberta: "que perfis existem nestes dados?". Para um fenômeno como a violência contra a mulher, cuja complexidade resiste a categorizações simples, essa abertura é valiosa, pois permite que a heterogeneidade real das situações emerja dos dados em vez de ser aprisionada em categorias definidas de antemão. A técnica, nesse sentido, é tão exploratória quanto humilde: reconhece que nem tudo o que importa foi previamente nomeado.

Há, porém, uma contrapartida a essa flexibilidade. A ausência de rótulos que orientem o aprendizado torna os resultados sensíveis às escolhas do analista — quais variáveis incluir, como medi-las, como definir a semelhança, quantos grupos considerar. Diferentes escolhas produzem diferentes agrupamentos, e não há um critério único e objetivo para eleger o melhor. Essa dependência das decisões metodológicas reforça que a técnica não substitui o julgamento, mas o desloca: da interpretação dos casos para a configuração do procedimento. A transparência quanto a essas escolhas torna-se, assim, condição de validade e de reprodutibilidade da análise.

A escolha das variáveis merece atenção particular, por ser determinante dos padrões que a análise poderá revelar. Um agrupamento só pode reconhecer semelhanças e diferenças nas dimensões que lhe são fornecidas: variáveis ausentes dos dados permanecem invisíveis ao algoritmo, por mais relevantes que sejam para a compreensão do fenômeno. No caso da violência contra a mulher, dimensões como o histórico de agressões anteriores, a dependência econômica em relação ao agressor ou o acesso a redes de apoio são frequentemente mal registradas ou ausentes das bases, o que limita a riqueza dos perfis que se podem construir. A qualidade da análise, assim, é limitada a montante pela própria concepção dos instrumentos de coleta.

Essa dependência remete a uma questão mais ampla, relativa ao desenho dos sistemas de informação. Bases de dados concebidas para fins administrativos ou de mera contabilização de casos nem sempre capturam as variáveis que uma análise de padrões exigiria, o que sugere que o aproveitamento analítico da aprendizagem de máquina não começa no algoritmo, mas na definição do que se registra e de como se registra. Aprimorar os instrumentos de notificação, incorporando dimensões relevantes e padronizando seu preenchimento, é condição para que as técnicas analíticas possam expressar todo o seu potencial — o que reafirma a integração entre a gestão da informação e a capacidade de análise.

3.3 APLICAÇÕES DA APRENDIZAGEM NÃO SUPERVISIONADA À VIOLÊNCIA CONTRA A MULHER

As aplicações documentadas na literatura ilustram o potencial das técnicas não supervisionadas. Estudos de agrupamento têm sido empregados para identificar perfis de vitimização e para caracterizar os grupos de mulheres mais expostas a diferentes formas de violência, subsidiando instrumentos de apoio à decisão de autoridades locais (PINTO-MUÑOZ; ZUÑIGA-SAMBONI; ORDOÑEZ-ERAZO, 2023). Ao reunir vítimas por semelhança de características — idade, escolaridade, tipo de violência, contexto de ocorrência —, esses estudos convertem grandes bases de notificação em tipologias interpretáveis, capazes de orientar a alocação de recursos e o desenho de intervenções dirigidas.

Revisões da literatura reforçam que a aprendizagem de máquina, ao detectar mudanças sutis e padrões em dados textuais e estruturados, constitui área promissora para a compreensão e a prevenção da violência doméstica (HUI; CONSTANTINO; LEE, 2023). No caso das técnicas não supervisionadas, essa promessa manifesta-se na capacidade de revelar agrupamentos e associações que a análise convencional, orientada por hipóteses prévias, tenderia a não formular — por exemplo, a coocorrência de determinados fatores que define um perfil de risco antes despercebido.

Transpostas para o contexto brasileiro, tais aplicações encontram terreno fértil nas bases de notificação e mortalidade já existentes. O agrupamento de registros poderia, em princípio, revelar perfis regionais de vitimização, identificar territórios de concentração de casos e apoiar a vigilância epidemiológica da violência, integrando-a às demais ações da saúde coletiva. A promessa, nesse sentido, não é substituir o julgamento humano, mas ampliar a capacidade de enxergar padrões em dados cuja escala ultrapassa a análise manual.

A literatura internacional oferece exemplos concretos que ajudam a dimensionar as possibilidades. Estudos aplicaram algoritmos de agrupamento para classificar mulheres segundo sua

vulnerabilidade e risco de sofrer diferentes formas de violência, integrando os resultados a instrumentos de apoio à decisão destinados a autoridades locais, o que demonstra a viabilidade de converter a análise de padrões em ferramenta operacional (PINTO-MUÑOZ; ZUÑIGA-SAMBONI; ORDOÑEZ-ERAZO, 2023). Ainda que produzidos em contextos distintos do brasileiro, esses trabalhos evidenciam um caminho: o da tradução de padrões estatísticos em subsídios para a ação, aproximando a análise de dados da prática dos serviços.

A transposição desses exemplos ao Brasil precisa, contudo, considerar as particularidades do território e dos dados nacionais. A grande extensão do país, a heterogeneidade regional e as desigualdades no acesso aos serviços produzem bases de dados cuja estrutura reflete não apenas o fenômeno, mas também a geografia da presença do Estado. Um agrupamento aplicado a essas bases poderia revelar tanto perfis de vitimização quanto padrões de cobertura dos serviços, e distinguir uns dos outros exigiria conhecimento contextual apurado. A promessa analítica, portanto, não dispensa — antes, exige — a articulação com o saber dos profissionais e dos gestores que conhecem cada território.

3.4 POTENCIAL ANALÍTICO PARA A SAÚDE COLETIVA

O potencial da aprendizagem não supervisionada para a saúde coletiva pode ser sintetizado em algumas contribuições. A primeira é a descoberta de perfis de vulnerabilidade: ao agrupar vítimas por semelhança, a técnica permite caracterizar tipologias que orientam a formulação de políticas mais específicas do que as concebidas para uma população homogênea e indiferenciada. A segunda é o apoio à vigilância: a identificação de agrupamentos temporais e territoriais de casos pode sinalizar situações que demandam atenção prioritária, aproximando a resposta do sistema da dinâmica real do fenômeno.

A terceira contribuição é a priorização. Em um cenário de recursos limitados, a capacidade de identificar os grupos e os territórios de maior risco permite direcionar esforços de prevenção e de proteção com maior eficiência. A quarta é a geração de hipóteses: ao revelar padrões não antecipados, a análise não supervisionada pode inspirar novas perguntas de pesquisa e novas linhas de investigação, funcionando como etapa exploratória que precede e orienta estudos mais dirigidos. Em conjunto, essas contribuições sugerem que a técnica, longe de ser mera curiosidade computacional, pode integrar o instrumental analítico da saúde coletiva.

Cumpre, todavia, delimitar o alcance dessas promessas. A aprendizagem não supervisionada é uma ferramenta exploratória: os agrupamentos que produz não têm significado intrínseco, e sua interpretação depende do conhecimento substantivo do fenômeno. Um perfil identificado pelo algoritmo só se torna útil quando lido à luz da compreensão dos determinantes sociais da violência, o

que reafirma a centralidade do julgamento humano e do diálogo interdisciplinar. A técnica amplia a visão, mas não dispensa quem a interprete.

Vale sublinhar, ainda, a natureza integradora dessas contribuições. A descoberta de perfis, o apoio à vigilância, a priorização e a geração de hipóteses não são funções isoladas, mas etapas de um mesmo ciclo analítico: os perfis identificados alimentam a vigilância, que orienta a priorização, que por sua vez suscita novas hipóteses a serem investigadas. Quando integrada a esse ciclo, a aprendizagem não supervisionada deixa de ser uma técnica pontual e passa a compor um sistema de inteligência para a saúde coletiva, no qual os dados retroalimentam continuamente a compreensão e a ação. É nessa integração, mais do que em qualquer aplicação isolada, que reside seu maior potencial.

Essa visão sistêmica, porém, depende de condições institucionais que não são triviais. Exige a integração de bases de dados hoje frequentemente fragmentadas entre setores; exige capacidade técnica instalada nos serviços e nas instâncias de gestão; e exige uma cultura de uso de dados que valorize a análise como subsídio à decisão, e não como mero cumprimento burocrático. A ausência dessas condições pode fazer com que o potencial permaneça no plano da possibilidade, restrito a experiências acadêmicas isoladas sem tradução em política pública. Reconhecer o potencial, portanto, é também reconhecer a agenda institucional que sua realização pressupõe.

Um exemplo hipotético ajuda a concretizar essas contribuições. Imagine-se a aplicação de uma técnica de agrupamento ao conjunto de notificações de um estado ao longo de uma década. O procedimento poderia revelar, por exemplo, um grupo de casos caracterizado por vítimas jovens, com baixa escolaridade, agredidas por parceiros no ambiente doméstico e com notificações concentradas em determinados períodos; outro grupo poderia reunir casos de violência sexual praticada por desconhecidos em contextos distintos. Cada agrupamento sugeriria estratégias de prevenção e de proteção específicas, dificilmente formuláveis a partir de uma leitura indiferenciada do total de casos. O valor da técnica está precisamente nessa capacidade de decompor um agregado heterogêneo em perfis acionáveis.

Esse exemplo, contudo, também expõe os limites. A interpretação dos grupos exige conhecimento do contexto, sob pena de leituras equivocadas; a concentração temporal de notificações pode refletir mudanças no registro, e não no fenômeno; e a caracterização detalhada das vítimas, se mal protegida, aproxima-se perigosamente da possibilidade de reidentificação. O mesmo exemplo que ilustra o potencial ilustra, portanto, a inseparabilidade entre a promessa analítica e os cuidados éticos, que devem caminhar juntos em qualquer aplicação real. É essa inseparabilidade que a seção seguinte examina em detalhe.

3.5 LIMITES ÉTICOS: VIÉS, OPACIDADE, PRIVACIDADE E PROTEÇÃO DE DADOS

Ao potencial analítico contrapõem-se limites éticos que não podem ser tratados como secundários. O primeiro é o do viés. Os modelos de aprendizagem de máquina aprendem a partir dos dados que lhes são fornecidos, de modo que, se esses dados refletem preconceitos ou desigualdades existentes, o modelo tende a reproduzi-los, produzindo resultados que podem prejudicar justamente os grupos mais vulneráveis (LEONEL *et al.*, 2024). No caso da violência contra a mulher, o viés de subnotificação e as desigualdades no acesso aos serviços podem fazer com que determinados grupos apareçam sub-representados nos dados, e que os padrões descobertos reflitam mais as lacunas do registro do que a realidade do fenômeno.

O segundo limite é o da opacidade. Muitos sistemas de inteligência artificial funcionam como caixas-pretas, cujos critérios de decisão são difíceis de auditar, o que dificulta a identificação de vieses e compromete a transparência exigida de qualquer instrumento de política pública (LEONEL *et al.*, 2024). A dificuldade de explicar por que um modelo agrupou determinados casos ou sinalizou determinado risco coloca em xeque a prestação de contas e a possibilidade de contestação, valores caros ao Estado democrático de direito.

O terceiro limite, particularmente sensível, é o da privacidade e da proteção de dados. Dados sobre violência são dados pessoais sensíveis, cujo tratamento é submetido, no Brasil, às exigências da Lei Geral de Proteção de Dados, que estabelece princípios como a finalidade, a necessidade, a adequação e a não discriminação, bem como salvaguardas específicas para os dados sensíveis (BRASIL, 2018). A análise de grandes bases de notificação, ainda que anonimizadas, comporta o risco de reidentificação das vítimas, sobretudo quando o cruzamento de variáveis permite singularizar indivíduos em contextos de pequena população — risco que exige medidas técnicas e organizacionais rigorosas de proteção.

Esses limites não invalidam o uso da técnica, mas condicionam-no. A adoção da aprendizagem não supervisionada em saúde coletiva exige transparência metodológica, que permita auditar os procedimentos; governança de dados, que assegure o tratamento conforme a lei; e uma postura crítica quanto aos vieses de origem, que impeça a naturalização de resultados enviesados. Sem essas salvaguardas, a promessa de ampliar a proteção das vítimas pode converter-se em seu contrário, expondo-as a novos riscos decorrentes do próprio uso de seus dados.

Há, por fim, um limite de ordem mais ampla, relativo ao lugar da tecnologia no enfrentamento de um problema social. A literatura sobre inteligência artificial em saúde adverte contra a tentação de tratar problemas estruturais como se fossem problemas técnicos, solucionáveis por ferramentas computacionais, quando suas raízes estão em desigualdades sociais que nenhum algoritmo dissolve

(LEONEL *et al.*, 2024). A violência contra a mulher é fenômeno enraizado em relações históricas de poder e desigualdade de gênero, e o risco de que a sofisticação analítica desloque a atenção das causas para os sintomas — dos determinantes para os padrões — é real. A técnica pode iluminar o problema, mas não substitui as políticas que o enfrentam em sua origem.

Esse conjunto de limites recomenda uma postura que não é de rejeição, mas de vigilância crítica. Cada aplicação deve ser interrogada quanto à qualidade e à representatividade de seus dados, quanto à transparência de seus métodos, quanto às salvaguardas de proteção que adota e quanto ao lugar que ocupa em uma estratégia mais ampla de enfrentamento. Essa interrogação permanente, longe de inviabilizar o uso da técnica, é o que o torna legítimo, ao subordinar a ferramenta às finalidades que deveriam justificá-la — a proteção das vítimas e a redução das desigualdades.

Cabe registrar que a discussão ética não se restringe ao momento da análise, mas percorre todo o ciclo de vida dos dados — da coleta ao descarte. A legislação de proteção de dados estrutura-se justamente sobre essa lógica de ciclo, estabelecendo princípios que incidem sobre cada etapa do tratamento e conferindo direitos aos titulares dos dados (BRASIL, 2018). No caso da violência contra a mulher, esse enfoque de ciclo implica cuidados que vão da forma como o profissional registra a notificação, passando pelo armazenamento seguro e pela anonimização adequada, até a definição de quem pode acessar os dados e para quais finalidades. A responsabilidade ética, assim, é distribuída ao longo de toda a cadeia, e não concentrada em um único ponto.

Essa perspectiva de ciclo tem implicação prática relevante: a proteção efetiva das vítimas depende menos de uma medida isolada do que da coerência entre as diversas etapas. De pouco vale uma anonimização sofisticada se o acesso aos dados é irrestrito, ou uma política de acesso rigorosa se o registro original expõe informações desnecessárias. A governança de dados, nesse sentido, é um sistema, cujo elo mais frágil determina o nível de proteção do conjunto. Construir esse sistema exige articulação entre os setores de saúde, de segurança e de assistência, que hoje frequentemente operam de forma fragmentada, com bases de dados que não se comunicam e com culturas distintas de tratamento da informação.

Há, ainda, uma dimensão de responsabilização que a discussão ética não pode ignorar. Quando decisões que afetam a vida das pessoas passam a apoiar-se, ainda que parcialmente, em análises automatizadas, coloca-se a questão de quem responde por eventuais erros ou danos. A opacidade de certos modelos agrava essa dificuldade, ao tornar árduo identificar as causas de uma decisão equivocada. Para o campo da saúde coletiva, comprometido com a prestação de contas à sociedade, a definição clara de responsabilidades e a manutenção da supervisão humana sobre as decisões são

exigências incontornáveis, que impedem que a delegação a sistemas automatizados se converta em diluição da responsabilidade.

4 SÍNTESE E DISCUSSÃO

A integração das fontes analisadas permite responder ao objetivo central e identificar convergências, tensões e lacunas. O quadro que emerge é o de uma ferramenta de considerável potencial analítico, cuja adoção responsável depende do enfrentamento de riscos éticos igualmente consideráveis.

4.1 CONVERGÊNCIAS

A primeira convergência é o reconhecimento do potencial da aprendizagem de máquina para a compreensão e a prevenção da violência. Tanto as revisões da literatura quanto os estudos aplicados coincidem em identificar nas técnicas computacionais um recurso capaz de revelar padrões e de subsidiar decisões, ampliando o alcance da análise para além do que as abordagens descritivas convencionais permitem (HUI; CONSTANTINO; LEE, 2023). Essa convergência confere legitimidade à incorporação dessas técnicas ao instrumental da saúde coletiva, desde que observadas as devidas cautelas.

A segunda convergência é a advertência ética. A literatura sobre inteligência artificial em saúde é consistente em apontar o viés, a opacidade e os riscos à privacidade como desafios centrais, que exigem regulação, transparência e salvaguardas (LEONEL *et al.*, 2024). Essa preocupação não é periférica ao uso das técnicas, mas constitutiva de sua adoção responsável, e atravessa toda a literatura examinada.

A terceira convergência, de ordem propositiva, é o reconhecimento de que a solução para os riscos não está na renúncia à tecnologia, mas na sua governança. A literatura sobre proteção de dados e ética da inteligência artificial converge em apontar a transparência, a auditabilidade e a proteção dos dados sensíveis como pilares de um uso legítimo, aplicáveis também ao campo da violência (BRASIL, 2018). Essa convergência desloca o debate do falso dilema entre adotar ou rejeitar a técnica para a questão mais produtiva de como adotá-la de modo responsável, o que orienta a construção de arranjos institucionais capazes de conciliar o aproveitamento analítico com a proteção das pessoas.

4.2 TENSÕES

A principal tensão identificada opõe o potencial analítico aos limites éticos, e não se resolve pela supressão de um dos polos. Renunciar à técnica significaria abrir mão de um recurso capaz de

aprimorar a proteção das vítimas; adotá-la sem salvaguardas significaria expô-las a novos riscos. A saída não está em escolher entre potencial e ética, mas em construir arranjos que permitam realizar o primeiro sem violar a segunda — o que remete à governança de dados, à transparência metodológica e à participação de múltiplos atores na definição dos usos aceitáveis.

Uma segunda tensão diz respeito à qualidade dos dados. O potencial da análise é diretamente proporcional à qualidade e à representatividade dos registros, mas estes são marcados pela subnotificação e por vieses de origem. Existe, assim, o risco de que a sofisticação metodológica crie uma aparência de objetividade que encubra as limitações dos dados subjacentes, conferindo autoridade indevida a resultados que refletem, em parte, as lacunas do próprio registro.

Uma terceira tensão, menos evidente, opõe a padronização exigida pelos métodos computacionais à singularidade de cada situação de violência. Para serem processados por algoritmos, os casos precisam ser reduzidos a variáveis padronizadas, o que inevitavelmente descarta parte da complexidade e do contexto de cada história. Esse processo de simplificação é condição da análise, mas também seu limite: o que se ganha em capacidade de reconhecer padrões agregados perde-se em sensibilidade à particularidade de cada caso. Para a saúde coletiva, que precisa articular a visão populacional à atenção individual, essa tensão exige que a análise de padrões nunca substitua a escuta e o cuidado singulares, mas os complemente.

Essas tensões, tomadas em conjunto, desenham um campo de decisões que não comporta soluções puramente técnicas. Definir quando o ganho analítico justifica o risco à privacidade, quando a padronização é aceitável e quando ela apaga o essencial, quando um padrão descoberto merece tradução em política — todas essas são decisões que exigem deliberação, e que envolvem valores sobre os quais a técnica, por si, nada diz. Reconhecer isso é reconhecer que a governança da aprendizagem de máquina em saúde coletiva é, antes de tudo, uma questão de deliberação coletiva e de responsabilidade pública, e não um problema de otimização algorítmica.

4.3 LACUNAS

A primeira lacuna é a escassez de aplicações não supervisionadas especificamente voltadas às bases brasileiras de notificação e mortalidade, o que limita a validação do potencial aqui discutido no contexto nacional. A segunda é a ausência de protocolos consolidados de governança e de anonimização adaptados ao tratamento de dados de violência, que compatibilizem o aproveitamento analítico com as exigências da proteção de dados. A terceira é a insuficiente incorporação da perspectiva das próprias mulheres e dos serviços na definição dos usos legítimos dessas técnicas, o que

aponta para a necessidade de abordagens participativas que evitem a imposição, de cima para baixo, de soluções tecnológicas.

A quarta lacuna diz respeito à formação. A aplicação responsável dessas técnicas exige profissionais capazes de transitar entre a competência computacional, o conhecimento epidemiológico e a sensibilidade ética, perfil ainda raro e pouco contemplado pelos currículos tradicionais. A ausência dessa formação interdisciplinar tende a produzir dois desvios simétricos: análises tecnicamente sofisticadas, porém cegas às particularidades do fenômeno e aos riscos éticos; ou abordagens sensíveis ao problema social, mas incapazes de aproveitar o instrumental analítico disponível. Superar essa lacuna exige investimento em formação que integre esses saberes, condição para que a incorporação da tecnologia se dê com competência e responsabilidade.

A quinta lacuna, de natureza regulatória, refere-se à ausência de diretrizes específicas para o uso de inteligência artificial no campo da violência e da saúde. Embora a legislação de proteção de dados forneça um arcabouço geral, faltam orientações que traduzam seus princípios em procedimentos concretos para o tratamento de dados de violência — desde os critérios de anonimização até os limites dos usos admissíveis. A construção dessas diretrizes, com a participação dos diversos atores envolvidos, constituiria salvaguarda importante contra usos indevidos e contribuiria para a segurança jurídica das iniciativas legítimas.

A sexta lacuna, de caráter avaliativo, é a ausência de estudos que meçam o efeito real da incorporação dessas técnicas sobre os desfechos da violência — a redução de casos, a melhoria da proteção, a satisfação das vítimas com os serviços. A literatura concentra-se, majoritariamente, na demonstração da viabilidade técnica e na descrição dos padrões descobertos, mas raramente avança até a avaliação de impacto. Sem essa etapa, permanece em aberto a pergunta decisiva: se e em que medida a sofisticação analítica se traduz em benefício concreto para as mulheres. Preencher essa lacuna exige desenhos de avaliação que acompanhem as iniciativas ao longo do tempo e que comparem seus resultados aos de abordagens convencionais.

4.4 IMPLICAÇÕES

As implicações desta análise são simultaneamente metodológicas e éticas. Metodologicamente, sugere-se que a aprendizagem não supervisionada seja compreendida como etapa exploratória, geradora de hipóteses a serem validadas, e não como fonte de verdades acabadas. Eticamente, impõe-se que sua adoção seja precedida e acompanhada de mecanismos de governança de dados, de transparência e de mitigação de vieses, sob pena de que a tecnologia reproduza as desigualdades que

se propõe a enfrentar. A articulação entre a competência computacional, o conhecimento epidemiológico e a reflexão ética é, nesse sentido, condição de um uso legítimo.

Cumpre, por fim, resguardar-se contra os extremos do entusiasmo e da recusa. Nem a técnica é panaceia capaz de resolver por si a violência contra a mulher, nem é ameaça a ser rejeitada em bloco. O que a evidência autoriza afirmar é que a aprendizagem não supervisionada oferece um recurso analítico valioso para a saúde coletiva, cujo aproveitamento responsável depende de uma governança que coloque a proteção das vítimas — e não a mera exploração de seus dados — no centro de suas finalidades.

Convém, ainda, situar essas implicações no marco dos princípios que orientam a saúde coletiva. O compromisso com a equidade, que define o campo, impõe que qualquer instrumento analítico seja avaliado não apenas por sua eficiência, mas por seus efeitos distributivos: se amplia ou reduz as desigualdades no acesso à proteção. Sob esse critério, a aprendizagem de máquina só se justifica quando contribui para alcançar quem está mais desprotegido, e não quando, ao apoiar-se em dados enviesados, reforça a invisibilidade dos grupos já negligenciados. A avaliação de impacto equitativo deveria, assim, integrar o próprio desenho de qualquer aplicação, como critério de sua legitimidade.

Do mesmo modo, o princípio da participação social, caro à saúde coletiva brasileira, sugere que a definição dos usos legítimos dessas técnicas não seja monopólio de especialistas, mas envolva os movimentos de mulheres, os serviços e as instâncias de controle social. A incorporação dessas vozes não é obstáculo à inovação, mas condição de sua legitimidade democrática, e pode contribuir para identificar riscos e usos indevidos que escapariam a uma avaliação puramente técnica. A tecnologia, nesse marco, é meio a serviço de finalidades socialmente deliberadas, e não um fim que se imponha por sua própria disponibilidade.

5 CONCLUSÃO

A análise realizada permite responder ao objetivo central: a aprendizagem de máquina não supervisionada oferece potencial analítico relevante para a compreensão dos padrões de violência contra a mulher no Brasil, ao permitir a descoberta de perfis de vulnerabilidade, o reconhecimento de padrões de vitimização e a priorização de grupos de risco a partir das extensas bases de dados hoje subaproveitadas. Esse potencial, contudo, é indissociável de limites éticos substantivos — o viés, a opacidade e os riscos à privacidade — que condicionam sua adoção responsável.

Retomando os objetivos específicos, verificou-se que a magnitude da violência e a riqueza dos dados de notificação justificam o interesse por técnicas analíticas mais sofisticadas; que a aprendizagem não supervisionada, ao dispensar categorias prévias, é especialmente adequada à

descoberta de padrões latentes; que suas aplicações à violência de gênero, ainda que concentradas em outros contextos, demonstram viabilidade e utilidade; e que seus limites éticos, longe de serem acessórios, são constitutivos de qualquer uso legítimo.

As implicações para a saúde coletiva são claras. A incorporação dessas técnicas deve ser compreendida como oportunidade condicionada: oportunidade de ampliar a capacidade analítica do sistema, condicionada à observância de salvaguardas éticas e jurídicas que protejam as populações vulneráveis. A ausência dessas salvaguardas converteria uma promessa de proteção em fonte de novos riscos, o que reforça a centralidade da governança de dados e da transparência.

Reconhecem-se as limitações inerentes a uma revisão integrativa, cujo alcance depende da literatura disponível e cuja transposição ao contexto brasileiro permanece, em parte, no plano das possibilidades. Delineiam-se, assim, perspectivas para pesquisas futuras: a aplicação de técnicas não supervisionadas às bases nacionais de notificação e mortalidade, com validação epidemiológica; o desenvolvimento de protocolos de governança e de anonimização específicos para dados de violência; e a construção de abordagens participativas que incorporem a perspectiva das vítimas e dos serviços. Em síntese, a aprendizagem de máquina não supervisionada apresenta-se como recurso promissor para a saúde coletiva, cujo valor dependerá menos de sua sofisticação técnica do que da responsabilidade ética com que for empregada.

REFERÊNCIAS

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Diário Oficial da União, 2018.

HUI, V.; CONSTANTINO, R. E.; LEE, Y. J. **Harnessing machine learning in tackling domestic violence**: an integrative review. International Journal of Environmental Research and Public Health, v. 20, n. 6, e4984, 2023. DOI: 10.3390/ijerph20064984.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA; FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA. **Atlas da violência 2024**. Brasília: Ipea; São Paulo: FBSP, 2024.

LEONEL, J. S. *et al.* Inteligência artificial: desafios éticos e futuros. **Revista Bioética**, v. 32, e3739, 2024. DOI: 10.1590/1983-803420243739PT.

PINTO-MUÑOZ, C. C.; ZUÑIGA-SAMBONI, J. A.; ORDOÑEZ-ERAZO, H. A. **Machine learning applied to gender violence**: a systematic mapping study. Revista Facultad de Ingeniería, v. 32, n. 64, e15944, 2023. DOI: 10.19053/01211129.v32.n64.2023.15944.

SOUZA, M. T.; SILVA, M. D.; CARVALHO, R. Revisão integrativa: o que é e como fazer. **Einstein (São Paulo)**, v. 8, n. 1, p. 102-106, 2010. DOI: 10.1590/S1679-45082010RW1134.

STOCHERO, L.; PINTO, L. W. Characterization of notifications of violence against women living in rural contexts in Brazil from 2011 to 2020. **Revista Brasileira de Epidemiologia**, v. 27, e240059, 2024. DOI: 10.1590/1980-549720240059.