

## Uso de mídias sociais para detecção de rumores em saúde pública

*Using social media to detect public health rumors*

Leandro da Silva Gonçalves<sup>1</sup>, Danilo Pestana de Freitas<sup>1</sup>, Alessandro Jatobá<sup>2</sup>, Paulo Victor Rodrigues de Carvalho<sup>1</sup>

DOI: 10.1590/2358-28982025E19919P

**RESUMO** O cenário epidemiológico contemporâneo, com sua rápida disseminação de informações e desinformação, requer melhorias nos sistemas de vigilância em saúde, especialmente na identificação de rumores. Plataformas como as mídias sociais oferecem um grande volume de dados que, mesmo sendo de complexo acesso, podem ser utilizados para o monitoramento de rumores em saúde. Este estudo explora a aplicação de recursos tecnológicos e propõe um framework conceitual para otimizar o uso de informações de mídias sociais pelas equipes de vigilância em saúde, facilitando a análise e detecção de rumores. A validação do framework, por meio de instâncias práticas com dados do Twitter, identificou um conjunto significativo de mensagens relevantes para detecção de rumores de síndrome diarreica e respiratória, otimizando o tempo de análise em comparação com buscas manuais. Conclui-se que o framework proposto oferece uma abordagem estruturada e promissora para otimizar o uso de informações de mídias sociais na vigilância em saúde, auxiliando na ágil identificação de rumores. Esta pesquisa contribui para o desenvolvimento de um mecanismo para classificação dessas mensagens, a formulação de planos de busca e análises adaptáveis, e a criação de repositórios para gestão e reutilização de dados e parâmetros, elevando a eficiência na vigilância em saúde pública.

**PALAVRAS-CHAVE** Mídias sociais. Vigilância em saúde pública. Emergências.

**ABSTRACT** *The contemporary epidemiological scenario, with its rapid dissemination of information and misinformation, necessitates enhancements in health surveillance systems, particularly in detecting rumors. Platforms such as social media offer a large volume of data that, although challenging to access, can be utilized to monitor health rumors. This study explores the application of technological resources and proposes a conceptual framework to optimize the utilization of social media information by health surveillance teams, facilitating the analysis and detection of rumors. Validation of the framework, using practical examples with Twitter data, identified a significant set of messages relevant to the detection of rumors related to diarrheal and respiratory syndromes, thereby optimizing analysis time compared to manual searches. We conclude that the proposed framework offers a structured and promising approach to optimizing the use of social media information in health surveillance, aiding in the rapid identification of rumors. This research contributes to the development of a mechanism for classifying these messages, the formulation of adaptable search and analysis plans, and the creation of repositories for managing and reusing data and parameters, thereby increasing the efficiency of public health surveillance.*

**KEYWORDS** *Social media. Public health surveillance. Emergencies.*

<sup>1</sup>Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro (RJ), Brasil.  
lsg.nce@gmail.com

<sup>2</sup>Fundação Oswaldo Cruz (Fiocruz), Centro de Estudos Estratégicos da Fiocruz Antonio Ivo de Carvalho (CEE) - Rio de Janeiro (RJ), Brasil.



## Introdução

A complexidade crescente das emergências em saúde pública tem mostrado grandes desafios à capacidade de planejamento e respostas das agências governamentais pelo mundo<sup>1,2</sup>. A preocupação reside em mitigar possíveis mortes em face das instabilidades sanitárias, socioeconômica e política, com a intenção de preservar a saúde coletiva<sup>3</sup>. A disponibilidade das informações úteis, em tempo hábil<sup>4</sup>, mostra-se fundamental para o aumento da capacidade dos sistemas de vigilância em saúde em se antecipar e apoiar ações preventivas ou mesmo de contenção desde o surgimento de possíveis eventos em saúde<sup>5-7</sup>.

O grande volume de dados gerados pelas mídias sociais representa um campo fértil para vigilância em saúde pública ao oferecer potencial para o monitoramento de tendências, a identificação de surtos e a disseminação de informações de saúde ao público. Contudo, o sucesso na exploração desses dados requer a superação de alguns desafios comuns ao ambiente on-line, como o grande volume de informações, a diversidade de linguagens existentes e a variabilidade na credibilidade das fontes, além da ausência de estruturação das mensagens. Estudos científicos têm consistentemente apontado para a grande necessidade de soluções tecnológicas inovadoras capazes de processar e tratar essa vasta quantidade de informações, identificando dados oportunos e relevantes nas mídias sociais<sup>8-15</sup>.

No contexto da vigilância em saúde, observa-se que os órgãos de vigilância em saúde são responsáveis pela pronta resposta a eventos de saúde pública e pelo monitoramento de diversas fontes de informação, com dados oficiais e não oficiais (rumores)<sup>16-18</sup>. O modelo tradicional de vigilância em saúde, também conhecido como Vigilância Baseada em Indicadores (VBI)<sup>19</sup>, é um modelo voltado para uso de fontes oficiais de informações, que frequentemente

necessitam de mais agilidade e complemento de informações. Em contrapartida, a grande disponibilidade de rumores foi um dos motivos que apoiaram o desenvolvimento de abordagens como a Vigilância Baseada em Eventos (VBE)<sup>20</sup> e, recentemente, a Inteligência Epidemiológica (IE)<sup>21</sup>. Essas inovações têm o propósito de melhorar a sensibilidade da detecção prévia de possíveis ocorrências de eventos em saúde e acelerar a emissão de prováveis alertas.

Com isso, as mídias sociais, em especial o Twitter, vêm se destacando diante das demais ferramentas para busca ou compartilhamento de grande variedade de informações, incluindo sobre saúde<sup>22</sup>. Diversos fatores contribuem para esse destaque: boa parte dos posts são públicos, mensagens mais curtas que facilitam o processamento, a comunicação em tempo real permitindo acompanhar possíveis eventos e seus comportamentos e a cultura de compartilhamento de informações entre os usuários. É comum encontrar mensagens nas quais os usuários mencionam seu estado de saúde<sup>23</sup>. Esses aspectos tornam o Twitter uma rica fonte de informações para identificação de mensagens relevantes para a vigilância em saúde pública.

Essas informações representam uma grande contribuição para a diminuição do esforço de resposta a uma emergência de saúde pública, pois representam as informações de pessoas diretamente afetadas, capazes de aumentar a consciência situacional das equipes envolvidas<sup>24</sup>.

É importante ressaltar, contudo, que a plataforma antes conhecida como Twitter passou por mudanças no ano de 2023, adotando o nome de 'X'. Entre as mudanças, estão alterações nas políticas de acesso a sua Interface de Programação de Aplicativos (API), o que tem gerado discussões e desafios para a comunidade acadêmica quanto à coleta de dados de pesquisa. Embora este estudo tenha utilizado dados coletados antes dessas mudanças, é importante que

pesquisas futuras considerem o impacto dessas mudanças e explorem metodologias adaptadas para obtenção e análise de informações dessa mídia social e de outras<sup>25,26</sup>.

A capacidade de resposta e a resiliência da vigilância em saúde são cruciais e necessitam dessas informações oportunas. Além disso, as ações de gerenciamento e resposta a emergências em saúde pública são bastante complexas e urgentes, e qualquer apoio é importante. Assim, a coordenação e o controle adequado permitirão melhor articulação das ações e melhor aproveitamento dos recursos para o alcance de melhores resultados<sup>27</sup>.

Por meio do monitoramento constante de informações de diversas fontes e da realização de investigações para identificar possíveis ameaças à saúde pública, as equipes responsáveis preservam a integridade dos sistemas de vigilância em saúde e contribuem para a evolução do aprendizado, auxiliando na melhora de ações futuras ao lidar com novas perturbações<sup>28,29</sup>. Consideradas como uma grande fonte de informações úteis e oportunas<sup>30</sup>, o uso das informações de mídias sociais tem se mostrado promissor no processo de vigilância e na realização de respostas bem-sucedidas<sup>30,31</sup>.

No entanto, pela falta de uma estrutura que organize o grande volume de informações disponíveis, é necessário um imenso esforço para identificar aquelas que são úteis e disponibilizá-las em tempo hábil<sup>12,32,33</sup>. É possível encontrar iniciativas isoladas de algumas instituições e profissionais em usar essas informações para apoiar a tomada de decisões em cenários de eventos em saúde pública, mas a solução passa por um sistema padronizado de tratamento e análise destas informações.

Em complemento ao levantamento bibliográfico, as discussões aprofundadas com profissionais de vigilância em saúde com experiência no cotidiano operacional confirmaram a existência dos desafios na identificação e no processamento de dados úteis nas mídias sociais. Assim, revelou-se

a necessidade de uma abordagem mais estruturada para o tratamento dessas informações, visando aprimorar a detecção de rumores sobre síndromes. Nesse contexto, esta pesquisa propõe um *framework* conceitual que aborda desde a coleta de mensagens até o tratamento para a identificação de conteúdo de maior relevância. A validação do *framework* foi realizada mediante a construção de duas instâncias de aplicação, enfocadas no monitoramento e na detecção de rumores em saúde pública.

Nesse cenário, esta pesquisa tem como objetivo apresentar o desenvolvimento e a validação de um *framework* conceitual, cuja estrutura e proposta conceitual inicial foram delineadas em um trabalho anterior<sup>34</sup>, capaz de estruturar as informações de mídias sociais para otimizar o processo de detecção de rumores em saúde pública. Para isso, busca-se responder à seguinte questão de pesquisa: ‘como a estruturação da informação das mídias sociais pode apoiar a detecção de rumores em saúde pública?’.

## Material e métodos

Este estudo empregou uma metodologia que combinou uma investigação com abordagens qualitativas.

Foi realizado um levantamento da literatura científica que permitiu uma melhor compreensão das práticas correntes em vigilância em saúde, abrangendo aspectos regulatórios, estruturais e as estratégias empregadas. Esse levantamento também foi fundamental para identificar as principais lacunas de conhecimento e os desafios que se apresentam às equipes de saúde pública na estruturação de informações relevantes das mídias sociais.

Posteriormente, para aprofundar a compreensão sobre os fluxos de trabalho existentes na detecção de rumores e as demandas das equipes, foram conduzidas entrevistas não estruturadas com profissionais

e especialistas na área<sup>35</sup>. A partir do conhecimento adquirido, observou-se a necessidade de uma solução que aplique uma melhor estruturação para o uso das informações de mídias sociais.

Para condução das entrevistas não estruturadas, visando ao levantamento de requisitos e posterior avaliação do *framework*, foram selecionados três profissionais com base em sua experiência na área de detecção de rumores em saúde pública, ou seja, um, três e nove anos de experiência. As entrevistas foram compostas principalmente por questões abertas, no intuito de deixar os entrevistados falarem livremente<sup>36,37</sup>. Os temas explorados incluíram os desafios correntes da detecção de rumores, a percepção sobre o potencial das informações de mídias sociais e as funcionalidades desejáveis em ferramentas de apoio. Observou-se que a sobrecarga de informações e a ausência de ferramentas para auxiliar o gerenciamento de grandes volumes de informações são itens recorrentes entre os entrevistados. Todos os princípios éticos aplicáveis à pesquisa foram respeitados, os participantes foram devidamente informados sobre os objetivos do estudo e a confidencialidade dos seus dados, tendo consentido livremente em participar.

Com base no levantamento realizado e nas necessidades identificadas, foi proposto um *framework* conceitual capaz de viabilizar o desenvolvimento de sistemas nesse domínio complexo por meio da reunião das funcionalidades e elementos estudados, suas interações e tudo aquilo que contribui para o contexto abordado<sup>38</sup>. Após a construção do *framework*, esta pesquisa apresentou a solução elaborada ao mesmo grupo de especialistas entrevistados anteriormente. Assim, foram apresentadas as fases do *framework*, suas funcionalidades, seu escopo, utilidade e objetivos pretendidos. Ao final da apresentação, foram coletadas as percepções e sugestões de melhorias dos participantes.

Em seguida, para mostrar o

comportamento do *framework* proposto em um cenário real, esta pesquisa elaborou duas instâncias de busca por rumores baseadas em dados reais do Twitter. Os resultados obtidos foram apresentados e discutidos com os especialistas para validação tanto dos resultados quanto da contribuição para responder à questão abordada nesta pesquisa. As seções seguintes mostrarão a composição do *framework*, as instâncias aplicadas, os resultados obtidos e a avaliação deles.

## O *framework*

O *framework* está distribuído em fases distintas que preservam a flexibilidade durante o desenvolvimento dos sistemas, possibilitando o desenvolvimento evolutivo, acoplado e coeso das funcionalidades dos sistemas. Pela necessidade de um apoio mais complexo, a proposta não aborda tratamento de imagens, vídeos ou áudios.

Para o máximo aproveitamento das informações das mídias sociais, esta solução propõe o armazenamento das mensagens em repositórios de dados para preservação do histórico. Isso permite a reutilização das informações e a reaplicação de critérios e parâmetros de maneira integral ou parcial, para ampliar o nível de refinamento das análises e tratamentos já realizados ou utilizar essas informações como base para uma nova busca. É importante ressaltar que a reaplicação de critérios e parâmetros se refere à capacidade de analisar o conteúdo das mensagens coletadas e armazenadas, aplicando diferentes lógicas de análises sindrômicas, refinando filtros de descarte ou ajustando parâmetros de análise, sem implicar uma nova etapa de captura de mensagens.

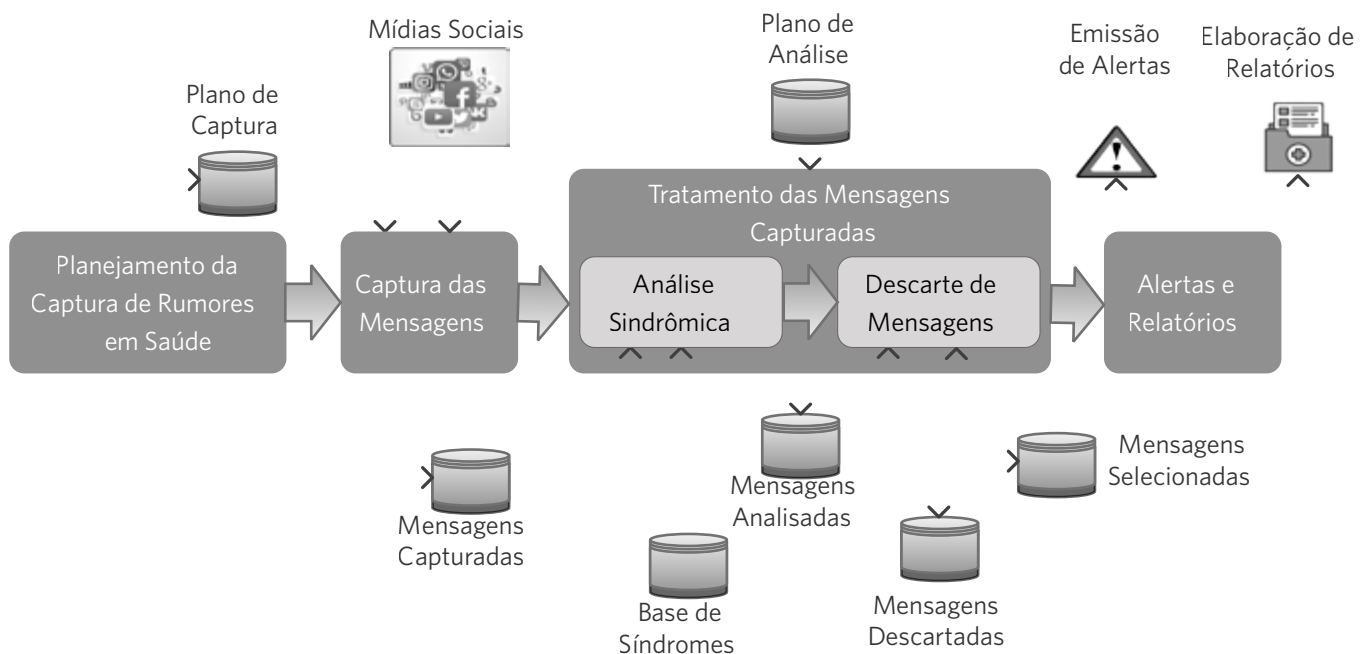
A automação das funções proposta auxilia na redução do esforço das equipes para capturar e tratar um volume suficiente de informações capazes de estabelecer a existência de um rumor. Esses planos podem ser aplicados de forma automática, tanto na captura quanto na análise das mensagens, pois serão

armazenados em repositórios de dados que permitem reuso. Assim, o *framework* orienta que soluções que apoiem as equipes de monitoramento e identificação de rumores possibilitem um refinamento da análise das mensagens por meio de modificações dos parâmetros aplicados sobre as mensagens com o objetivo de otimizar os resultados obtidos.

Para o *framework*, os sintomas ou sinais de uma síndrome se dividem em dois tipos: os primários e os secundários. Os primários são aqueles que se referem a características de identificação obrigatória para que seja apontada a existência de uma determinada síndrome. Já os secundários são características de identificação facultativa em que sua ausência não descaracteriza a presença de uma determinada síndrome, mas sua presença aumenta a precisão no apontamento da sua existência.

O processo do *framework* (figura 1) se inicia pelo Planejamento da Captura de Rumores em Saúde, no qual são estabelecidos os parâmetros que uma mensagem deve ter para ser capturada por meio das mídias sociais. Os parâmetros englobam conteúdo textual, ou seja, palavras-chave que representam sintomas, sinais, e respectivos sinônimos e variações regionais, considerados primários ou obrigatórios em uma mensagem relacionados com síndrome sob investigação seguidos pelos parâmetros relativos à geolocalização ou data de postagem da mensagem. Os parâmetros são armazenados em um repositório denominado Plano de Captura. Observe que, na maioria das buscas, os sinais e sintomas primários de cada síndrome devem ser tratados no Plano de Captura, pois são as palavras que necessariamente identificam rumores das síndromes pesquisadas.

Figura 1. Visão Geral



Fonte: elaboração própria.

Em seguida, há a captura das mensagens conforme conteúdo do plano de captura. O *framework* propõe que o plano pode ser aplicado tanto para buscar mensagens em tempo real quanto mensagens já postadas. Durante a captura, a solução aderente ao *framework* deve ser capaz de realizar um pré-processamento no conteúdo textual (retirada de *tokens* desnecessários, *emojis*, preposições, expressões etc.) antes do seu armazenamento no repositório denominado Mensagens Capturadas, para facilitar a comparação de seus conteúdos com os parâmetros de análise.

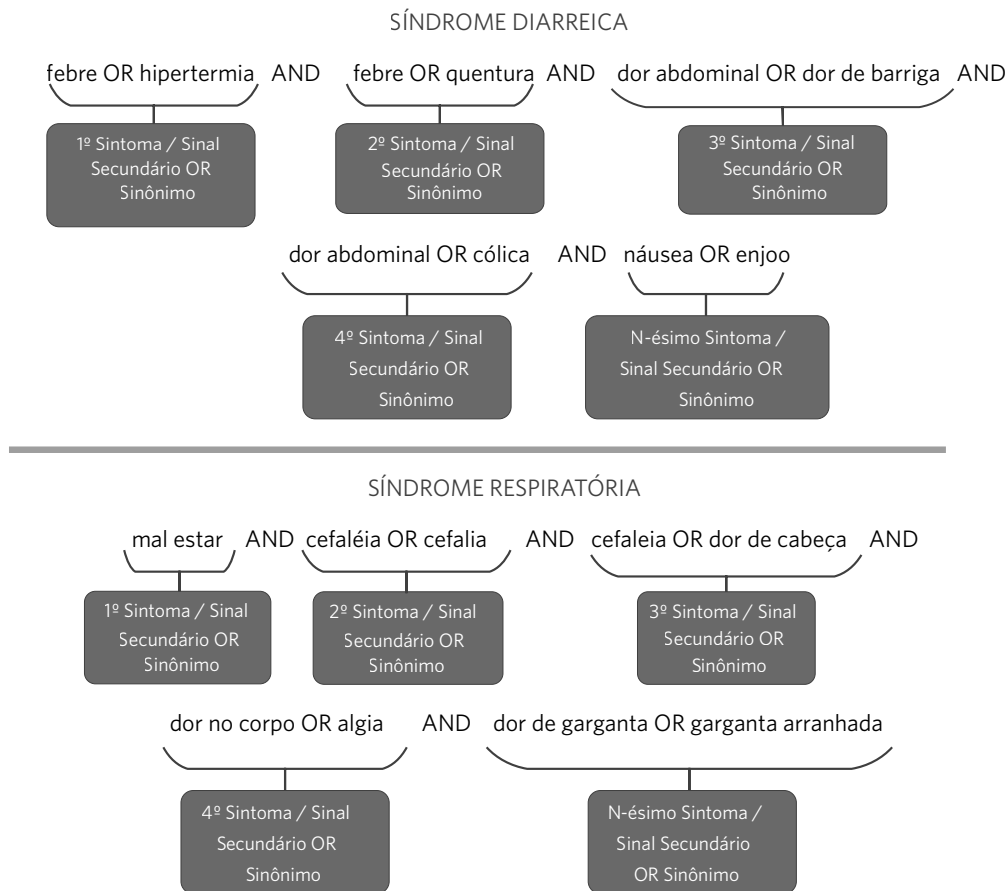
Após a captura, essas mensagens são analisadas para classificação conforme o seu grau de relacionamento com a síndrome investigada. Nessa fase, o *framework* sugere um plano de análise. Este é armazenado no repositório de dados denominado Plano de análise, e contém os parâmetros necessários para realização de duas análises: Análise Síndrômica e Descarte de Mensagens.

Por intermédio da funcionalidade de Análise Síndrômica, as mensagens são classificadas segundo sua relação com a síndrome que está sob alvo da investigação. Para isso, ele considera que a solução desenvolvida deve ter acesso a um repositório de dados preexistentes construído por especialistas na área da saúde, chamado de base de síndromes. Essa base contém a relação de cada uma das síndromes com seus sintomas e sinais considerados secundários ou facultativos, seus respectivos sinônimos e

possíveis variações regionais existentes que estão armazenados de maneira logicamente organizada. Cada organização lógica distinta dessas informações sobre uma síndrome forma uma lógica síndrômica que pode ser utilizada para identificar informações relevantes durante a análise. Por meio da utilização dos operadores AND, OR e NOT, é possível estabelecer uma ou mais lógicas síndrômicas sobre uma mesma síndrome, bastando apenas alterar a disposição das informações e dos operadores. Ao considerar a mudança na combinação do conteúdo das lógicas síndrômicas registradas, o *framework* contempla possíveis atualizações de parâmetros e diferentes níveis de restrição lógica sobre a análise.

Cada lógica síndrômica, da síndrome investigada, representa um parâmetro para a análise em andamento. A *figura 2* mostra as estruturas das lógicas síndrômicas utilizadas por esta pesquisa para identificar informações relevantes sobre síndrome diarreica e síndrome respiratória. Por intermédio dessa figura, é possível observar que, a cada sintoma ou sinal de uma síndrome, há uma construção lógica com seus respectivos sinônimos com uso dos operadores AND, OR, NOT, para contemplar possíveis variações regionais. Depois, cada uma dessas construções é novamente interligada com os mesmos operadores lógicos para compor uma das lógicas síndrômicas possíveis de determinada síndrome.

Figura 2. Exemplos de lógicas sindrômicas



Fonte: elaboração própria.

Ao observar a *figura 2*, é possível notar a existência de duas lógicas sindrômicas com estruturas um pouco diferentes. Além de estarem direcionadas a síndromes distintas, a lógica sindrômica referente à síndrome respiratória não pesquisou por sinônimos de ‘mal-estar’, deixando apenas este termo no primeiro sinal/sintoma dessa lógica sindrômica. Ao final da aplicação das lógicas sindrômicas sobre as mensagens capturadas, as mensagens resultantes dessa parte da análise são armazenadas em um repositório denominado Mensagens Analisadas.

Além dos parâmetros da análise sindrômica, o plano de análise contempla um conjunto de parâmetros preestabelecidos que será

aplicado sobre as mensagens já analisadas para identificar aquelas com pouca ou nenhuma contribuição para a investigação e retirá-las do processo. Isso auxilia na eliminação de qualquer ruído nas informações que serão usadas pelos especialistas. Os parâmetros englobam conteúdo textual, remetentes não confiáveis, geolocalização ou data de postagem da mensagem. O *framework* possibilita a realização de uma reanálise, quando houver a necessidade de aprofundamentos (ou refinamentos) nas análises já realizadas, com o estabelecimento de novos parâmetros. Armazenadas em um repositório de dados denominado

Mensagens Descartadas, essas mensagens não são excluídas de fato, e sim desconsideradas para uma determinada análise, mas elas são passíveis de serem analisadas em outro momento.

Ao final, o *framework* propõe uma estrutura para disseminação das informações resultantes mediante a elaboração de relatórios e emissão de alertas. No primeiro, com uso de gráficos e relatórios, a solução disseminará o resultado que pode servir de insumo para outros modelos já existentes. Quando todo o processo confirma a existência de um rumor sobre uma síndrome em determinada região, o *framework* sugere também a emissão de alertas que são documentos que identificam a concretização do rumor investigado e a necessidade de uma averiguação mais apurada no local pelas equipes de campo. É importante ressaltar que todo o histórico dos planos e mensagens é preservado para auxiliar nas capturas e análises futuras.

## Resultados

Nesta seção, são apresentados os principais achados da pesquisa, incluindo os resultados da aplicação prática do *framework* em duas instâncias de detecção de rumores e os resultados da avaliação qualitativa por especialistas em vigilância em saúde.

### Instâncias do *framework*

Com o *framework* ajustado conforme o *feedback* dos especialistas, foram construídas duas instâncias com utilização de informações reais obtidas por meio do aplicativo Twitter – isso para demonstrar e validar sua aplicabilidade<sup>39</sup>. Cada uma delas foi aplicada com escopo e localidades distintas. As duas instâncias foram realizadas no mesmo período de sete dias

anteriores à data de captura (1º a 7 de agosto de 2021), período considerado razoável para observação do comportamento epidemiológico. As instâncias referenciaram uma mesma área geográfica, a região compreendida na área programática 2.1 do Rio de Janeiro<sup>40</sup>. Os resultados específicos de cada instância serão detalhados a seguir.

A primeira instância concentrou-se na síndrome diarreica. A aplicação do plano de captura foi composta pelas informações de posicionamento geográfico e período de captura já mostrados. Essa instância utilizou 49 palavras-chave (*figura 3*) relacionadas com sintomas, sinais, sinônimos e variações regionais primárias dessa síndrome, resultando em 1.941 *tweets* capturados. Após a análise sindrômica, aplicou 59 palavras-chave referentes a sintomas, sinais, sinônimos e variações regionais consideradas relevantes pelos especialistas (lógica sindrômica). Assim, 133 mensagens (6,85% do total capturado) foram classificadas como analisadas, e as 1.808 (93,15% do total capturado) mensagens não classificadas pelos parâmetros estabelecidos de análise ficaram disponíveis para possíveis reclassificações após refinamento dos parâmetros dos planos.

Em seguida, o processo de descarte utilizou 57 palavras-chave capazes de identificar mensagens com pouca ou nenhuma contribuição para a investigação em andamento. Por exemplo, palavras como ‘jogo’ ou ‘genocida’ podem identificar mensagens que estejam relacionadas com um jogo de algum esporte ou com uma crítica política. Também há palavras que podem caracterizar disseminadores de ‘fake news’ ou mesmo um ‘bot’ gerador de mensagens irrelevantes para a investigação. Isso reduziu o conjunto de mensagens relevantes para apenas 45 (2,32% do total capturado), que foram consideradas pertinentes e encaminhadas para avaliação especializada e a possível emissão de alertas (*figura 3*).

Figura 3. Parâmetros da Instância de Síndrome Diarreica

| PLANEJAMENTO DA CAPTURA (49 PALAVRAS-CHAVE) |  |
|---|--|
| SINTOMAS E SINAIS PRIMÁRIOS                 | SINÔNIMOS  |
| diarreia OR                                 | afitamento OR afito OR borra OR cagado OR cagando OR caganeira OR cagar OR câmaras OR cambras OR carreirinha OR caseira OR corredeira OR corrução OR defecação OR defecado OR defecando OR defecar OR dejeção OR dejetar OR desaranjado OR desaranjo OR destempero OR disenteria OR evacua OR evacuacao OR evacuado OR evacuando OR evacuar OR excreção OR excretado OR excretar OR excreto OR fezes amolecida OR fezes aquosa OR fezes liquida OR fezes mole OR fezes pastosa OR incontidência OR ligeira OR maculo OR mal de bicho OR obrado OR obrar OR piriri OR reira OR sedeca OR soltura OR soltura do ventre |

| PLANEJAMENTO DA ANÁLISE (59 PALAVRAS-CHAVE) |  |
|---|--|
| SINTOMAS E SINAIS SECUNDÁRIOS               | SINÔNIMOS  |
| febre OR                                    | Febril OR Febrento OR Febrão OR Febrinha OR Febrilizar OR Febrícula OR Febriólogo OR Febroso OR Hipetermia OR Hiperpiese OR hipertermico OR Pirexia OR Quentura OR Temperatura corporal Elevada OR Temperatura OR corporal alta OR |
| dor abdominal OR                            | dor no abdome OR dor de barriga OR dor na barriga OR colica OR abdominalgia OR obstrução OR espasmos OR  |
| sangue nas fezes OR                         | fezes escuras OR fezes pretas OR cheiro forte OR hematoquezia OR melena OR sangramento OR sangrento OR   |
| muco nas fezes OR                           | cor das fezes OR fezes pretas OR fezes verdes OR coco OR merda OR bosta OR   |
| vomito OR                                   | vomitado OR vomitando OR vomitar OR vomição OR golfada OR golfo OR bolçado OR jorro OR lanço OR êmene OR desembucho OR regurgito OR nojo OR  |
| nausea OR                                   | enjoo OR enjojo OR ânsia OR asco   |

| PARÂMETROS DE DESCARTE (57 PALAVRAS-CHAVE)   |
|--|
| as merdas OR cagando na cabeça OR estar cagado OR Lula OR time OR bolsonaro OR cagando para OR essa bosta que OR merda pra cagar OR merda que OR Borra bosta OR cagando pra OR falar merda OR mesma bosta OR militante OR pra cagar OR cagar pra OR faz merda OR mesma merda OR mito OR bosta que OR cagando se OR ficar cagado OR comida caseira OR olimpiada OR caga regra OR cagando solenemente OR ficar na merda OR defecar pela OR os merdas OR cagando regra OR cagar no pau OR fosse merda OR é uma merda OR passar coco OR cagado de algo OR comidinha caseira OR futebol OR se fosse merda OR politico OR cagando de OR eleitor OR genocida OR stf OR presidente OR cagando de medo OR essa merda OR governo OR stm OR cagando de rir OR esta cagando OR ligeira OR ta cagado OR cagando e andando OR esta merda OR ligeira impressão OR vasco |

Fonte: elaboração própria.

A segunda instância foi voltada à síndrome respiratória. Utilizando os mesmos parâmetros de geolocalização e período de captura, esse plano de captura se diferencia pelo uso de 27 palavras-chave relacionadas com sintomas, sinais, sinônimos e variações

regionais primários dessa síndrome (figura 4), obtendo 826 *tweets* capturados. Em seguida, a análise utilizou outras 27 palavras-chave relacionadas com sintomas, sinais ou sinônimos secundários também consideradas relevantes por especialistas



Figura 5. Exemplos de mensagens selecionadas pelas instâncias

**MENSAGENS SELECIONADAS - SÍNDROME DIARREICA**

- amanha eu acordo cinco e meia isso sao horas de ter dor de barriga estou me acabando em cagar jesus cristo
- caralho acordei agora do nada com maior caganeira e fiquei ai agora q lembrei q comemos batatas fritas cheiiii di oleo linguica e eu ainda comi um ovo de pascoa inteiro sozinha
- @matheusalvescrf eu passo muito mal da caganeira e vomito ao mesmo tempo sai por cima e por baixo deus me livre nem me arrisco a comer algo com isso mais
- ja acordei com a mirella cagando a casa toda e o henrique vomitando
- @raqueldrehmer @feltrinoficial quer dizer got e um cocozao parrudaco himym e mais uma diarrei a rala
- @josiasteofi ofo @kimpaim nando nao manda nem nas fezes dele ta todo cagado
- cheia de dor de barriga mas n consigo cagar q agonia mane
- acordei vomitando com dor de cabeça tontura di arreia febre corona nao itaipava
- bati meu recorde de tempo sentada no vaso tentando aliviar uma colica de 30min fui pra 1 h 40min
- deu ate dor na barriga ja cheguei em casa cagando mermao
- @maktubdourado teve caganeira ou so vomito
- com um pouco de febre e di arreia
- plantao eva desabafo hoje ela ja vomitou e a diarrei a voltou ta mais triste na caminha dela essa noite acordou varias vezes pra fazer coco orando pra ela comer bem daqui a pouco porque o remedio dela chega amanha

**MENSAGENS SELECIONADAS - SÍNDROME RESPIRATÓRIA**

- qualquer coisinha a eu com aperto no peito falta de ar dor de cabeça
- chuva a dor de cabeça nariz entupido oi linda vamos querer
- @profeta calado @youtube amem toda enfermidade que esta no meu toda dor no corpo toda dor de cabeça dor nas juntas e nariz entupido todo mal estar todo cansaco em nome de jesus sejam curados agora em nao de jesus em nome de jesus
- @igoreuristenes fadiga e dor no corpo tbm
- @igoreuristenes fadiga febre dor no corpo dura por 2 3 dias
- dor de cabeça e falta de ar e nao para de cair ligacao pqp
- e nariz entupido e colica e dor de cabeça e eu espirrando toda hr
- chega essa hora nao consigo dormir nariz entupido dor de cabeça e falta de ar eu q lute
- to com uma angustia no peito uma falta de ar horrivel
- a pior parte de chorar muito e o nariz entupido e a dor de cabeça
- vou descansar to cheia de dor de cabeça e pra piorar com nariz entupido
- garganta ruim nariz entupido dor de cabeça
- eu como e parece que estou infartando depois um mal estar tenho que ficar uma hora sentada pelo menos que fadiga horrivel

Fonte: elaboração própria.

## Avaliação do *framework*

A percepção dos especialistas em vigilância em saúde sobre o *framework* foi coletada via entrevistas, discussões e a realização de outro questionário que também aplicou questões

abertas e seguiu os mesmos princípios éticos já mencionados para levantar críticas e sugestões para a solução apresentada.

Posteriormente, o *framework* foi apresentado aos especialistas para uma avaliação inicial. Mesmo com observações sobre a necessidade

de alguns ajustes, os especialistas demonstraram uma boa aceitação quanto ao conceito do *framework* desenvolvido. As funcionalidades foram avaliadas pelos mesmos especialistas por meio de apresentação de um questionário, ressaltando clareza, aplicabilidade e relevância.

Foram abordadas questões sobre a relevância dos parâmetros de captura, como palavras-chave, localização e data; a estratégia de classificar sintomas primários para captura e secundários para a análise sindrômica foi considerada um ótimo direcionador para identificar e analisar mensagens durante o processo de detecção de rumores. Segundo os especialistas, os parâmetros e a localização estabelecidos para as instâncias foram considerados de boa abrangência para identificar um maior volume de informações relevantes na etapa de captura, contemplando: o maior número de variações nas escritas dos sintomas; a importância dos critérios de descarte para refinar os resultados; a utilidade do armazenamento de informações para análises futuras, permitindo uma análise sobre um maior volume de informações com menos esforço pelas equipes de vigilância em saúde; e a capacidade do *framework* em identificar mensagens relevantes ao final do processo, capazes de realçar a necessidade de emissão de alertas sobre a existência de um rumor em saúde pública.

Cada avaliação ocorreu de maneira remota e individual. Dessa maneira, foi possível levantar as críticas e sugestões de cada participante sem a interferência dos demais. Em cada uma das avaliações, o participante pôde assistir à demonstração do *framework* proposto e seu funcionamento com liberdade de interrupção para questionamento.

Considerando que as instâncias foram aplicadas em uma área de abrangência correspondente a toda a área programática 2.1 no Rio de Janeiro, e não a apenas um bairro específico, os resultados foram satisfatórios.

## Discussões e trabalhos futuros

A motivação desta pesquisa surgiu por meio do questionamento sobre como estruturar as informações de mídias sociais para otimizar o esforço das equipes de vigilância em saúde pública em suas atividades de detecção de rumores sobre síndromes.

A existência de informações úteis nas mídias sociais foi apurada por diversas pesquisas que também apontaram a dificuldade em manipular tais informações. Logo, identificar um processo de agregação, refinamento e geração de relatórios estimulou esta pesquisa.

Esta investigação considera a necessidade da atuação de especialistas em vigilância epidemiológica e saúde pública. O conhecimento desses profissionais é fundamental para a correta definição dos sintomas primários e secundários, para a elaboração das lógicas sindrômicas e para a interpretação dos alertas gerados, assegurando a relevância e a aplicabilidade do *framework*.

Mesmo com diversos estudos e soluções desenvolvidas que buscam explorar as informações de mídias sociais na área da saúde, algumas não apresentaram uma estrutura na qual o conhecimento do especialista pudesse ser explorado, dando-lhes a possibilidade de analisar diferentes resultados por meio do refinamento de suas análises<sup>33,41,42</sup>. O *framework* aqui proposto busca superar essa limitação ao incorporar explicitamente a experiência dos especialistas em saúde nas etapas de configuração dos planos de captura e análise e ao permitir o refinamento iterativo desses planos.

Outras pesquisas não apresentaram um tratamento das informações de maneira mais estruturada para identificar a sua relevância<sup>12,43</sup>, porém, este trabalho propôs que o tratamento das informações mediante a aplicação de uma organização lógica seja composto de alguns dos sintomas e seus respectivos sinônimos.

Essa abordagem estruturada que distingue sintomas primários e secundários, e utiliza lógicas sindrômicas flexíveis, representa um avanço em relação a métodos que aplicam apenas buscas por palavras-chave de forma mais genérica.

Durante a pesquisa, foi interessante observar os diferentes resultados por meio da aplicação de diferentes lógicas sindrômicas de uma mesma síndrome sobre uma mesma massa de informações já capturadas e como uma lógica preexistente pode servir de base para construção de outras que irão identificar diferentes síndromes.

Dentre as principais contribuições desta pesquisa, destacam-se: um mecanismo de classificação de mensagens mais direcionado e eficaz para identificar informações relevantes; a manutenção de um histórico completo de mensagens e parâmetros de análise, promovendo reusabilidade; e a introdução da análise sindrômica baseada em um repositório de lógicas criadas por especialistas, que organiza sintomas e sinais de forma lógica para maior precisão na detecção de rumores. A flexibilidade para criar e reutilizar múltiplas lógicas sindrômicas a partir de um repositório dedicado é um diferencial que otimiza as investigações realizadas.

Esta pesquisa reconhece a limitação de ter sido voltada apenas para o tratamento textual das informações do Twitter devido à complexidade do processamento de mídias como vídeos e imagens. A abordagem priorizou o conhecimento dos profissionais e especialistas em saúde sem o uso de ontologias ou vocabulários controlados para a classificação, dada a deficiência de soluções tecnológicas que atendam às necessidades imediatas deles.

Para trabalhos futuros, esta pesquisa planeja expandir a validação e o enriquecimento conceitual do *framework* com a participação de: especialistas de outras áreas da saúde para refinar as lógicas sindrômicas e a relevância dos sintomas; profissionais de comunicação e análise de mídias sociais, para aprimorar a compreensão dos padrões de disseminação de

informação e desinformação, com objetivo de melhorar as estratégias de captura; especialistas em gestão do conhecimento e ciência de dados, para otimizar os repositórios e explorar técnicas analíticas para o tratamento de grande volume de informações; especialistas em Inteligência Artificial e Machine Learning, para aprimorar o tratamento das mensagens após sua captura, otimizando a relevância das mensagens resultantes das análises.

## Conclusões

Esta pesquisa elaborou e avaliou um *framework* conceitual inovador, com objetivo de melhorar o processo de detecção de rumores relacionados com a síndromes nas mídias sociais. A proposta visa apoiar as equipes de vigilância em saúde pública para lidar de maneira eficaz com o grande volume e considerável complexidade das informações vindas dessas mídias, para transformá-las em um recurso oportuno para a realização de uma resposta preventiva a possíveis eventos em saúde.

Após o levantamento da literatura e entrevistas abertas com especialistas de vigilância em saúde, foi percebido o desafio das equipes de saúde pública no processo de identificação de rumores. Assim, esta pesquisa apresentou um *framework* que estabelece parâmetros para captura e análise de mensagens, permitindo a realização de uma classificação assertiva com base em seu grau de contribuição e investigação. A arquitetura proposta inclui a definição de funções essenciais e a organização de repositórios de dados mais robustos, para garantir a persistência do histórico de informações e parâmetros das análises realizadas. Sob a forma de planos de captura e análise, a organização confere maior rastreabilidade e flexibilidade, possibilitando a reutilização e o refinamento das análises sobre os armazenados, até que o nível de informação desejado seja obtido.

Depois de sua elaboração, o *framework* foi avaliado pelos mesmos especialistas

participantes da pesquisa. Após a realização de uma apresentação detalhada da ferramenta e das instancias com informações reais, foram realizadas discussões aprofundadas e a aplicação de um questionário composto de questões abertas. Dessa forma, foi possível, coletar críticas e sugestões que confirmaram que a solução apresentada agradou aos especialistas que puderem observar o seu potencial uso com dados reais.

## Colaboradores

Gonçalves LS (0000-0003-3900-0748)\*, Freitas DP(0000-0001-8759-8719)\*, Jatobá A (0000-0002-7059-6546)\* e Carvalho PVR (0000-0002-9276-8193)\* contribuíram igualmente para a elaboração do manuscrito. ■

## Referências

1. Enaya MF, Klingbeil T, Krüger J, et al. A case study on the development of the German Corona-Warn-App. *J Syst Softw.* 2024;213:112020. DOI: <https://doi.org/10.1016/j.jss.2024.112020>
2. Greiner AL, Stehling-Ariza T, Bugli D, et al. Challenges in Public Health Rapid Response Team Management. *Heal Secur.* 2020;18(S1):S-8-S-13. DOI: <https://doi.org/10.1089/hs.2019.0060>
3. Mekone YM. Public health emergency declarations in the Ethiopian federal system: understanding the scope of state and federal emergency declarations and authorities. *Law, Democr Dev.* 2022;26(1):1-25. DOI: <https://doi.org/10.17159/2077-4907/2021/ldd.v26.1>
4. World Bank. Change cannot wait: Building resilient health systems in the shadow of COVID-19 [Internet]. Washington, DC: World Bank; 2022 [acesso em 2025 abr 3]. Disponível em: <https://hdl.handle.net/10986/38233>
5. Ferguson NM, Cummings DAT, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature.* 2005;437(7056):209-214. DOI: <https://doi.org/10.1038/nature04017>
6. Longini IM, Nizam A, Xu S, et al. Containing Pandemic Influenza at the Source. *Science.* 2005;309(5737):1083-7. DOI: <https://doi.org/10.1126/science.1115717>
7. Mota E, Teixeira MG. Vigilância Epidemiológica e a pandemia da Covid-19 no Brasil: elementos para entender a resposta brasileira e a explosão de casos e mortes. *Saúde debate.* 2020;44(Esp 4):130-45. DOI: <https://doi.org/10.1590/0103-11042020E408>
8. Freitas DP, Borges MRS, Carvalho PVR. A conceptual framework for developing solutions that organise social media information for emergency response teams. *Behav Inf Technol.* 2020;39(3):360-78. DOI: <https://doi.org/10.1080/0144929X.2019.1621933>
9. Lwin MO, Lu J, Sheldenkar A, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Heal Surveill.* 2020;6(2):e19447. DOI: <https://doi.org/10.2196/19447>
10. Martí P, Serrano-Estrada L, Nolasco-Cirugeda A. Social Media data: Challenges, opportunities and limitations in urban studies. *Comput Environ Urban Syst.* 2019;74:161-74. DOI: <https://doi.org/10.1016/j.compenvurbysys.2018.11.001>
11. Mondal T, Roy T, Bhattacharya I, et al. A Study on Rumor Propagation Trends and Features in a Post Disaster Situation. In: Proceedings of the 20th In-

\*Orcid (Open Researcher and Contributor ID).

- ternational Conference on Distributed Computing and Networking. New York, NY: ACM; 2019. p. 381-84. DOI: <https://doi.org/10.1145/3288599.3295581>
12. Simões MD, Neves ARM. Mineração de Texto No Twitter: Uma Ferramenta Auxiliar Na Detecção de Epidemias. In: 21º Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS); 2021; Porto Alegre. Porto Alegre: Sociedade Brasileira de Computação; 2021. p. 281-92. DOI: <https://doi.org/10.5753/sbcas.2021.16072>
  13. Stieglitz S, Mirbabaie M, Ross B, et al. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *Int J Inf Manage.* 2018;39:156-68. DOI: <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>
  14. Ramos MC, Gomes DF, Mello NF, et al. Big Data e Inteligência Artificial para pesquisa translacional na Covid-19: revisão rápida. *Saúde debate.* 2022;46(135):1202-14. DOI: <https://doi.org/10.1590/0103-1104202213518>
  15. Souza EP, Gomes CM, Barroso DH, et al. Aplicações do Deep Learning para diagnóstico de doenças e identificação de insetos vetores. *Saúde debate.* 2019;43(Esp 2):147-54. DOI: <https://doi.org/10.1590/0103-11042019S211>
  16. Agência Nacional de Vigilância Sanitária (BR). Regulamento Sanitário Internacional (2005): versão em português [Internet]. Brasília, DF; 2009 [acesso em 2025 abr 3]. Disponível em: [www.gov.br/anvisa/pt-br/assuntos/paf/regulamento-sanitario-internacional](http://www.gov.br/anvisa/pt-br/assuntos/paf/regulamento-sanitario-internacional)
  17. World Health Organization. Early detection, assessment and response to acute public health events: Implementation of early warning and response with a focus on event-based surveillance. Geneva: WHO; 2014.
  18. Santos SSBS, Melo CMM, Barboni AR, et al. Estudo de linha de base avaliativa sobre capacidade de gestão descentralizada da Vigilância Epidemiológica. *Saúde debate.* 2018;42(116):73-86. DOI: <https://doi.org/10.1590/0103-1104201811606>
  19. Maramraj KK, Naidu S, Ray S, et al. Navigating through COVID-19 Waves: Pandemic Response using Indicator-based Epidemiological Surveillance Tool (PRIEST). *J Mar Med Soc.* 2021;23(1):4-9. DOI: [https://doi.org/10.4103/jmms.jmms\\_89\\_21](https://doi.org/10.4103/jmms.jmms_89_21)
  20. Balajee SA, Salyer SJ, Greene-Cramer B, et al. The practice of event-based surveillance: concept and methods. *Glob Secur Heal Sci Policy.* 2021;6(1):1-9. DOI: <https://doi.org/10.1080/23779497.2020.1848444>
  21. Paquet C, Coulombier D, Kaiser R, et al. Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. *Eurosurveillance.* 2006;11(12):5-6. DOI: <https://doi.org/10.2807/esm.11.12.00665-en>
  22. Nanath K, Joy G. Leveraging Twitter data to analyze the virality of Covid-19 tweets: a text mining approach. *Behav Inf Technol.* 2023;42(2):196-214. DOI: <https://doi.org/10.1080/0144929X.2021.1941259>
  23. Shakeri Hossein Abad Z, Kline A, Sultana M, et al. Digital public health surveillance: a systematic scoping review. *NPJ Digit Med.* 2021;4(1):41. DOI: <https://doi.org/10.1038/s41746-021-00407-6>
  24. Palen L, Hiltz SR, Liu SB. Online forums supporting grassroots participation in emergency preparedness and response. *Commun ACM.* 2007;50(3):54-8. DOI: <https://doi.org/10.1145/1226736.1226766>
  25. Blakey E. The day data transparency died: How Twitter/X cut off access for social research. *Contexts.* 2024;23(2):30-3. DOI: <https://doi.org/10.1177/15365042241252125>
  26. Murthy D. Sociology of Twitter/X: Trends, challenges, and future research directions. *Annu Rev Sociol.* 2024;50(1):169-90. DOI: <https://doi.org/10.1146/annurev-soc-031021-035658>
  27. Ministério da Saúde (BR). Plano de Resposta às Emergências em Saúde Pública. Brasília, DF: Ministério da Saúde; 2014.
  28. Jatobá A, Carvalho PVR. Resiliência em saúde pública: preceitos, conceitos, desafios e perspectivas. *Saúde debate.* 2022;46(Esp 8):130-39. DOI: <https://doi.org/10.1590/0103-11042022e810>

29. Walker BH. Resilience: what it is and is not. *Ecol Soc*. 2020;25(2):art11. DOI: <https://doi.org/10.5751/ES-11647-250211>
30. Kim J, Bae J, Hastak M. Emergency information diffusion on online social media during storm Cindy in U.S. *Int J Inf Manage*. 2018;40:153-65. DOI: <https://doi.org/10.1016/j.ijinfomgt.2018.02.003>
31. Van de Walle B, Brugghehans B, Comes T. Improving situation awareness in crisis response teams: An experimental analysis of enriched information and centralized coordination. *Int J Hum Comput Stud*. 2016;95:66-79. DOI: <https://doi.org/10.1016/j.ijhcs.2016.05.001>
32. Hagen L, Neely S, Scharf R, et al. Social media use for crisis and emergency risk communications during the Zika health crisis. *Digit Gov Res Pract*. 2020;1(2):1-21. DOI: <https://doi.org/10.1145/3372021>
33. Anwar M, Khoury D, Aldridge AP, et al. Using Twitter to Surveil the Opioid Epidemic in North Carolina: An Exploratory Study. *JMIR Public Heal Surveill*. 2020;6(2):e17574. DOI: <https://doi.org/10.2196/17574>
34. Gonçalves LS, Freitas DP, Carvalho PVR. Uso de mídias sociais para detecção de rumores em saúde pública. *Esc Reg Sist Informação do Rio Janeiro*. 2021;1:1-4.
35. Minayo MCS, Deslandes SF, Gomes R. *Pesquisa Social: Teoria, Método e Criatividade*. 28ª ed. Petrópolis, RJ: Vozes; 2009.
36. Schallenburger M, Schwartz J, Icks A, et al. Triggers of intensive care patients with palliative care needs from nurses' perspective: a mixed methods study. *Crit Care*. 2024;28(1):181. DOI: <https://doi.org/10.1186/s13054-024-04969-1>
37. Saglam Y. Which data gathering method is superior: An open-ended questionnaire or a semi-structured interview? *Int J Stud Educ*. 2024;6(3):375-86. DOI: <https://doi.org/10.46328/ijonse.220>
38. Partelow S. What is a framework? Understanding their purpose, value, development and use. *J Environ Stud Sci*. 2023;13(3):510-19. DOI: <https://doi.org/10.1007/s13412-023-00833-w>
39. Gregor S, Hevner AR. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Q*. 2013;37(2):337-55. DOI: <https://doi.org/10.25300/MISQ/2013/37.2.01>
40. Freitas DA, Souza-Santos R, Wakimoto MD. Acesso aos serviços de saúde por pacientes com suspeita de dengue na cidade do Rio de Janeiro, Brasil. *Ciênc saúde coletiva*. 2019;24(4):1507-16. DOI: <https://doi.org/10.1590/1413-81232018244.11252017>
41. Wakamiya S, Kawai Y, Aramaki E. Twitter-Based influenza detection after flu peak via Tweets with indirect information: Text mining study. *JMIR Public Heal Surveill*. 2018;4(3):e65. DOI: <https://doi.org/10.2196/publichealth.8627>
42. Duval FV, Silva FAB. O uso do Twitter como minerador de eventos adversos de medicamentos de combate à malária: o caso da doxiciclina. *Cad Saúde Pública*. 2019;35(5):1-16. DOI: <https://doi.org/10.1590/0102-311x00033417>
43. Pessanha GRG, Fidelis TO, Freire CD, et al. #Fiqueemcasa: análise de sentimento dos usuários do Twitter em relação ao Covid19. *Holos*. 2020;5:1-20. DOI: <https://doi.org/10.15628/holos.2020.11147>

---

Recebido em 28/10/2024

Aprovado em 02/06/2025

Conflito de interesses: inexistente

Disponibilidade de dados: os dados de pesquisa estão contidos no próprio manuscrito

Suporte financeiro: não houve

**Editora responsável:** Paula de Castro Nunes