

Estado da publicação: Não informado pelo autor submissor

Modelos de vigilância em saúde apoiados em dados de redes sociais são uma realidade viável?

Kleber Rodrigues dos Santos, Daielly Mantovani, Celso Machado Jr, Guilherme A. Leal

<https://doi.org/10.1590/SciELOPreprints.11847>

Submetido em: 2025-05-03

Postado em: 2025-05-21 (versão 1)

(AAAA-MM-DD)

Modelos de vigilância em saúde apoiados em dados de redes sociais são uma realidade viável?

Kleber R. dos Santos <https://orcid.org/0000-0003-2123-5426>

Universidade de São Paulo, Brasil

Daielly M. N. Mantovani <https://orcid.org/0000-0001-6320-3268>

Universidade de São Paulo, Brasil

Celso Machado Jr. <https://orcid.org/0000-0003-3835-2979>

Universidade Municipal de São Caetano do Sul, Brasil

Guilherme Arevalo Leal <https://orcid.org/0000-0002-8415-0916>

Universidade de São Paulo - Brasil

RESUMO

Este artigo apresenta uma revisão sistemática da literatura com o objetivo de analisar como dados provenientes de redes sociais, em especial da plataforma X (antigo Twitter), têm sido empregados na formulação de modelos preditivos aplicados à vigilância sanitária, utilizando técnicas de aprendizado de máquina. Foram examinados 23 estudos publicados entre 2018 e 2021, período exatamente antes e após a pandemia de COVID-19. Esses estudos aplicaram algoritmos analíticos a postagens com georreferenciamento, com o propósito de detectar surtos, monitorar epidemias e apoiar processos decisórios em saúde pública. Os resultados sugerem que esses modelos podem antecipar focos de risco com até quatro semanas de antecedência, contribuindo para respostas mais oportunas por parte dos gestores. Contudo, observam-se limitações importantes para a adoção desses sistemas, tais como a escassez de dados geolocalizados, a vulnerabilidade a ruídos informacionais e os entraves técnicos e éticos relacionados ao processamento automatizado da linguagem natural. A revisão indica que, embora essas soluções revelem aplicabilidade crescente, sua consolidação institucional depende de estratégias integradas de validação, governança algorítmica e aprimoramento da infraestrutura digital. Este estudo busca contribuir para o campo da Saúde Digital ao oferecer uma leitura crítica sobre os caminhos e obstáculos associados ao uso de redes sociais e inteligência artificial na vigilância sanitária.

Palavras-chave: Saúde coletiva; Vigilância sanitária; Redes sociais; Inteligência artificial; Aprendizado de máquina; Revisão sistemática.

Are Health Surveillance Models Supported by Social Media Data a Viable Reality?

ABSTRACT

This article presents a systematic literature review aimed at analyzing how data from social media platforms — particularly X (formerly Twitter)— have been employed in the formulation of predictive models applied to health surveillance, using machine learning techniques. Twenty-three studies published between 2018 and 2021 were examined, covering the period immediately before and after the COVID-19 pandemic. These studies applied analytical algorithms to georeferenced posts in order to detect outbreaks, monitor epidemics, and support decision-making processes in public health. The findings suggest that such models may anticipate risk clusters up to four weeks in advance, enabling more timely responses from health authorities. However, significant limitations to the adoption of these systems were identified, such as the scarcity of geolocated data, susceptibility to informational noise, and technical and ethical constraints related to automated natural language processing. The review indicates that, although the applicability of these solutions is expanding, their institutional consolidation depends on integrated strategies for validation, algorithmic governance, and the enhancement of digital infrastructure. This study seeks to contribute to the field of Digital Health by offering a critical perspective on the pathways and barriers associated with the use of social media and artificial intelligence in health surveillance.

Keywords: Public health; Health surveillance; Social networks; Artificial intelligence; Machine learning; Systematic review.

Resumen

Este artículo presenta una revisión sistemática de la literatura con el objetivo de analizar cómo los datos provenientes de redes sociales—en particular de la plataforma X (anteriormente Twitter)—han sido empleados en la formulación de modelos predictivos aplicados a la vigilancia sanitaria, utilizando técnicas de aprendizaje automático. Se examinaron veintitrés estudios publicados entre 2018 y 2021, abarcando el periodo inmediatamente anterior y posterior a la pandemia de COVID-19. Estos estudios aplicaron algoritmos analíticos a publicaciones georreferenciadas con el fin de detectar brotes, monitorear epidemias y respaldar los procesos decisorios en salud pública. Los hallazgos sugieren que dichos modelos pueden

anticipar focos de riesgo con hasta cuatro semanas de antelación, lo que permite respuestas más oportunas por parte de las autoridades sanitarias. No obstante, se identificaron limitaciones relevantes para la adopción de estos sistemas, como la escasez de datos geolocalizados, la vulnerabilidad al ruido informacional y los desafíos técnicos y éticos asociados al procesamiento automático del lenguaje natural. La revisión indica que, aunque la aplicabilidad de estas soluciones está en expansión, su consolidación institucional depende de estrategias integradas de validación, gobernanza algorítmica y fortalecimiento de la infraestructura digital. Este estudio busca contribuir al campo de la Salud Digital al ofrecer una perspectiva crítica sobre los caminos y obstáculos asociados al uso de redes sociales e inteligencia artificial en la vigilancia sanitaria.

Palabras clave: Salud colectiva; Vigilancia sanitaria; Redes sociales; Inteligencia artificial; Aprendizaje automático; Revisión sistemática.

1 Introdução

A vigilância em saúde constitui um dos pilares estruturantes da saúde pública contemporânea, fundamental para a detecção precoce de surtos, mitigação de riscos epidemiológicos e formulação de respostas estratégicas à população (Borle, Reichel, Niebuhr, & Voelter-Mahlknecht, 2021; Hassan Zadeh, Zolbanin, Sharda, & Delen, 2019). Sistemas de vigilância tradicionalmente dependem de dados oriundos de atendimentos em serviços de saúde, os quais estão sujeitos a atrasos de notificação, lacunas geográficas e vieses institucionais (Masri et al., 2019; Xavier, Olenski, Acosta, Sallum, & Saraiva, 2020). A incorporação de dados extraídos de redes sociais, como a plataforma X (anteriormente Twitter), surge como alternativa para mitigar essas limitações ao proporcionar fluxos contínuos e automatizados de informações, captando em tempo real percepções, sintomas e comportamentos da população (Allam, Dey, & Jones, 2020; Amirzadeh, Sobhaninia, Buckman, & Sharifi, 2023; Thirunavukarasu et al., 2023; Zhang et al., 2022).

No campo da saúde pública, evidências mostram o potencial das redes sociais na detecção de surtos e na análise do impacto de desinformações sobre políticas públicas e campanhas de vacinação (Machado Júnior et al., 2023; Sharevski, Alsaadi, Jachim, & Pieroni, 2022). Essa perspectiva se articula com diretrizes internacionais, como a agenda da Organização Mundial da Saúde para a transformação digital na atenção primária (WHO, 2024), e com a Portaria GM/MS nº 3.691/2024, que reformula o conceito de Saúde Digital no Brasil,

integrando o uso de Inteligência Artificial (IA) nos processos decisórios do Sistema Único de Saúde (SUS).

Apesar dos avanços, a adoção de modelos de vigilância baseados em dados não convencionais enfrenta desafios técnicos e éticos, como a qualidade das postagens, a ausência de georreferenciamento em larga escala e a necessidade de validação por especialistas humanos (Edo-Osagie, Smith, Lake, Edeghere, & De La Iglesia, 2019; Rashid & Wang, 2021). Diante disso, este artigo desenvolve uma revisão sistemática da literatura com o objetivo de identificar de que maneira modelos analíticos baseados em dados oriundos do Twitter têm sido aplicados à vigilância sanitária e quais barreiras se impõem à sua incorporação pela gestão pública em saúde. A pesquisa pretende contribuir com o campo da Saúde Digital ao reunir evidências empíricas sobre o potencial dessas soluções, bem como os limites operacionais que desafiam sua adoção em políticas públicas sustentadas por dados não tradicionais.

2 Metodologia: Revisão sistemática de literatura

Esta revisão sistemática da literatura foi conduzida com base nos princípios metodológicos propostos por Kitchenham (2004), com ênfase em transparência, reprodutibilidade e rigor analítico. O objetivo principal foi identificar e analisar estudos que utilizaram dados de redes sociais, em especial do X, para o desenvolvimento de modelos preditivos aplicados à vigilância sanitária. O delineamento da busca e os critérios de inclusão e exclusão dos estudos foram definidos a partir de uma etapa exploratória da literatura (Dall, Fellingner, & Holzinger, 2022; Finch et al., 2016), com o protocolo metodológico encontra-se detalhado na seção subsequente. A aplicação do método PRISMA foi adaptada para garantir uma triagem sistemática e fundamentar a síntese dos achados.

2.1 Condução da revisão

As palavras-chave foram definidas com base em uma análise preliminar da literatura (Dall et al., 2022; Finch et al., 2016). A string de busca combinou os termos: (“Health Surveillance”) AND (“Social Network” OR “Social Media”) AND (“Machine Learning” OR “Big Data” OR “Artificial Intelligence” OR “predictive model”) AND (“Twitter”) além dos critérios de inclusão e exclusão apresentados na Tabela 1.

A escolha da plataforma X (anteriormente Twitter) como foco da pesquisa fundamenta-se em suas características técnicas e comunicacionais, como a alta frequência de postagem, a brevidade das mensagens e o histórico de uso em contextos de saúde pública, desastres e monitoramento social (Dredze, Paul, Bergsma, & Tran, 2013; Martínez & Pascual,

2020). O intervalo de análise buscou abarcar dois momentos-chave: (i) o período pré-pandemia (2018–2019), marcado por um uso ainda exploratório de redes sociais em vigilância sanitária, e (ii) a fase inicial e de intensificação da pandemia de COVID-19 (2020–2021), que estimulou uma produção científica mais intensa e pragmática em saúde digital. Essa delimitação visa compreender tanto a evolução metodológica quanto os contextos de aplicação dos modelos examinados (Machado Júnior et al., 2023).

Tabela 1 - Critérios de inclusão e exclusão

Critérios de inclusão	Critérios de exclusão
<ol style="list-style-type: none"> 1. Estudos no domínio de gestão da saúde pública, que utilizaram dados de redes sociais para criação de alertas sanitários. 2. Tipos de documentos: artigos publicados em peer-reviewed journals e anais de eventos. 3. Estudos empíricos e/ou experimentais. 4. Artigos que utilizaram dados de redes sociais. 5. Artigos publicados em inglês, português e Espanhol. 6. Artigos publicados a partir de 2018 até 2021. 7. Artigos com identificação espacial ou geográfica. 8. Artigos que aplicaram algoritmos analíticos ou de machine learning na análise dos dados. 	<ol style="list-style-type: none"> 1. Estudos em que o foco do uso de algoritmos seja no entendimento das doenças e não em sua gestão. 2. Estudos que modelem dados que não sejam fruto de interações em redes sociais. 3. Capítulos de livros, editoriais e revisões de literatura (incluindo revisões sistemáticas). 4. Artigos teóricos. 5. Artigos cujas versões completas não estivessem disponíveis eletronicamente. 6. Artigos em outras línguas que não português, inglês e espanhol. 7. Artigos que não aplicaram algoritmos analíticos ou de machine learning em suas análises de dados. 8. Artigos que não tenham identificação geográfica ou espacial.

Fonte: Elaborada pelos autores (2025).

A busca foi conduzida nas bases Scopus, Periódicos Capes, ScienceDirect, Web of Science e Google Scholar, considerando artigos publicados em inglês, português e espanhol. Inicialmente, foram identificados 3.212 documentos (Fase 1). Após aplicação dos filtros temporais e temáticos, restaram 1.845 estudos (Fase 2). A triagem por títulos e palavras-chave eliminou artigos que não apresentavam georreferenciamento, não aplicavam algoritmos analíticos ou não descreviam o processo de obtenção e processamento dos dados, resultando em 457 documentos (Fase 3). A leitura dos resumos e conclusões reduziu a amostra a 60 estudos (Fase 4). Por fim, a leitura integral e detalhada de cada texto permitiu selecionar 23 artigos compatíveis com os critérios definidos (Fase 5). O processo de seleção foi conduzido por dois avaliadores independentes e, em caso de divergência, um terceiro revisor foi consultado, conforme protocolo descrito por (Kitchenham, 2004).

Os dados extraídos foram organizados em uma planilha contendo informações como: título, autores, país, enfermidade estudada, técnicas analíticas utilizadas, abordagem de

geolocalização e objetivo do modelo. A Tabela 2 sintetiza as etapas e os critérios adotados no processo de triagem e seleção dos estudos..

Tabela 2 - Fases adotadas no processo de seleção dos artigos aderentes aos objetivos desta pesquisa

Fase	Critério	Número Artigos
1	Bases = Scopus, Periódicos Capes, Science Direct, Web of Science e Google Scholar Palavras chaves = (“Health Surveillance”) AND ("Social Network" OR "Social Media") AND ("Machine Learning" OR "Big Data" OR "Artificial Intelligence" OR "predictive model") AND (“Twitter”). Idiomas = inglês, português e espanhol.	N = 3.212
2	Data publicação: janeiro 2018 a junho 2021	N = 1.845
3	Leitura dos títulos dos documentos relacionados ao objetivo de pesquisa. Foram excluídos artigos que não incluíam visualização espacial e temporal das informações, não descreviam os algoritmos de Machine Learning utilizados. Foram considerados somente os artigos que utilizaram a rede social Twitter e que descreviam como os dados foram obtidos e processados.	N = 457
4	Leitura dos resumos e conclusões dos documentos relacionados ao objetivo de pesquisa	N = 60
5	Leitura detalhada dos documentos e seleção dos artigos para referências à pesquisa	N = 23

Fonte: Elaborada pelos autores (2025).

3 Análise e categorização dos estudos

Após a seleção dos 23 artigos que compõem a base final desta revisão sistemática, procedeu-se à análise descritiva e categorização dos estudos. A categorização foi guiada por três dimensões centrais: a) região geográfica da análise, b) enfermidade-alvo dos modelos, e c) finalidade da aplicação dos dados oriundos das redes sociais. A delimitação geográfica é fundamental para interpretar o contexto epidemiológico, tecnológico e político em que os estudos foram conduzidos, sendo também um critério para avaliar a aplicabilidade de seus resultados em diferentes realidades socioeconômicas. A análise da doença estudada permite aferir o potencial de replicabilidade dos modelos em surtos com diferentes padrões de comportamento. Por fim, a finalidade do uso da rede social — se para vigilância geral ou alertas específicos — é central para avaliar o grau de sofisticação dos modelos e seu potencial de integração à gestão pública em saúde. Na tabela

Com base nessa estrutura, os dados foram extraídos e organizados, possibilitando identificar dois temas principais: (1) a finalidade dos modelos baseados em dados de redes sociais e (2) os desafios para sua implementação no contexto real da saúde pública. O primeiro tema foi classificado em dois tipos de aplicação: modelos orientados ao monitoramento de temas gerais de saúde, baseados em menções recorrentes a doenças, sintomas ou percepções sociais (Chatfield, Scholl, & Brajawidagda, 2013; Verjovsky et al., 2023); e modelos com foco preditivo, orientados à emissão de alertas precoces de doenças específicas, com base na correlação entre o volume e o conteúdo das postagens e dados oficiais de casos (Cuomo,

Purushothaman, Li, Cai, & Mackey, 2021; Heth, Bemis, & Christiansen, 2018; L. Li et al., 2021).

Tabela 3 - Artigos selecionados e seus objetivos

#Artigo	Primeiro Autor	Ano	Objetivo
1	(Adriani, Azzahro, & Hidayanto, 2020)	2020	Desenvolve <i>framework</i> para identificar emergências sanitárias pelo Twitter e gera informações em mapas na Indonésia.
2	(AlAgha, 2021)	2021	Utiliza modelo não supervisionado para análise de sentimento e classificação por tema dos tweets sobre COVID-19.
3	(Cuomo et al., 2021)	2021	Desenvolve modelo com SVM para mapa de risco de contaminação de COVID-19.
4	(Edo-Osagie et al., 2019)	2019	Propõe modelo supervisionado para classificação de tweets sintomáticos e filtra a relevância.
5	(Euzebio et al., 2020)	2020	Desenvolve algoritmo que analisa uma pequena amostra do Twitter para monitorar Dengue em Ribeirão Preto.
6	(Fazeli et al., 2021)	2021	Apresenta um framework para coleta, análise e mapeamento de conteúdos sobre a pandemia de COVID-19.
7	(Hassan Zadeh et al., 2019)	2019	Apresenta modelo de acompanhamento de Influenza utilizando redes sociais e faz correlação temporal.
8	(Heth et al., 2018)	2018	Análise correlação (Pearson) da evolução dos casos semanais de influenza com tweets.
9	(Jain & Cherrickallil, 2018)	2018	Apresenta ferramenta Mendinsights que transforma tweets em informações para ações na saúde.
10	(L. Li et al., 2021)	2021	Desenvolve criação de um sinal de alerta por estado dos EUA usando PLN e ML.
11	(Mackey et al., 2020)	2020	Usa ML não supervisionado (BTM) e desenvolve sinais de casos positivos de Covid durante o início da pandemia.
12	(Martínez & Pascual, 2020)	2020	Apresenta forma de geolocalização dos tweets com Processamento de Linguagem Natural.
13	(Masri et al., 2019)	2019	Desenvolve 2 modelos calibrados por auto-regressão usando Twitter para estimar casos com 1 semana de antecedência.
14	(Nguyen, Nguyen, & Nguyen, 2021)	2021	Desenvolve índice de saúde da população usando Twitter e modelo de redes neurais convolucionais.
15	(Pruss et al., 2019)	2019	Examina com modelo LDA as discussões no Twitter sobre a epidemia de Zikavirus em 2015.
16	(Rashid & Wang, 2021)	2021	Introdução da ferramenta CovidSens que faz uso das redes sociais para alertas sobre COVID.
17	(Şerban, Thapen, Maginnis, Hankin, & Foot, 2019)	2019	Descreve software criado para monitorar propagação de doenças pelo Twitter.
18	(Sidana et al., 2018)	2018	Desenvolve 2 modelos para análise temporal de possíveis doenças analisando informações das redes sociais.
19	(Souza, Assunção, Oliveira, Neill, & Meira, 2019)	2019	Analisa e detecta pontos de risco de Dengue em 2 cidades do Brasil.
20	(Spurlock & Elgazzar, 2020)	2020	Desenvolve aplicação que agrupa usuários por propensão a contaminação de COVID-19 de acordo com sua rede de contatos nas redes sociais.
21	(Tufts et al., 2018)	2018	Análise a correlação da incidência de 14 doenças na Pensilvânia com menções no Twitter no período de 2012 a 2015.
22	(Xavier et al., 2020)	2020	Faz análises exploratórias de dados do Twitter e demonstra o potencial das Redes Sociais como ferramenta de vigilância sanitária.
23	(Z. Li et al., 2020)	2020	Desenvolve modelo usando redes neurais com dados das redes sociais e dados socioeconômicos.

Fonte: Elaborado pelos autores (2025)

O segundo eixo refere-se aos obstáculos práticos relatados nos estudos quanto à adoção institucional desses modelos. Entre os principais estão a confiabilidade e representatividade dos dados, a ausência frequente de georreferenciamento, a dependência de validações humanas mesmo em modelos automatizados, e a interferência de ruídos informacionais, como fake news ou variações de humor social causadas por anúncios oficiais (Masri et al., 2019; Rashid & Wang, 2021; Xavier et al., 2020). As seções seguintes aprofundam esses dois temas centrais, sintetizando os achados da literatura.

3.1 Análise descritiva dos estudos incluídos

Do total da amostra, 44% dos estudos foram publicados antes da pandemia de COVID-19 e 56% durante sua ocorrência, o que indica que o interesse por modelos preditivos já existia anteriormente, mas foi intensificado no contexto da emergência sanitária global (Machado Júnior et al., 2023).

Em relação à qualidade das publicações, identificou-se que cerca de 35% dos artigos foram publicados em periódicos de alto fator de impacto, como *Artificial Intelligence Review*, *IEEE Transactions on Knowledge and Data Engineering* e *Information Processing & Management*. A maior parte dos trabalhos foi publicada em periódicos da área médica ou de ciência da computação aplicada, com limitada presença em revistas especializadas em gestão pública e saúde coletiva, o que evidencia uma lacuna relevante na interlocução entre tecnologia e políticas de saúde pública. Adicionalmente, apenas um dos estudos analisados foi publicado em língua portuguesa, demonstrando que a produção nacional, embora existente, ainda é minoritária na literatura internacional (Pruss et al., 2019).

Do ponto de vista geográfico, observa-se forte concentração de estudos realizados nos Estados Unidos (65%), seguidos de pesquisas localizadas no Brasil, Reino Unido, Indonésia, Espanha, França e Índia. Regiões como África e Oceania estão ausentes da amostra, o que limita a representatividade global dos resultados (AlAgha, 2021; Cuomo et al., 2021; Heth et al., 2018; Jain & Cherikkallil, 2018; Masri et al., 2019; Nguyen et al., 2021; Spurlock & Elgazzar, 2020). Essa concentração em países desenvolvidos evidencia a desigualdade de infraestrutura e acesso à tecnologia, evidenciando uma barreira adicional à adoção desses sistemas por gestores de saúde em regiões mais vulneráveis.

4 Achados e Análises

4.1 Tema 1 – Finalidade do Sistema Criado

A análise dos estudos revelou duas abordagens predominantes em relação à finalidade dos sistemas baseados em dados do Twitter: (i) modelos voltados ao monitoramento geral de temas relacionados à saúde e (ii) modelos preditivos voltados à geração de alertas específicos para doenças com padrões conhecidos. A primeira abordagem visa captar menções recorrentes a tópicos de saúde, sintomas, percepções sociais ou preocupações coletivas, sem necessariamente vincular os dados a uma enfermidade específica. Nesse caso, os sistemas atuam como sensores comportamentais, permitindo que gestores de saúde identifiquem rapidamente mudanças na atenção pública, circulação de desinformação ou demandas emergentes de informação e serviços (Machado Júnior et al., 2023; Sharevski et al., 2022; Xavier et al., 2020).

A segunda abordagem concentra-se em modelos com capacidade preditiva, que correlacionam padrões linguísticos e volumétricos das postagens com séries temporais de casos notificados, gerando sinais antecipados de possíveis surtos. Esses modelos buscam não apenas descrever comportamentos passados, mas apoiar a tomada de decisão em tempo oportuno, a partir da análise automatizada do conteúdo, da localização e da evolução temporal das mensagens (Cuomo et al., 2021; Heth et al., 2018; L. Li et al., 2021).

A distinção entre essas abordagens tem implicações diretas na forma como os sistemas são concebidos. Modelos voltados ao monitoramento geral operam com vocabulário amplo e sensível às tendências sociais, sendo úteis para captar percepção pública, avaliar políticas ou rastrear rumores. Já os modelos preditivos exigem vocabulário específico, validação cruzada com dados oficiais e calibração de algoritmos, assegurando maior precisão na geração de alertas direcionados (Edo-Osagie et al., 2019; Hassan Zadeh et al., 2019; Masri et al., 2019).

4.1.1 Estudos que propõe modelos preditivos para Alertas precoces

Entre os estudos analisados, destaca-se um grupo que buscou desenvolver modelos com capacidade de gerar alertas sanitários antecipadamente, utilizando dados do Twitter como insumo primário para detecção de sinais de surtos epidemiológicos. Esses modelos foram majoritariamente aplicados à COVID-19, Influenza e Zika, e validaram suas previsões por meio de correlação com dados oficiais de saúde ou com outros indicadores indiretos, como buscas no Google ou menções em notícias jornalísticas.

O estudo de L. Li et al., (2021) analisou 14 milhões de tweets postados nos Estados Unidos entre 20 de janeiro e 10 de março de 2020, período anterior à declaração oficial da pandemia pela OMS. O modelo gerou 24 alertas antecipados em nível estadual e apresentou correlações fortes com buscas no Google ($r = 0,8$) e com notícias ($r = 0,7$), com defasagem de 16 e 17 dias, respectivamente.

Cuomo et al., (2021) também propuseram um sistema de alerta para COVID-19, utilizando tweets geolocalizados nos condados norte-americanos entre 3 e 17 de março de 2020. Após filtragem de postagens que relatavam sintomas ou experiências com a doença, aplicou-se um classificador SVM, que apresentou correlação significativa com os dados oficiais de incidência e um intervalo de antecedência de 13 dias.

Heth et al., (2018), por sua vez, exploraram o uso de tweets para antecipação de surtos de Influenza em um condado dos EUA em 2016. A partir de postagens contendo o termo “influenza”, os autores construíram uma série temporal correlacionada com dados de testes positivos e atendimentos hospitalares, obtendo correlações moderadas ($r = 0,54$ e $r = 0,61$) e um intervalo de antecipação de uma semana.

Já o modelo de Hassan Zadeh et al., (2019), aplicado à Influenza entre 2013 e 2015, classificou 1.000 tweets para distinguir postagens informativas daquelas que indicavam infecção. Os dados serviram para gerar mapas de calor com base na geolocalização dos tweets e apresentaram correlação elevada ($r = 0,9$) com os dados oficiais, com defasagem temporal de um mês.

Para a Zika, Masri et al., (2019) analisaram tweets de 2016 contendo as palavras “zika” ou “mosquito” e desenvolveram modelos autorregressivos para prever casos confirmados na Flórida e nos EUA. O modelo com melhor desempenho, que combinava número de tweets com dados epidemiológicos das três semanas anteriores, obteve R^2 de 74% para a Flórida e 70% para os EUA, com previsão de uma semana de antecedência.

No contexto latino-americano, Pruss et al., (2019) utilizaram 3,9 milhões de tweets sobre Zika em três idiomas (português, inglês e espanhol), oriundos de 42 países das Américas. Apesar da correlação fraca entre tweets e casos confirmados ($r = 0,398$), o estudo evidenciou que países com maior incidência da doença apresentaram discussões mais intensas sobre seus efeitos negativos, demonstrando coerência temática entre a realidade epidemiológica e a narrativa pública.

Por fim, Edo-Osagie et al., (2019) aplicaram modelos supervisionados e semi-supervisionados para identificar sintomas de asma no Reino Unido, com base em 10 milhões

de tweets. O uso de emojis contribuiu para a acurácia da classificação, e o modelo obteve correlação significativa com dados oficiais ($r = 0,417$), com defasagem de sete dias. Esses estudos demonstram que, embora com variações metodológicas e contextuais, modelos baseados em dados do Twitter podem prever picos de doenças com antecedência significativa, oferecendo suporte relevante para a gestão pública em contextos de emergência.

4.1.2 Estudos que propõe modelos de monitoramento na Vigilância em Saúde

Além dos modelos com capacidade preditiva, a revisão identificou um conjunto expressivo de estudos cujo foco principal era o monitoramento de temas gerais relacionados à saúde pública por meio da análise de postagens no Twitter. Esses estudos não tinham como objetivo prever surtos específicos, mas compreender tendências de comportamento, sentimentos e discursos sociais associados a doenças, sistemas de saúde ou campanhas governamentais, o que também se mostra estratégico para ações de educação em saúde, gestão da comunicação de risco e formulação de políticas públicas.

Allam & Jones, (2020) investigaram a relação entre infodemia e desinformação durante a pandemia de COVID-19, o modelo identificou que uma parcela significativa das postagens analisadas continha informações não confiáveis ou rumores, reforçando a necessidade de monitoramento contínuo da qualidade da informação circulante nas redes.

Amirzadeh et al., (2023) aplicaram modelos de aprendizado de máquina para analisar a percepção pública em relação à vacinação contra COVID-19 nos Estados Unidos. Os dados revelaram flutuações de sentimento relacionadas a anúncios oficiais e eventos adversos reportados, indicando o valor do monitoramento contínuo para ajustar estratégias de comunicação em tempo real.

Nguyen et al., (2021) utilizaram uma abordagem baseada em análise de sentimentos para examinar a reação do público à política de distanciamento social no Vietnã. Os resultados mostraram que a aceitação da política variava significativamente conforme o nível de engajamento da população com as campanhas de informação do governo, demonstrando como o Twitter pode atuar como um termômetro da opinião pública.

Cuomo et al., (2021), além de seu modelo preditivo, também analisaram postagens relacionadas à COVID-19 para identificar padrões discursivos associados à desinformação e à ansiedade social. A partir de análise textual automatizada, detectaram correlações entre o volume de postagens alarmistas e o aumento de buscas por medicamentos não comprovados,

como a hidroxicloroquina, sugerindo que o monitoramento tem papel relevante para mitigar riscos decorrentes de boatos e falsas promessas terapêuticas.

Verjovsky et al., (2023) desenvolveram um modelo para monitoramento da comunicação sobre temas de saúde, com foco em campanhas de vacinação no Brasil. O estudo evidenciou que picos de volume de postagens negativas antecediam quedas na cobertura vacinal, o que reforça a importância do Twitter como fonte estratégica para o planejamento de ações educativas e intervenções preventivas.

Esses estudos indicam que os sistemas de monitoramento baseados em dados de redes sociais oferecem aos gestores públicos ferramentas úteis para avaliar o impacto de suas ações comunicacionais, monitorar sentimentos sociais e rastrear rumores em circulação. Embora não gerem previsões epidemiológicas diretas, seu uso contínuo pode informar decisões estratégicas, especialmente em contextos de crise sanitária, politização da saúde ou hesitação vacinal.

4.2 Tema 2 – Desafios de gestão

Apesar do potencial evidenciado nos estudos revisados, a aplicação prática de modelos baseados em dados do Twitter enfrenta desafios estruturais, técnicos e éticos que dificultam sua adoção em larga escala pelos sistemas públicos de vigilância sanitária. A seguir, são discutidos os principais obstáculos identificados na literatura.

O primeiro desafio refere-se à qualidade e à representatividade dos dados disponíveis na plataforma. A ausência de critérios padronizados para a coleta e georreferenciamento de tweets compromete a precisão dos modelos, especialmente em países em desenvolvimento, onde o acesso à rede e a frequência de uso variam entre grupos populacionais e regiões geográficas (Edo-Osagie et al., 2019; Heth et al., 2018). Além disso, muitos usuários mantêm seus perfis privados ou optam por não permitir a geolocalização, o que reduz o volume de dados espacialmente qualificados para análises territoriais (Masri et al., 2019).

Outro fator limitante reside na dependência de validação humana, mesmo em sistemas automatizados. Diversos estudos exigiram a leitura manual de milhares de tweets para classificar amostras, treinar algoritmos e validar categorias, o que compromete a escalabilidade das soluções e as torna suscetíveis a vieses interpretativos (Cuomo et al., 2021; L. Li et al., 2021; Rashid & Wang, 2021).

A influência de ruídos informacionais é outro desafio recorrente. Postagens impulsionadas por campanhas políticas, humor público, desinformação e até ironia textual

dificultam a extração de sinais válidos por meio de modelos de *machine learning*, sobretudo quando não há refinamento semântico ou uso avançado de PLN (Allam & Jones, 2020; Sharevski et al., 2022). Além disso, alterações no algoritmo do Twitter e suas regras de acesso à API afetam diretamente a replicabilidade dos estudos e a continuidade das pesquisas.

Do ponto de vista ético e legal, poucos estudos abordam com profundidade os riscos relacionados à privacidade dos usuários, à reidentificação por metadados ou ao uso discriminatório de dados sensíveis por sistemas automatizados (Popkova & Sergi, 2022). A ausência de marcos normativos específicos para vigilância digital em redes sociais abre espaço para usos abusivos ou mal interpretados, o que exige atenção redobrada por parte dos gestores públicos e pesquisadores.

Por fim, destaca-se o desafio relacionado a infraestrutura e qualificação profissional nos serviços públicos de saúde, especialmente em países do Sul Global. A implementação de sistemas baseados em IA requer infraestrutura tecnológica robusta, integração com bases de dados oficiais e equipes capacitadas para interpretar os resultados e adequar os modelos às especificidades locais (Amirzadeh et al., 2023; Xavier et al., 2020).

Em conjunto, esses obstáculos indicam que, embora os modelos de vigilância baseados em redes sociais apresentem viabilidade técnica e consistência científica, sua adoção institucional requer um ecossistema de governança, padronização e qualificação ainda em consolidação, sobretudo no âmbito do SUS.

5 Discussões e Oportunidades de Pesquisas Futuras

A presente revisão evidenciou que o uso de dados de redes sociais, especialmente da plataforma X (antigo Twitter), em sistemas de vigilância sanitária baseados em inteligência artificial apresenta viabilidade técnica, respaldo científico e aplicabilidade potencial na gestão pública. No entanto, sua consolidação como ferramenta institucional depende de avanços em diferentes frentes de pesquisa.

Primeiramente, destaca-se a necessidade de desenvolver modelos sensíveis às variações socioculturais, considerando diferenças no uso das redes sociais, nos padrões linguísticos e nas condições tecnológicas locais. A concentração de estudos em países desenvolvidos impõe o desafio de adaptar algoritmos às realidades do Sul Global, de modo a ampliar representatividade e equidade.

Em segundo lugar, é necessário fomentar pesquisas sobre governança algorítmica e ética da vigilância digital, com atenção à proteção da privacidade, mitigação de vieses discriminatórios e formulação de marcos legais adequados ao uso de dados não estruturados em políticas públicas (Popkova & Sergi, 2022).

Ademais, observa-se uma lacuna na integração entre sistemas de saúde e dados de redes sociais em tempo real, o que abre espaço para estudos-piloto voltados à implementação desses modelos em plataformas públicas, com avaliações de custo-efetividade, impacto operacional e aceitabilidade por parte dos profissionais de saúde.

São igualmente escassos os estudos que abordam a percepção da população sobre o uso de dados postados em redes sociais para fins sanitários, o que evidencia a importância de investigações qualitativas sobre confiança pública e engajamento cidadão em ambientes digitais.

Finalmente, futuras pesquisas podem aprofundar a aplicação de modelos híbridos, que combinem PLN avançado com dados clínicos e ambientais, além de explorar o uso de plataformas alternativas ao Twitter, como WhatsApp, YouTube e TikTok — sobretudo em contextos nos quais o micro blog não é predominante.

O fortalecimento da Saúde Digital requer, portanto, um esforço coordenado entre áreas distintas do conhecimento, promovendo diálogo entre epidemiologistas, cientistas de dados, juristas, gestores e cidadãos na construção de soluções tecnológicas responsáveis, transparentes e orientadas ao interesse público.

6 Conclusões e Implicações para a Gestão Pública em Saúde

Os resultados desta revisão sistemática evidenciam que modelos de vigilância sanitária baseados em dados do Twitter representam uma estratégia complementar relevante e promissora aos sistemas tradicionais de notificação e monitoramento epidemiológico. Os estudos analisados demonstram que, quando bem calibrados, esses modelos podem antecipar surtos com até quatro semanas de antecedência, monitorar percepções públicas sobre políticas de saúde, rastrear desinformações e identificar territórios com maior risco ou resistência a intervenções sanitárias. Essa capacidade de resposta ágil, orientada por dados massivos e não estruturados, está alinhada aos princípios da Saúde Digital previstos na Portaria GM/MS nº 3.691/2024, especialmente no que se refere ao apoio à tomada de decisão em tempo real e ao uso de Inteligência Artificial em processos de gestão.

Contudo, a implementação desses sistemas na gestão pública em saúde exige a superação de desafios técnicos, éticos e estruturais ainda pouco resolvidos na literatura. A escassez de dados georreferenciados, a necessidade de validação humana intensiva, os riscos de vieses algorítmicos e a fragilidade de marcos regulatórios sobre privacidade digital impõem restrições significativas à adoção institucional desses modelos. Além disso, a concentração geográfica da produção científica — majoritariamente centrada em países desenvolvidos — reduz a aplicabilidade dos achados a contextos como o brasileiro, caracterizados por desigualdade digital e assimetrias de infraestrutura entre regiões.

Neste cenário, gestores públicos e formuladores de políticas precisam considerar tais modelos não como substitutos dos sistemas tradicionais, mas como instrumentos complementares de apoio à vigilância integrada, especialmente em tempos de crise sanitária, campanhas de vacinação ou eventos ampla repercussão social. A incorporação desses sistemas ao SUS dependerá da articulação entre regulação ética, investimento tecnológico, qualificação de equipes e parcerias interinstitucionais para o desenvolvimento de modelos adaptados à realidade nacional. A governança desses dados — inclusive quanto à transparência, auditabilidade e segurança — constitui elemento central para garantir sua legitimidade e efetividade.

Esta revisão contribui para o campo da Saúde Coletiva ao reunir evidências sobre a viabilidade e os limites do uso de redes sociais e algoritmos de aprendizado de máquina no fortalecimento da vigilância sanitária. Os achados sugerem que há base científica suficiente para o desenvolvimento de políticas-piloto e testes controlados no SUS, desde que sustentados por um ecossistema consistente de validação científica, controle ético e compromisso com a transparência pública.

7 Declaração de Conflito de Interesses

Os autores declaram que não existem conflitos de interesse relacionados à elaboração, condução ou publicação deste artigo. A pesquisa foi desenvolvida de forma autônoma, sem qualquer interferência de natureza financeira, comercial, institucional ou pessoal que pudesse influenciar os resultados apresentados.

8 Contribuições dos Autores

Este manuscrito foi elaborado com base na taxonomia CRediT (Contributor Roles Taxonomy), garantindo a transparência sobre as contribuições individuais de cada autor:

Kleber R. dos Santos: concepção da pesquisa, definição da metodologia, curadoria dos dados, análise formal, redação da versão inicial do manuscrito e elaboração das visualizações.

Daielly M. N. Mantovani: concepção da pesquisa, definição da metodologia, análise formal, supervisão do estudo e revisão crítica do manuscrito.

Celso Machado Jr: análise formal e revisão crítica do manuscrito.

Guilherme A. Leal: análise formal e revisão crítica do manuscrito.

Todos os autores participaram ativamente das etapas do estudo, revisaram e aprovaram a versão final submetida, assumindo responsabilidade conjunta pela integridade e originalidade do conteúdo apresentado.

9 Referências

- Adriani, M., Azzahro, F., & Hidayanto, A. N. (2020). Disease surveillance in Indonesia through Twitter posts. *Journal of Applied Research and Technology*, 18(3). <https://doi.org/10.22201/icat.24486736e.2020.18.3.1091>
- AlAgha, I. (2021). Topic Modeling and Sentiment Analysis of Twitter Discussions on COVID-19 from Spatial and Temporal Perspectives. *Journal of Information Science Theory and Practice*, 9(1), 35–53. <https://doi.org/10.1633/JISTaP.2021.9.1.3>
- Allam, Z., Dey, G., & Jones, D. (2020). Artificial Intelligence (AI) Provided Early Detection of the Coronavirus (COVID-19) in China and Will Influence Future Urban Health Policy Internationally. *AI*, 1(2), 156–165. <https://doi.org/10.3390/ai1020009>

- Allam, Z., & Jones, D. S. (2020). Pandemic stricken cities on lockdown. Where are our planning and design professionals [now, then and into the future]? *Land Use Policy*, *97*, 104805. <https://doi.org/10.1016/j.landusepol.2020.104805>
- Amirzadeh, M., Sobhaninia, S., Buckman, S. T., & Sharifi, A. (2023). Towards building resilient cities to pandemics: A review of COVID-19 literature. *Sustainable Cities and Society*, *89*, 104326. <https://doi.org/10.1016/j.scs.2022.104326>
- Borle, P., Reichel, K., Niebuhr, F., & Voelter-Mahlknecht, S. (2021). How Are Techno-Stressors Associated with Mental Health and Work Outcomes? A Systematic Review of Occupational Exposure to Information and Communication Technologies within the Technostress Model. *International Journal of Environmental Research and Public Health*, *18*(16), 8673. <https://doi.org/10.3390/ijerph18168673>
- Chatfield, A. T., Scholl, H. J. (Jochen), & Brajawidagda, U. (2013). Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services. *Government Information Quarterly*, *30*(4), 377–386. <https://doi.org/10.1016/j.giq.2013.05.021>
- Cuomo, R. E., Purushothaman, V., Li, J., Cai, M., & Mackey, T. K. (2021). A longitudinal and geospatial analysis of COVID-19 tweets during the early outbreak period in the United States. *BMC Public Health*, *21*(1), 793. <https://doi.org/10.1186/s12889-021-10827-4>
- Dall, M., Fellingner, J., & Holzinger, D. (2022). The link between social communication and mental health from childhood to young adulthood: A systematic review. *Frontiers in Psychiatry*, *13*, 944815. <https://doi.org/10.3389/fpsy.2022.944815>
- Dredze, M., Paul, M. J., Bergsma, S., & Tran, H. (2013). *Carmen: A Twitter Geolocation System with Applications to Public Health*.

- Edo-Osagie, O., Smith, G., Lake, I., Edeghere, O., & De La Iglesia, B. (2019). Twitter mining using semi-supervised classification for relevance filtering in syndromic surveillance. *PLOS ONE*, *14*(7), e0210689. <https://doi.org/10.1371/journal.pone.0210689>
- Euzebio, C., Agy, S., Boldorini Jr., C., Porto, L., Alcarás, J. R., Martinez, A., & Ruiz, E. (2020). Statistical analysis of small twitter data collection to identify dengue outbreaks. *Anais Do Symposium on Knowledge Discovery, Mining and Learning (KDMiLe 2020)*, 17–24. Brasil: Sociedade Brasileira de Computação. <https://doi.org/10.5753/kdmile.2020.11954>
- Fazeli, S., Zamanzadeh, D., Ovalle, A., Nguyen, T., Gee, G., & Sarrafzadeh, M. (2021). *COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations*.
- Finch, K. C., Snook, K. R., Duke, C. H., Fu, K.-W., Tse, Z. T. H., Adhikari, A., & Fung, I. C.-H. (2016). Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. *Natural Hazards*, *83*(1), 729–760. <https://doi.org/10.1007/s11069-016-2327-8>
- Hassan Zadeh, A., Zolbanin, H. M., Sharda, R., & Delen, D. (2019). Social Media for Nowcasting Flu Activity: Spatio-Temporal Big Data Analysis. *Information Systems Frontiers*, *21*(4), 743–760. <https://doi.org/10.1007/s10796-018-9893-0>
- Heth, Z., Bemis, K., & Christiansen, D. (2018). Correlation of Tweets Mentioning Influenza Illness and Traditional Surveillance Data. *Online Journal of Public Health Informatics*, *10*(1), 2579. <https://doi.org/10.5210/ojphi.v10i1.8773>
- Jain, A., & Cherikkallil, S. (2018). Medinsights: Twitter Based Platform for Health Care Analytics. *Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, (Icirca)*, 1104–1109. <https://doi.org/10.1109/ICIRCA.2018.8597360>

- Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.
- Li, L., Gao, L., Zhou, J., Ma, Z., Choy, D. F., & Hall, M. A. (2021, abril 17). *Can Social Media Data Be Utilized to Enhance Early Warning: Retrospective Analysis of the U.S. Covid-19 Pandemic*. Health Informatics. <https://doi.org/10.1101/2021.04.11.21255285>
- Li, Z., Li, X., Porter, D., Zhang, J., Jiang, Y., Olatosi, B., & Weissman, S. (2020). Monitoring the Spatial Spread of COVID-19 and Effectiveness of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics. *JMIR Research Protocols*, 9(12), e24432. <https://doi.org/10.2196/24432>
- Machado Júnior, C., Mantovani, D. M. N., De Sandes-Guimarães, L. V., Romeiro, M. D. C., Furlaneto, C. J., & Bazanini, R. (2023). Volatility of the COVID-19 vaccine hesitancy: Sentiment analysis conducted in Brazil. *Frontiers in Public Health*, 11, 1192155. <https://doi.org/10.3389/fpubh.2023.1192155>
- Mackey, T., Purushothaman, V., Li, J., Shah, N., Nali, M., Bardier, C., ... Cuomo, R. (2020). Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Inveillance Study. *JMIR Public Health and Surveillance*, 6(2), e19509. <https://doi.org/10.2196/19509>
- Martínez, N. J. F., & Pascual, C. P. (2020). Knowledge-based rules for the extraction of complex, fine-grained locative references from tweets. *Revista Electrónica de Lingüística Aplicada*, 19(1). Recuperado de Url: <http://www.aesla.org.es/ojs/index.php/RAEL/article/view/368>
- Masri, S., Jia, J., Li, C., Zhou, G., Lee, M.-C., Yan, G., & Wu, J. (2019). Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. *BMC Public Health*, 19(1), 761. <https://doi.org/10.1186/s12889-019-7103-8>

- Nguyen, H., Nguyen, T., & Nguyen, D. T. (2021). A graph-based approach for population health analysis using Geo-tagged tweets. *Multimedia Tools and Applications*, *80*(5), 7187–7204. <https://doi.org/10.1007/s11042-020-10034-0>
- Popkova, E. G., & Sergi, B. S. (2022). Digital public health: Automation based on new datasets and the Internet of Things. *Socio-Economic Planning Sciences*, *80*, 101039. <https://doi.org/10.1016/j.seps.2021.101039>
- Pruss, D., Fujinuma, Y., Daughton, A. R., Paul, M. J., Arnot, B., Albers Szafir, D., & Boyd-Graber, J. (2019). Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLOS ONE*, *14*(5), e0216922. <https://doi.org/10.1371/journal.pone.0216922>
- Rashid, M. T., & Wang, D. (2021). CovidSens: A vision on reliable social sensing for COVID-19. *Artificial Intelligence Review*, *54*(1), 1–25. <https://doi.org/10.1007/s10462-020-09852-3>
- Şerban, O., Thapen, N., Maginnis, B., Hankin, C., & Foot, V. (2019). Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing and Management*, *56*(3), 1166–1184. <https://doi.org/10.1016/j.ipm.2018.04.011>
- Sharevski, F., Alsaadi, R., Jachim, P., & Pieroni, E. (2022). Misinformation warnings: Twitter’s soft moderation effects on COVID-19 vaccine belief echoes. *Computers & Security*, *114*, 102577. <https://doi.org/10.1016/j.cose.2021.102577>
- Sidana, S., Amer-Yahia, S., Clausel, M., Rebai, M., Mai, S. T., & Amini, M.-R. (2018). Health Monitoring on Social Media over Time. *IEEE Transactions on Knowledge and Data Engineering*, *30*(8), 1467–1480. <https://doi.org/10.1109/TKDE.2018.2795606>
- Souza, R. C. S. N. P., Assunção, R. M., Oliveira, D. M., Neill, D. B., & Meira, W. (2019). Where did I get dengue? Detecting spatial clusters of infection risk with social network

- data. *Spatial and Spatio-temporal Epidemiology*, 29, 163–175.
<https://doi.org/10.1016/j.sste.2018.11.005>
- Spurlock, K., & Elgazzar, H. (2020). Predicting COVID-19 Infection Groups using Social Networks and Machine Learning Algorithms. *2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2020*, 0245–0251. <https://doi.org/10.1109/UEMCON51285.2020.9298093>
- Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>
- Tufts, C., Polsky, D., Volpp, K. G., Groeneveld, P. W., Ungar, L., Merchant, R. M., & Pelullo, A. P. (2018). Characterizing Tweet Volume and Content About Common Health Conditions Across Pennsylvania: Retrospective Analysis. *JMIR Public Health and Surveillance*, 4(4), e10834. <https://doi.org/10.2196/10834>
- Verjovsky, M., Barreto, M. P., Carmo, I., Coutinho, B., Thomer, L., Lifschitz, S., & Jurberg, C. (2023). Political quarrel overshadows vaccination advocacy: How the vaccine debate on Brazilian Twitter was framed by anti-vaxxers during Bolsonaro administration. *Vaccine*, 41(39), 5715–5721. <https://doi.org/10.1016/j.vaccine.2023.07.075>
- WHO. (2024). *Digital transformation handbook for primary health care: Optimizing person-centred point of service systems*.
- Xavier, F., Olenski, J. R. W., Acosta, A. L., Sallum, M. A. M., & Saraiva, A. M. (2020). Análise de redes sociais como estratégia de apoio a vigilância em saúde durante a Covid-19. *Estudos Avancados*, 34(99), 261–282. <https://doi.org/10.1590/S0103-4014.2020.3499.016>

Zhang, Y., Chen, K., Weng, Y., Chen, Z., Zhang, J., & Hubbard, R. (2022). An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. *Expert Systems with Applications*, 198, 116882. <https://doi.org/10.1016/j.eswa.2022.116882>

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.