



INSTITUTO DE MEDICINA INTEGRAL PROF. FERNANDO FIGUEIRA

FACULDADE PERNAMBUCANA DE SAÚDE

EXIVIR: EXTRATOR DE INFORMAÇÕES CLÍNICAS E EPIDEMIOLÓGICAS EM VIROSES RESPIRATÓRIAS

Artigo elaborado para conclusão do curso
de graduação em Medicina 2025.

Autora: Izaura Elias Porto

Orientadora: Cristiane Campello Bresani Salvi - Fiocruz Pernambuco

Co-orientador: Frederico Antônio Pereira Ramos

Co-autoras: Luana Lira de Carvalho Plauto, Maria Luiza Lima Costa

RECIFE

2025

Pesquisadores Responsáveis

Estudante Candidato: Izaura Elias Porto

Função: Graduanda do décimo segundo período do curso de Medicina da Faculdade
Pernambucana de Saúde

ORCID: 0009-0007-9450-7476

Estudante Colaborador: Luana Lira de Carvalho Plauto

Função: Graduanda do décimo segundo período do curso de Medicina da Faculdade
Pernambucana de Saúde

ORCID: 0000-0003-2285-0156

Estudante Colaborador: Maria Luiza Lima Costa

Função: Graduanda do décimo primeiro período do curso de Medicina da Faculdade
Pernambucana de Saúde

ORCID: 0000-0001-9067-5703

Orientadora: Cristiane Campello Bresani Salvi

Função: Médica formada pela Universidade de Pernambuco (UPE) em 2000, especialista em Clínica Médica pela Secretaria Estadual de Saúde de Pernambuco, mestre em Saúde Materno-Infantil pelo IMIP desde 2007 e PhD em Saúde Pública pela Fundação Oswaldo Cruz-Fiocruz desde 2013.

Pesquisadora associada ao Grupo de Estudos em Nutrição e Saúde Integral do IMIP e pesquisadora clínica do Laboratório de Virologia e Terapia Experimental (LaViTE) da Fiocruz Pernambuco.

Coorientador: Frederico Antônio Pereira Ramos

Função: Docente e Pesquisador da Faculdade Pernambucana de Saúde, Médico Clínico e Pneumologista. Doutor em Medicina Translacional pela UNIFESP/ IMIP.

ORCID: 0000-0002-5893-512

ARTIGO ORIGINAL

EXIVIR: EXTRATOR DE INFORMAÇÕES CLÍNICAS E EPIDEMIOLÓGICAS EM VIROSES RESPIRATÓRIAS EMERGENTES

Izaura Elias Porto¹

Luana Lira de Carvalho Plauto¹

Maria Luiza Lima Costa¹

Frederico Antônio Pereira Ramos¹

Cristiane Campello Bresani Salvi^{1,3}

¹Faculdade Pernambucana de Saúde (FPS). Av. Mal. Mascarenhas de Moraes, 4861, Recife - PE, Brasil. CEP: 51150-000

²FIOCRUZ Fundação Oswaldo Cruz – Instituto Aggeu Magalhães. Avenida Professor Moraes Rêgo, s/n, Cidade Universitária, Recife – PE. CEP: 50740-465

³Grupo de Estudos em Nutrição e Saúde Integral do IMIP. Rua dos Coelhos, 300, Boa Vista, Recife– PE. CEP: 50070-902

RESUMO

OBJETIVOS: Desenvolver e validar um programa computacional para extração automatizada de dados clínicos de prontuários eletrônicos hospitalares, com foco em síndromes respiratórias virais. **MÉTODO:** Série de casos retrospectiva de pacientes internados com COVID-19 em dois hospitais de referência na primeira onda da pandemia em Recife, compreendendo 315 prontuários eletrônicos, utilizados para extração manual de dados (n 91) e para identificação de termos relacionados aos sinais e sintomas de COVID-19 (n 104). Os termos compuseram o *pipeline* do programa do Extrator de Informações Clínicas e Epidemiológicas em Viroses Respiratórias (EXIVIR), utilizando linguagem *python* e Processamento de Linguagem Natural. **RESULTADOS:** Entre 91 pacientes, a maioria era do sexo masculino, de meia idade, com manifestações de trato respiratório inferior (tosse, dispneia e hipóxia). Sintomas gripais e alterações olfativas e gustativas foram pouco frequentes. Pelo menos 10% dos casos apresentaram sinais de comprometimento do sistema nervoso central. A extração automatizada de dados de 21 prontuários apresentou índice de concordância global de 68% (coeficiente Kappa 0,452), quando comparado à extração manual. **CONCLUSÃO:** O desenvolvimento de ferramentas computacionais para extração estruturada de dados clínicos a partir de prontuários eletrônicos representa um avanço para a pesquisa clínica no contexto do Sistema Único de Saúde, com potencial de ampliar a capacidade de resposta frente a emergências sanitárias.

PALAVRAS-CHAVE: Prontuário Eletrônico; Informática Médica/Mineração de Dados; Inteligência Artificial; Gripe; COVID-19; Sinais e Sintomas.

ABSTRACT

OBJECTIVES: To develop and validate a computational program for the automated extraction of clinical data from hospital electronic health records, focusing on viral respiratory syndromes. **METHODS:** A retrospective case series of patients hospitalized with COVID-19 in two referral hospitals during the first wave of the pandemic in Recife, including 315 electronic health records. These were used for manual data extraction (n = 91) and for identifying terms related to COVID-19 signs and symptoms (n = 104). The identified terms were incorporated into the pipeline of the Clinical and Epidemiological Information Extractor for Respiratory Viral Infections (EXIVIR), developed in Python and based on Natural Language Processing techniques. **RESULTS:** Among the 91 patients, most were middle-aged men presenting with lower respiratory tract symptoms (cough, dyspnea, and hypoxia). Flu-like symptoms and chemosensory changes (anosmia and ageusia) were uncommon. At least 10% of cases showed signs of central nervous system involvement. Automated data extraction from 21 records achieved an overall agreement rate of 68% (Kappa coefficient 0.452) when compared with manual extraction. **CONCLUSION:** The development of computational tools for structured clinical data extraction from electronic health records represents an important step forward for clinical research within Brazil's Unified Health System, with the potential to enhance responsiveness during public health emergencies.

KEYWORDS: Electronic Health Records; Medical Informatics/Data Mining; Artificial Intelligence; Influenza; COVID-19; Signs and Symptoms.

INTRODUÇÃO

Em dezembro de 2019, surgiram casos de pneumonia grave de etiologia desconhecida na cidade de Wuhan, na China¹. O agente etiológico foi descrito como um novo betacoronavírus da síndrome respiratória aguda grave tipo 2 (SARS-CoV-2), de natureza altamente contagiosa². Esse fator, em combinação com infecções subclínicas frequentes, contribuiu para a rápida disseminação da COVID-19 na China, chegando a atingir mais de 58 países em março de 2020³, quando a OMS declarou oficialmente a pandemia.

Nesse contexto, a comunidade científica global direcionou esforços para fortalecer pesquisas clínicas e epidemiológicas, e estratégias de gestão voltadas ao controle da COVID-19⁴. No entanto, durante a pandemia, houve rápido esgotamento dos recursos humanos e materiais dos sistemas de saúde e instituições de pesquisa no mundo inteiro⁵, especialmente, nos países de baixa e média renda⁶. Não obstante, a precisão com a qual a epidemiologia da pandemia de COVID-19 foi representada é passível de questionamentos, particularmente em países com subnotificação dos sistemas de dados secundários, como o Brasil⁷.

No que concerne ao quadro clínico da COVID-19, os abundantes estudos já nos primeiros meses de 2020, derivam predominantemente de casuísticas chinesas, como observado por uma metanálise de 16 estudos. Essa revisão categorizou a doença, conforme alguns sinais e sintomas, em formas leves (tosse, odinofagia, febre, diarreia, cefaleia, mialgia/artralgia, fadiga e perda de olfato e paladar) e graves (pneumonia por COVID-19: dispneia, anorexia, confusão, dor ou pressão torácica e temperatura acima de 38°C)⁸. No Brasil, o primeiro caso de COVID-19 foi confirmado em fevereiro de 2020⁷, seguindo-se a disseminação em todos os estados do país. Em maio de 2020, o Brasil

tornou-se o epicentro da pandemia na América do Sul⁹, porém, não há dados sobre o comportamento clínico da COVID-19 na população brasileira.

É possível que esse cenário se dê pela dificuldade de acesso a dados clínicos primários do Sistema Único de Saúde (SUS) para fins de pesquisa. De fato, em contextos de emergências em saúde pública, a disponibilidade oportuna e confiável de dados clínico-epidemiológicos é essencial, *a priori* para se conhecerem as manifestações clínicas e a história natural do agravo, mas também para estabelecer estratégias de enfrentamento direcionadas às necessidades das diferentes populações. Portanto, este estudo desenvolveu uma ferramenta computacional para mineração de informações clínicas de prontuários eletrônicos de pacientes com infecções virais respiratórias.

MÉTODOS

O objetivo central do estudo foi desenvolver e validar um Extrator de Informações Clínicas e Epidemiológicas em Vírus Respiratórias (EXIVIR), utilizando uma série de casos retrospectiva de pacientes internados com COVID-19 na primeira onda da pandemia em Pernambuco. O projeto teve aprovação prévia do Comitê de Ética em Pesquisa do Instituto Aggeu Magalhães-Fiocruz (CAAE 36538320.9.0000.5190; parecer nº 4.343.605). O processo de consentimento livre e esclarecido foi realizado, retrospectivamente por meios digitais, através de contato telefônico com os participantes, incluindo envio e aceite do termo de consentimento livre e esclarecido (TCLE), ou dispensa do mesmo quando do insucesso do contato.

Foram incluídos prontuários eletrônicos de pacientes (PEP) adultos com confirmação laboratorial de infecção pelo SARS-CoV-2, admitidos entre abril e junho de

2020 em dois hospitais públicos de referência no estado: Hospital Getúlio Vargas (HGV) e Hospital Agamenon Magalhães (HAM). Foram excluídos prontuários incompletos, duplicados ou que não puderam ser processados corretamente pelo EXIVIR. Não houve cálculo amostral, pois se trata de um estudo de série de casos, com o objetivo de incluir todos os pacientes elegíveis no período pré-definido, sem intenção de testar hipóteses.

O processo de amostragem foi realizado por conveniência identificando-se, inicialmente, as internações em leitos COVID-19 no período definido, em ambos os serviços. Posteriormente, foram selecionados os casos com diagnóstico confirmado de COVID-19. Os casos foram definidos com base em resultados do teste molecular para SARS-CoV-2 (RT-PCR) em swab nasofaríngeo, que foram acessados na plataforma Gerenciador de Ambiente Laboratorial (GAL) do Laboratório Central de Saúde de Pernambuco (LACEN-PE).

As variáveis de interesse basearam-se no quadro clínico de síndromes virais respiratórias, abrangendo manifestações clínicas gerais, osteomusculares, urinárias, respiratórias, gastrointestinais, cardiovasculares, neurológicas, e dermatológicas (**Quadro 1**). Também foram registradas as seguintes variáveis epidemiológicas: idade, sexo biológico, procedência, dias de sintomas até a admissão e até a realização do swab.

Quadro 1. Variáveis correspondentes às manifestações clínicas das síndromes virais respiratórias consideradas para o desenvolvimento computacional do Extrator de Informações Clínicas e Epidemiológicas em Virose Respiratórias (EXIVIR).

Manifestações	Variáveis de interesse
---------------	------------------------

Gerais	Febre, astenia, rinorreia, obstrução nasal, odinofagia
Osteomusculares	Mialgia, artralgia, artrite, dorsalgia, lombalgia
Gastrointestinais	Náuseas, vômitos, dor abdominal, epigastralgia, diarreia, disenteria
Respiratórias	Tosse, expectoração, dor torácica, dor pleurítica, dificuldade de respirar, dispneia, taquipneia, hipóxia, ventilação mecânica assistida e ventilação não invasiva
Neurológicas	Cefaleia, tonturas, síncope, alteração do paladar, alteração do olfato, convulsões, paresia, parestesia, alteração nível consciência, alteração conteúdo consciência
Dermatológicas	Prurido, rash

Ambos os hospitais utilizam o sistema de prontuário eletrônico MVPEP®, através do qual 315 prontuários foram obtidos em formato PDF (**Figura 1**). Desses, 91 prontuários do HAM tiveram suas evoluções médicas lidas na íntegra, por dois pesquisadores da equipe, para a extração manual dos dados clínicos e epidemiológicos, com registro em planilha de dados, no formato Excel. Por sua vez, outros três pesquisadores realizaram a leitura de 104 prontuários (85 do HGV e 19 do HAM) para verificar como as variáveis de interesse estavam descritas. Nessa etapa, foram rastreados os termos exatos das variáveis (**Quadro 1**), assim como, termos sinônimos, com erros ortográficos, siglas e abreviações, cuja lista foi utilizada para compor o *pipeline* do programa do EXIVIR, juntamente com termos adicionais sugeridos por especialistas clínicos da equipe.

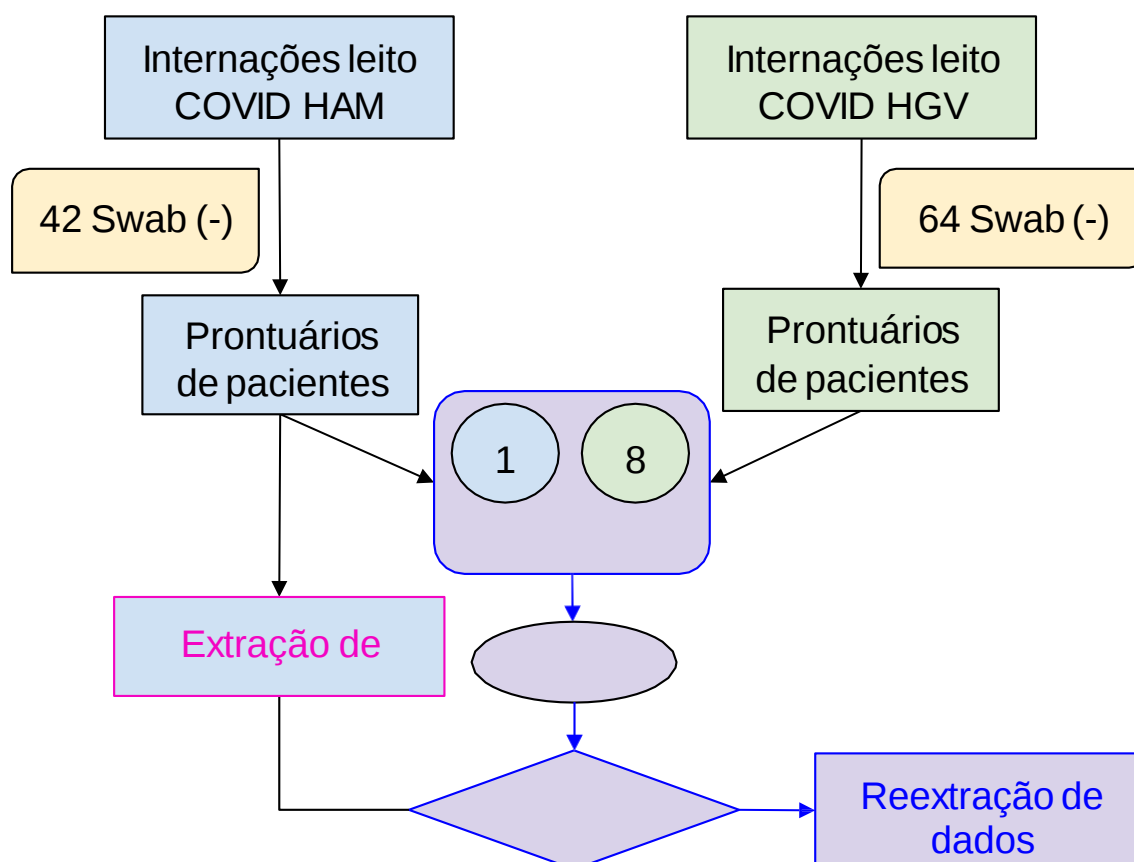


Figura 1. Fluxograma de seleção e leitura dos prontuários eletrônicos de pacientes (PEP) no desenvolvimento do Extrator de Informações Clínicas e Epidemiológicas em Vírus Respiratórias (EXIVIR).

O programa do EXIVIR, contendo a *persona*, as instruções e o *pipeline*, foi desenvolvido em linguagem python, utilizando o modelo de Processamento de Linguagem Natural (PLN) Phi-4 (Microsoft). Para melhorar a eficácia do programa, adotou-se o método *few shot learning*, que utiliza exemplos de textos artificiais classificados para ‘ensinar’ como a tarefa deve ser realizada. Para proceder à etapa de leitura automatizada dos prontuários, primeiramente, lançou-se mão da ferramenta de Reconhecimento de Caractere Óptico (OCR) para que os arquivos de PDF, que se encontravam em formato não editável, fossem processados de imagens para textos manipuláveis em código.

A partir daí, os textos de 21 prontuários foram processados com o EXIVIR para a extração automatizada de dados (presença ou ausência de cada variável de interesse),

resultando em uma base de dados de formato idêntico à planilha com dados extraídos manualmente. Todas as ferramentas adotadas no desenvolvimento do EXIVIR são de código aberto, o que permitiu que o programa fosse executado localmente em ambiente seguro, garantindo a segurança dos dados e sigilo das informações.

O conjunto de evoluções médicas de um mesmo paciente foi tratado como uma única nota a ser processada pelo programa, de modo que cada variável de interesse foi considerada presente se identificada em pelo menos um texto do prontuário. Sendo assim, os dados foram computados como transversais. Em princípio, os dados extraídos manualmente foram utilizados para cálculos das frequências e médias das variáveis de interesse. Posteriormente, foi calculado o índice de concordância geral para todos os resultados obtidos na extração automatizada dos dados com o EXIVIR, em comparação com os respectivos resultados da extração humana. Como teste de hipóteses, lançou-se mão do teste Kappa de Cohen.

RESULTADOS

Na fase de identificação dos termos, foram identificados 287 diferentes palavras, siglas ou abreviações relacionadas aos sinais e sintomas de interesse, reportados nos textos de 104 prontuários. As variáveis do sistema respiratório contribuíram com a maior parte desses termos (60%, 169/287), entre as quais, dispneia (35 termos) e ventilação mecânica assistida (22 termos) apresentaram maior variabilidade de terminologia. As manifestações neurológicas foram o sistema com a segunda maior quantidade de termos (20%, 57/287), em especial na descrição de alteração do nível de consciência (43 termos), seguindo-se de sintomas gerais (12%, 35/287), liderados por febre (20 termos).

Entre os 91 prontuários eletrônicos utilizados para a extração manual de dados, a maioria era de pacientes do sexo masculino (64%), procedente da capital pernambucana (75%), com idade média de 59 anos, e duração dos sintomas de 13 dias à admissão e 7,5 dias à coleta do swab para o teste molecular de SARS-CoV-2, em média (**Tabela1**).

Tabela 1. Dados epidemiológicos identificados em 91 pacientes hospitalizados com COVID-19, através da leitura humana dos prontuários (abril-junho/2020 Recife, Pernambuco, Brasil).

Variável	Distribuição
Sexo (razão feminino:masculino)	33:58
Idade (média em anos)	59
Pacientes procedentes de Recife	75%
Dias de sintomas na admissão (média)	13
Dias de sintomas na coleta do swab (média)	7,5

Os sintomas respiratórios predominaram no quadro clínico, com destaque para a tosse e dificuldade de respirar, presentes em quase 80% e 90% dos casos, respectivamente, enquanto expectoração e dor torácica foram relatados, respectivamente, em 10% e 14% dos pacientes. A febre foi o segundo sintoma mais frequente, acometendo pouco mais da metade dos pacientes, seguido de astenia entre os sintomas gerais, e de mialgia, entre os sintomas osteomusculares, ambos relatados em 20% dos casos. Entre os sintomas gastrointestinais, predominou a diarreia e, entre os sintomas neurológicos, cefaleia, alteração de olfato e alteração do paladar estavam presentes em torno de 10%, cada.

Tabela 2. Sintomas de infecção respiratória viral identificados em 91 pacientes hospitalizados com COVID-19, através da leitura humana dos prontuários (abril-junho/2020 Recife, Pernambuco, Brasil).

Variável	n	Percentual
Sintomas Gerais		
Febre	50	55%
Astenia	21	23%
Rinorreia/obstrução nasal	4	4,4%
Odinofagia	3	3,2%
Sintomas Osteomusculares		
Mialgia	20	22%
Artralgia/artrite	0	-
Dorsalgia/lombalgia	3	3,2%
Sintomas Gastrointestinais		
Náuseas	0	-
Vômitos	3	3,2%
Dor abdominal/epigastria	4	4,4%
Diarréia	8	8,8%
Sintomas Respiratórios		
Tosse	72	79%
Expectoração	9	9,9%
Dor torácica	13	14%
Dificuldade de respirar	81	89%
Sintomas Neurológicos		
Cefaleia	11	12%
Tonturas	1	1,1%
Síncope	1	1,1%
Alteração do paladar	9	9,9%
Alteração do olfato	13	14%
Sintomas Dermatológicos		
Prurido	0	-

Entre os sinais clínicos (**Tabela 3**), as manifestações respiratórias foram as mais frequentes, em especial dispneia/taquipneia e hipóxia, presentes em quase 90% e 74%, respectivamente, seguidas de oxigenioterapia por cateter em quase 70% e ventilação

mecânica assistida em 15%. No sistema neurológico, mais de 5% dos pacientes apresentaram alterações no nível de consciência e quase 10%, do conteúdo da consciência.

Tabela 3. Sinais clínicos identificados em 91 pacientes hospitalizados com COVID-19, através da leitura humana dos prontuários (abril-junho/2020 Recife, Pernambuco, Brasil).

Variável	n	Percentual
Convulsões	1	1,1%
Paresia	0	-
Parestesia	0	-
Alteração do nível de consciência	5	5,5%
Alteração do conteúdo da consciência	9	9,9%
Dispneia/taquipneia	81	89%
Hipóxia	67	74%
Ventilação mecânica assistida	14	15%
Oxigenoterapia por cateter	63	69%
Rash	0	-

A avaliação da acurácia do EXIVIR na extração automática dos mesmos sinais e sintomas clínicos verificados na leitura humana, encontrou que, do total de 972 dados dicotômicos (sim ou não) extraídos de 21 prontuários, houve índice concordância global homem:máquina de 68%, com coeficiente Kappa de Cohen de 0,452 (erro padrão 0,042), indicando nível intermediário de confiabilidade¹⁰.

DISCUSSÃO

A presente casuística é composta de 91 pacientes hospitalizados em leito de enfermaria de dois hospitais de referência para COVID-19 na região metropolitana de Recife, com

infecção confirmada, em sua maioria do sexo masculino, de meia idade, na segunda semana de doença, com quadro clínico, predominantemente, de trato respiratório inferior, marcado por tosse, dispneia subjetiva ou objetiva e hipóxia com necessidade de oxigênio-terapia, na grande maioria dos casos. Excetuando-se febre, presente em metade dos casos, manifestações de síndrome gripal (astenia, mialgia, rinorreia, obstrução nasal, odinofagia) foram pouco frequentes, em especial as de vias aéreas superiores, presentes em menos de 5% dos casos. Manifestações olfativas e gustativas chegaram a acometer até 14% dos pacientes, enquanto até 10% apresentaram alterações de sistema nervoso central. Não se identificaram registros indicativos de exantema (rash, prurido).

Esses achados estão em consonância com o caráter clássico da COVID-19 e com outros poucos estudos nacionais que também descrevem a síndrome respiratória aguda grave como eixo central do quadro, inclusive em alguns subgrupos populacionais, tais como adultos e idosos com doença cardiovascular crônica,¹¹ crianças e adolescentes.¹² Enquanto isso, o destaque que as complicações de sistema nervoso central representaram na COVID-19,¹³ foi também encontrado em nossa amostra, distinguindo-se, portanto, da forma leve da doença síndrome gripal associada a alterações de olfato e paladar (AOP), características.⁸ É importante destacar que essas alterações neurosensoriais foram inicialmente relatadas em torno de 15% dos casos de populações asiáticas, no entanto, com o avançar da pandemia, se mostraram presentes na maioria dos casos europeus (>60%), o que resultou na inclusão global da perda de paladar e/ou olfato como um achado indicativo de COVID-19¹⁴. Séries de casos no Brasil encontraram altas frequências de alterações de paladar e olfato, acima de 50%¹⁵.

É possível que as diferenças geográficas na prevalência de AOP, encontradas na literatura, estejam relacionadas a distintas interações patógeno-hospedeiro, em função de subtipos e virulência do SARS-CoV-2 circulante, em conjunto com variações na

susceptibilidade humana, individuais, hereditárias e raciais. Por sua vez, uma metanálise mostrou que métodos subjetivos de identificação de AOP, tais como o autorrelato, resultam em menores frequências, quando comparados a questionários específicos ou testes de olfato ou paladar (74% vs 61%, p -value = 0.03). Em nossa casuística, pode-se dizer que a frequência de AOP está muito aquém do esperado.¹⁵ Tal achado, provavelmente, se deve ao método de identificação dos sintomas, primariamente registrados para fins de assistência, com base em auto-relato espontâneo dos pacientes ou em questionamento de anamnese não padronizada.

Algumas outras limitações metodológicas incluem a amostragem restrita a internamentos em enfermaria. Nesse contexto, infere-se que se tratam de casos com algum critério de gravidade por ter sido considerada a necessidade de hospitalização, o que condiz com a alta frequência de sintomas compatíveis com pneumonia da COVID-19.⁸ Por outro lado, não foram incluídos pacientes em terapia intensiva, subestimando a frequência de variáveis indicativas de maior gravidade clínica, como hipoxemia e ventilação mecânica. Além disso, a falta de estruturação dos prontuários eletrônicos somada à grande variabilidade ortográfica das variáveis de interesse, incluindo erros de grafia, siglas e abreviações não padronizadas, representa uma fonte de vieses de identificação, tanto na leitura humana, mas principalmente na extração automatizada dos dados.

A despeito desses fatores, que naturalmente desafiam o desempenho de ferramentas de extração automatizada de dados clínicos, o desempenho do EXIVIR, quanto à concordância global, mostrou-se com moderada confiabilidade, quando comparado à extração manual. Resultados melhores podem ser alcançados com maiores tamanhos amostrais. As limitações, no entanto, reforçam a relevância da proposta, ao evidenciar o quanto ferramentas, como o EXIVIR, treinadas no reconhecimento dos

sinais e sintomas em suas mais diversas descrições, no idioma português, são necessárias para a extração adequada de dados no Sistema Único de Saúde.

Iniciativas que buscaram estruturar dados clínicos para vigilância e manejo da COVID-19 no Brasil também evidenciaram desafios relacionados à heterogeneidade das informações registradas, como apontado em análises do sistema de vigilância de SRA-G¹⁶ e no desenvolvimento de plataformas de teleconsulta em enfermagem.¹⁷ Em escala global, trabalhos que aplicaram técnicas de inteligência artificial para predição da necessidade de ventilação mecânica,¹⁸ risco de óbito¹⁹ e para triagem clínica²⁰ destacam o potencial dessas abordagens, mas igualmente reconhecem que seu desempenho depende da qualidade e padronização dos dados de entrada.

Nesse sentido, revisões rápidas sobre o uso de Big Data e inteligência artificial em saúde durante a pandemia²¹ e análises sobre o desenvolvimento de tecnologias digitais no Brasil e no mundo²² reforçam a pertinência do desenvolvimento do EXIVIR, que merece destaque por se basear em notas do mundo real de pacientes com COVID-19 primária antes do início da vacinação, o que o torna singular. Apesar dos desafios técnicos e da necessidade de validação contínua, os resultados preliminares apontam para uma solução capaz de acelerar a produção científica, apoiar a vigilância epidemiológica e subsidiar políticas públicas em contextos de emergência sanitária, podendo constituir-se em futuro legado de inovação e aplicabilidade prática.

CONCLUSÕES

Nossos achados confirmam a centralidade do comprometimento pulmonar no quadro clínico de pacientes hospitalizados com COVID-19, com possibilidade de complicações do sistema nervoso central, e reforçam a pertinência de sistemas computacionais capazes de mapear e organizar dados clínicos primários de forma ágil. Nesse contexto, o uso de

inteligência artificial pode superar as limitações históricas de padronização e acesso a informações no SUS.

REFERÊNCIAS

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3.
2. Mo Yang H, Lombardi Junior LP, Yang AC. Modeling the transmission of the new coronavirus in São Paulo State, Brazil—assessing the epidemiological impacts of isolating young and elder persons. *Math Med Biol*. 2021 Jun 15;38(2):137-177. doi: 10.1093/imammb/dqaa015. PMID: 33434925; PMCID: PMC7928895.
3. Aguilar JB, Faust JS, Westafer LM, Gutierrez JB. A model describing COVID-19 community transmission taking into account asymptomatic carriers and risk mitigation. *medRxiv* [preprint]. 2020 Mar 18:2020.03.18.20037994v5. doi:10.1101/2020.03.18.20037994. .
4. Haleem A, Javaid M, Vaishya R. Effects of COVID-19 pandemic in daily life. *Curr Med Res Pract*. 2020 Mar-Apr;10(2):78-79. doi:10.1016/j.cmrp.2020.03.011. PMID:32292804; PMCID:PMC7147210.
5. Sohrabi C, Mathew G, Franchi T, Kerwan A, Griffin M, Soleil C Del Mundo J, Ali SA, Agha M, Agha R. Impact of the coronavirus (COVID-19) pandemic on scientific research and implications for clinical academic training - A review. *Int J Surg*. 2021 Feb;86:57-63. doi: 10.1016/j.ijssu.2020.12.008. Epub 2021 Jan 12. PMID: 33444873; PMCID: PMC7833269.

6. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020 May 1;368(6490):489-493. doi: 10.1126/science.abb3221. Epub 2020 Mar 16. PMID: 32179701; PMCID: PMC7164387.
7. Prado MF, Antunes BBP, Bastos LSL, Peres IT, Silva AAB, Dantas LF, Baião FA, Maçaira P, Hamacher S, Bozza FA; Analysis of COVID-19 Under-reporting in Brazil. *Rev Bras Ter Intensiva*. 2020;32(2):224-228. doi:10.5935/0103-507X.20200030.
8. Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeftang MM, Spijker R, Hooft L, Emperador D, Domen J, Tans A, Janssens S, Wickramasinghe D, Lannoy V, Horn SRA, Van den Bruel A; Cochrane COVID-19 Diagnostic Test Accuracy Group. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst Rev*. 2022 May 20;5(5):CD013665. doi: 10.1002/14651858.CD013665.pub3. PMID: 35593186; PMCID: PMC9121352.
9. WHO Coronavirus Disease (COVID-19) Dashboard. 2020. Disponível em <https://covid19.who.int/region/amro/country/br> . Acesso em: 3 jul. 2020.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159-74. PMID: 843571.
11. Paiva KM, Hillesheim D, Rech CR, Delevatti RS, Brown RVS, Gonzáles AI, et al. Prevalência e fatores associados à SRAG por COVID-19 em adultos e idosos com doença cardiovascular crônica. *Arq Bras Cardiol*. 2021;117(5):968-75. doi:10.36660/abc.20200955.

12. Santos AL. Características clínicas e fatores de risco que afetam a mortalidade de crianças e adolescentes com COVID-19: um estudo de coorte retrospectivo de âmbito nacional utilizando aprendizado de máquina [tese]. Belo Horizonte: Universidade Federal de Minas Gerais; 2024.
13. Ousseiran ZH, Fares Y, Chamoun WT. Neurological manifestations of COVID-19: a systematic review and detailed comprehension. *Int J Neurosci*. 2023 Jul;133(7):754-769. doi: 10.1080/00207454.2021.1973000. Epub 2021 Sep 27. PMID: 34433369; PMCID: PMC8506813.
14. Kumar AA, Lee SWY, Lock C, Keong NC. Geographical Variations in Host Predisposition to COVID-19 Related Anosmia, Ageusia, and Neurological Syndromes. *Front Med (Lausanne)*. 2021 Apr 29;8:661359. doi: 10.3389/fmed.2021.661359. PMID: 33996863; PMCID: PMC8116582
15. Poudineh M, Amirbeik A, Firouzabadi MD, Hajizadeh M, Kahe F, Babaniamansour S, Poopak A, Tajrishi FZ, Deravi N, Rahmanian M, Ghasemirad H, Malek M, Noroozi M, Fathi M, Keylani K, Firouzabadi AD, Abbasabad GD, Yazdanian F, Ramezanzpour S, Babaniamansour A, Tajrishi FZ, Mohammadi S, Ansari G, Firouzabadi FD, Yousem DM. Olfactory and Gustatory Recovery Time Evaluation of COVID-19: A Systematic Review and Meta-Analysis. *Acta Med Indones*. 2025 Jan;57(1):18-43. PMID: 40263685)
16. Leandro GCW, Cicchero LM, Massago M, Oliveira DG, Bortoli DM, Murillo RSG, Arcoverde MAM, Andrade L, Nihei OK. Avaliação do sistema de vigilância de infecções respiratórias agudas e fatores associados: estudo transversal, Brasil, 2009-2021. *Epidemiol Serv Saude*. 2025;34:e20240555.

17. Oliveira FBM, dos Santos MBL, dos Santos CJA, Sousa MS, Cunha HGS, e Silva LLL, Miranda Junior RNC, Assunção MJSM, Mallet JRS. Teleconsulta de enfermagem: desenvolvimento de plataforma para atendimento de casos de COVID-19. *Arquivos de Ciências da Saúde da UNIPAR*. 2023;27(2):931-47. doi:10.25110/arqsaude.v27i2.2023-023
18. Godoy MF, Chatkin JM, Rodrigues RS, Forte GC, Marchiori E, Gavenski N, Barros RC, Hochhegger B. Artificial intelligence to predict the need for mechanical ventilation in cases of severe COVID-19. *Radiol Bras*. 2023 Mar/Abr;56(2):81–85.
19. Silva R, Silva Neto DR. Inteligência artificial e previsão de óbito por Covid-19 no Brasil: uma análise comparativa entre os algoritmos Logistic Regression, Decision Tree e Random Forest. *Saúde Debate*. 2022;46(Spe8):118–129.
20. Pinasco GC, Silva Júnior JF, Lima Júnior JF, Lima Júnior JF. An interpretable machine learning model for COVID-19 screening. *J Hum Growth Dev*. 2022;32(2):268–274.
21. Ramos MC, Gomes DF, Mello NF, Silva EN, Barreto JOM, Shimizu HE. Big Data e Inteligência Artificial para pesquisa translacional na Covid-19: revisão rápida. *Saúde Debate*. 2022;46(135):1202–1214.
22. Celuppi IC. Uma análise sobre o desenvolvimento de tecnologias digitais em saúde para o enfrentamento da COVID-19 no Brasil e no mundo. *Cad Saúde Pública*. 2021;37(3):e00243220.

Contribuição dos autores: Bresani-Salvi CC: concepção do estudo; trabalho de campo e coleta de dados; organização e consolidação do banco de dados; supervisão do desenvolvimento computacional; análise, interpretação e discussão dos dados e escrita do

manuscrito. Porto IE, Costa MLL, Plauto LLC: participaram de trabalho de campo; coleta de dados; análise, interpretação e discussão dos dados; escrita do manuscrito. Ramos FAP: participou da concepção do estudo e da escrita do manuscrito. Todos os autores aprovaram a versão final do artigo e declaram não haver conflito de interesse.

