Journal of Health Informatics

J Health Inform. 2026;18:e1869
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1869

ARTIGO COMPLETO
CBIS 2026

Preparação de dados e implicações para modelagem preditiva na porta de entrada da APS-Rio

Data preparation and implications for predictive modeling at the primary care entry point in Rio de Janeiro

Elisane Adriana Santos Rodrigues[1], Alaneir de Fátima dos Santos[1], Renato Vimieiro[1], Mariangela Leal Cherchiglia[1], Davi Santos Rodrigues[1]

ORCID: 0000-0003-3588-1947 · 0000-0002-7674-0449 · 0000-0002-7911-2456 · 0000-0001-5622-567X · 0009-0003-6209-2182

[1] Universidade Federal de Minas Gerais, Belo Horizonte (MG), Brasil.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1869

DOI
10.59681/2175-4411.v18.iEspecial.2026.1869

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Elisane Adriana Santos Rodrigues
elisanesantos@pbh.gov.br

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Objetivo: Descrever a preparação e análise exploratória da base integrada da Atenção Primária à Saúde do Rio de Janeiro, visando subsidiar modelos de Inteligência Artificial para triagem na porta de entrada. **Método:** Estudo retrospectivo, quantitativo, com dados secundários do sistema Vitacare (2018–2024), a serem integrados às internações (SIH/SUS) e mortalidade (SIM). Avaliaram-se perfis de demanda, fluxos, completude e consistência das variáveis. **Resultados:** Analisaram-se 130.185.206 atendimentos de 6.414.289 pacientes em 240 unidades. Entre os atendimentos com modalidade codificada, identificou-se predominância de consultas agendadas (63,29%) e, no total da base, volume de visitas domiciliares (31,04%), acompanhados de incompletude de sinais vitais no acolhimento. **Conclusão:** Os paradoxos informacionais encontrados propõem reavaliação contínua e refinamento do banco pela equipe de pesquisa, e limitações em protocolos tradicionais direcionam o desenvolvimento da IA para modelagens alternativas baseadas em texto livre, variáveis contextuais e histórico longitudinal.

Descritores: Atenção Primária à Saúde; Inteligência Artificial; Registros Eletrônicos de Saúde.

ABSTRACT

Objective: To describe the preparation and exploratory analysis of the integrated Primary Health Care database of Rio de Janeiro, aiming to support Artificial Intelligence models for front-door triage. **Method:** A retrospective, quantitative study using secondary data from the Vitacare system (2018–2024), to be integrated with hospital admissions (SIH/SUS) and mortality (SIM) databases. Demand profiles, workflows, completeness, and variable consistency were evaluated. **Results:** A total of 130,185,206 appointments from 6,414,289 patients across 240 units were analyzed. Among appointments with coded access modalities, a predominance of scheduled visits (63.29%) and, in the full database, a high volume of home visits (31.04%) were identified, accompanied by incomplete vital signs data during reception/intake. **Conclusion:** The informational paradoxes identified suggest the need for continuous re-evaluation and database refinement by the research team. Furthermore, limitations in traditional protocols direct the development of AI toward alternative modeling based on free text, contextual variables, and longitudinal history.

Keywords: Primary Health Care; Artificial Intelligence; Electronic Health Records.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: Rodrigues EAS, dos Santos AF, Vimieiro R, Cherchiglia ML, Rodrigues DS. Preparação de dados e implicações para modelagem preditiva na porta de entrada da APS-Rio. J Health Inform 2026;18:e1869. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1869

1 INTRODUÇÃO

A Atenção Primária à Saúde (APS) é a principal porta de entrada do Sistema Único de Saúde (SUS), sendo responsável por acolher, acompanhar e coordenar o cuidado ao longo do tempo. Nos grandes centros urbanos brasileiros, entretanto, a APS convive com desafios persistentes como a sobrecarga devido à alta demanda e fragilidades na organização da porta de entrada. Esses fatores podem comprometer o atributo de acesso de primeiro contato e repercutir sobre a longitudinalidade, a integralidade e a coordenação do cuidado, ampliando o risco de iniquidades no uso dos serviços [1,2].

No município do Rio de Janeiro, a reforma dos serviços de cuidados primários iniciada em 2009 levou a uma expansão rápida da Estratégia Saúde da Família (ESF) e à adoção de prontuário eletrônico em rede. Dados oficiais mostram que a cobertura de saúde da família saltou de cerca de 8,5% para mais de 62% entre 2008 e 2017, com posterior retração para menos de 40% em 2020, em decorrência de mudanças de governos e de modelos de financiamento [3]. Paralelamente, a cidade adotou o sistema Vitacare como prontuário eletrônico da APS, que consolida um grande repositório de dados clínicos e administrativos, potencialmente útil para monitoramento, planejamento e pesquisa [4].

A Carteira de Serviços da APS-Rio, atualizada em 2021, define um escopo amplo de ações essenciais, estratégicas e ampliadas que devem estar disponíveis durante todo o horário de funcionamento das unidades, incluindo acolhimento, consultas médicas e de enfermagem, visitas domiciliares, imunização, procedimentos, exames básicos e ações de planejamento reprodutivo, entre outras [5]. Também explicita que os casos de urgência devem ser atendidos no mesmo dia por médico, enfermeiro ou cirurgião-dentista. No cotidiano dos serviços, contudo, a elevada demanda espontânea e a complexidade dos fluxos de acesso frequentemente exigem arranjos locais que inserem os Agentes Comunitários de Saúde (ACS) no acolhimento humanizado e na organização inicial da porta de entrada. Embora essa medida busque qualificar a escuta e aproximar o serviço da comunidade, a intensa demanda assistencial pode tensionar o equilíbrio entre as atividades internas e o trabalho no território, o que demanda esforços permanentes de gestão para evitar a sobrecarga das equipes e fortalecer a equidade no acesso [6].

Paralelamente, cresce internacionalmente a literatura sobre uso de inteligência artificial (IA) e aprendizado de máquina em APS, predominantemente em modelos de apoio a diagnóstico, estratificação de risco para condições específicas e predição de desfechos clínicos. No cenário brasileiro, contudo, a produção científica correlata ainda se mostra incipiente e escassa, verificando-se uma quase total ausência de aplicações descritas e validadas que sejam voltadas especificamente para a reestruturação da porta de entrada da APS, para a classificação informatizada de risco no momento da chegada ou para a definição automatizada de fluxos direcionados de atendimento, que consigam discriminar as demandas médicas, de enfermagem, vacinais ou puramente administrativas. As experiências nacionais de reorganização do acesso em saúde permanecem, em grande parte, restritas a arranjos puramente organizacionais e analógicos, como o Acolhimento por Equipe e as diretrizes do Acesso Avançado [7], os quais demandam elevado engajamento subjetivo dos trabalhadores, gestão analógica de agendas e processos tradicionais de triagem visual, sem explorar sistematicamente a capacidade preditiva contida nos dados rotineiros e o potencial da IA para subsidiar essas tomadas de decisão [1].

No contexto do projeto guarda-chuva "Utilização de modelos de aprendizado de máquina para a identificação de pacientes de alto risco na Atenção Primária à Saúde" (CAAE 79960224.3.0000.5149, aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais e coparticipação/anuência da Secretaria Municipal do Rio de Janeiro), a APS do Rio de Janeiro oferece um cenário singular para o desenvolvimento de modelos inteligentes. A viabilidade técnica de integração entre os microdados de atendimento da atenção primária (Vitacare), as internações na rede secundária e terciária e os registros vitais de mortalidade confere ao projeto a oportunidade não apenas de caracterizar os padrões vigentes de acesso e fluxo assistencial na porta de entrada, mas também, em etapas subsequentes e avançadas da pesquisa, correlacionar de forma preditiva as decisões de direcionamento adotadas na chegada dos usuários com a ocorrência de desfechos graves desfavoráveis e evitáveis em saúde.

Este artigo tem como objetivo apresentar resultados parciais da fase de análise e descrever a preparação de uma base de dados da APS do Rio de Janeiro como etapa inicial para o desenvolvimento de um modelo de IA automatizado na porta de entrada. O foco está em descrever a dimensão e as características da base, avaliar a qualidade e a completude das variáveis disponíveis, caracterizar o perfil da demanda e dos fluxos assistenciais e explicitar, com base nesses achados, o desenho metodológico planejado para um modelo de IA capaz de classificar a urgência presumida e recomendar o fluxo assistencial mais adequado no momento da chegada do usuário.

2 MÉTODOS

Trata-se de um estudo retrospectivo, observacional e de abordagem quantitativa, baseado em dados secundários oriundos de grandes bases administrativas e clínicas do SUS. A presente pesquisa corresponde a

uma etapa parcial do estudo guarda-chuva intitulado "Utilização de modelos de aprendizado de máquina para a identificação de pacientes de alto risco na Atenção Primária à Saúde", aprovado sob o CAAE 79960224.3.0000.5149. O foco deste artigo reside especificamente na preparação e na análise exploratória do banco de dados da APS do município do Rio de Janeiro, que posteriormente será integrado aos sistemas nacionais de internações hospitalares e de mortalidade, visando subsidiar a construção futura de modelos preditivos de triagem automatizada na porta de entrada das unidades assistenciais.

Para o desenvolvimento do estudo, foram utilizados dados do período de 01/01/2018 a 31/12/2024 a partir da Base da Atenção Primária do Rio de Janeiro, extraída do sistema de prontuário eletrônico Vitacare, contendo os registros de atendimentos realizados nas unidades básicas do município.

A população de referência do estudo foi composta por todos os indivíduos que apresentaram ao menos um atendimento registrado na base da atenção primária carioca ao longo do setênio analisado. Não foram aplicados critérios de exclusão baseados em idade ou condição clínica nesta fase exploratória, embora restrições analíticas, como o recorte exclusivo para a população adulta, possam ser adotadas nas etapas subsequentes de modelagem preditiva.

As informações estruturadas foram organizadas metodologicamente a partir de dois blocos principais de variáveis. O bloco de cadastro, mensurado no nível do paciente, consolidou 10.683.280 registros e abrangeu dimensões sociodemográficas como sexo, idade derivada da data de nascimento, raça ou cor declarada, escolaridade, situação profissional e renda familiar. Esse bloco também contemplou indicadores de condições de saúde (listagem de agravos e doenças crônicas), histórico de óbito, uso de plano de saúde privado, marcadores de vulnerabilidade social e condições de moradia (tipo de domicílio, posse, saneamento, destino do lixo e recebimento de benefícios sociais governamentais).

Por sua vez, o bloco de atendimento, estruturado no nível do episódio assistencial, totalizou 130.185.206 eventos e incluiu identificadores estruturados (código do episódio, chave única do paciente e Cadastro Nacional de Estabelecimentos de Saúde, CNES), atributos de classificação como tipo e subtipo de atendimento, natureza do estabelecimento e categoria profissional do prestador, na qual a atuação dos ACS e demais técnicos foi agregada na macroprofissão de apoio/outros profissionais da saúde ou devidamente mapeada nos códigos de ocupação. Este bloco também registrou componentes temporais que permitiram derivar o dia da semana, mês, ano e turno assistencial.

O conteúdo clínico e administrativo foi mapeado por meio do motivo do atendimento preenchido em campo não estruturado, desfecho da consulta, procedimentos e exames realizados, prescrições e medicamentos administrados, além de medidas clínicas e sinais vitais, que foram armazenados em uma estrutura de dados aninhada, contemplando altura, peso, índice de massa corporal, circunferência abdominal, pressão arterial, frequências cardíaca e respiratória, saturação de oxigênio e temperatura corpórea. Para dar suporte à avaliação dessas variáveis, foram adicionalmente utilizados resumos estatísticos pré-calculados e relatórios técnicos automatizados que servirão para estimar a distribuição e os valores típicos de cada coluna.

Para a caracterização do tipo de atendimento na porta de entrada, analisaram-se especificamente os registros classificados sob as modalidades puras de acesso (Demanda Agendada e Demanda Espontânea), totalizando um subconjunto de 35.841.656 episódios. A análise dos principais motivos de busca por atendimento foi delimitada para o mesmo subconjunto dos episódios de demanda agendada e espontânea. As proporções relativas dos quatro motivos prevalentes foram calculadas utilizando como denominador este subgrupo de acesso direto, permitindo caracterizar a demanda típica que chega à porta de entrada da APS sob esses dois fluxos operacionais.

O pré-processamento e o controle de qualidade dos dados foram executados em um ambiente analítico seguro. Nesse processo, realizou-se a padronização dos tipos de variáveis para formatos homogêneos de data, identificadores e categorias, seguida pela verificação e correção de inconsistências temporais, tais como datas de saídas anteriores à entrada ou registros em períodos futuros, acarretando a exclusão de dados manifestamente inválidos. Assegurou-se também a unicidade dos identificadores de episódios e das chaves de ligação dos pacientes na base analítica final. A avaliação de qualidade subsequente mensurou o percentual de valores nulos, o número de categorias únicas e a distribuição de frequências por coluna nos dois blocos principais. Esse procedimento permitiu a identificação formal de campos com completude criticamente baixa.

Constatou-se que o banco tinha baixa taxa de preenchimento de variáveis relacionadas às medidas antropométricas, sinais vitais, motivo do atendimento e desfecho da consulta. Para lidar com a incompletude, os registros sem informação direta foram categorizados analiticamente como "Não informado/Outros", permitindo manter a integridade do denominador global nas análises descritivas. O mapeamento detalhado dessa incompletude serviu de critério formal para delimitar quais variáveis possuem consistência para atuar como preditores no modelo algorítmico futuro e quais serão mantidas apenas para caracterização descritiva contextual.

Na etapa de derivação de variáveis, novas métricas foram sintetizadas a partir dos dados brutos para

refinar o poder explicativo da base. Na dimensão de tempo e sazonalidade, derivaram-se indicadores de ano, mês, dia da semana, hora e o período do turno assistencial. Para a análise da evolução temporal da demanda na série histórica (2018–2024), adotou-se a métrica de usuários-ano. Essa unidade de medida corresponde à contagem anual de pacientes únicos que realizaram ao menos um atendimento dentro de cada exercício financeiro/assistencial. Ressalta-se que o denominador acumulado do setênio (N=21.039.165 usuários-ano) reflete o somatório do fluxo de usuários ativos em cada ano, permitindo que um mesmo indivíduo seja contabilizado em múltiplos anos caso tenha mantido vínculo contínuo com os serviços de APS ao longo do período. O perfil de utilização dos serviços foi caracterizado pela contagem longitudinal de atendimentos por paciente no acumulado e de forma anualizada, segmentando-se o uso por tipo e subtipo de procedimento e computando-se a distribuição da carga assistencial por categorias profissionais mapeadas em macroprofissões. Os perfis de morbidade e as condições crônicas foram estruturados a partir da contagem de agravos ativos de cada paciente, isolando-se patologias prevalentes como hipertensão, diabetes, distúrbios vasculares periféricos, afecções da coluna vertebral, tuberculose e doença pulmonar obstrutiva crônica. Na dimensão de vulnerabilidade social, manteve-se o indicador binário original da base, prevendo-se a combinação futura dos marcadores habitacionais e de saneamento para a estruturação de um escore de privação sintético. Para os registros de campo não estruturado contidos no motivo da consulta, aplicou-se mapeamento semântico baseado em dicionários de palavras-chave, o qual fará a conversão das queixas em grandes categorias diagnósticas e de demanda, a exemplo de dor, sintomas respiratórios, solicitação de exames, renovação de receituário e vacinação.

A análise exploratória propriamente dita desenvolveu-se em três níveis metodológicos distintos. O primeiro nível examinou a dimensão e a cobertura temporal da base e quantificou o volume absoluto de atendimentos, pacientes únicos e a infraestrutura de rede composta por 240 estabelecimentos distribuídos entre Clínicas da Família, Centros Municipais de Saúde e demais arranjos. O segundo nível concentrou-se no comportamento dos fluxos e nos padrões temporais, ao avaliar a série histórica de atendimentos e a produtividade por categoria profissional e subtipo de procedimento. O terceiro nível mensurou a qualidade dos registros e suas implicações diretas para a modelagem estatística. Analisou-se a completude dos campos clínicos e demográficos para selecionar o conjunto de preditores robustos e identificar quais atributos clínicos relevantes não seriam viáveis para compor o núcleo algorítmico principal, devido à taxa de dados omissos. Todas as análises computacionais foram operacionalizadas por meio de técnicas de estatística descritiva complementadas por visualizações de dados em rotinas construídas nas linguagens Python e R em um ambiente computacional fechado e protegido.

Os aspectos éticos foram rigorosamente observados; a pesquisa está integrada ao projeto aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais sob o CAAE 79960224.3.0000.5149 e conta com a anuência das instâncias de ética em pesquisa da Secretaria Municipal de Saúde do Rio de Janeiro. Nesta fase de processamento, foram manipuladas exclusivamente bases de dados pseudonimizadas, nas quais quaisquer identificadores diretos dos cidadãos foram previamente removidos e substituídos por códigos artificiais de identificação. O procedimento de pareamento e *linkage* entre os bancos de dados da Atenção Primária está sob a responsabilidade de uma equipe técnica autorizada em um servidor institucional dedicado, garantindo que os pesquisadores acessem apenas a base analítica anonimizada para fins de exploração estatística. O desenvolvimento do estudo cumpre integralmente as normativas das Resoluções 466/2012 e 510/2016 do Conselho Nacional de Saúde, bem como as determinações legais da Lei Geral de Proteção de Dados Pessoais e mantém controles rígidos de acesso restrito aos dados, garantia de uso exclusivo para fins de investigação científica e previsão de descarte seguro do repositório após a conclusão regulamentar do projeto.

3 RESULTADOS

A análise descritiva da base integrada de dados da Atenção Primária à Saúde do município do Rio de Janeiro (APS-Rio), que abrangeu o período de 01/01/2018 a 31/12/2024, consolidou o volume de 130.185.206 registros de episódios assistenciais vinculados a 6.414.289 pacientes únicos. Essa produção foi gerada em 240 estabelecimentos de saúde, cuja distribuição tipológica demonstrou o predomínio das Clínicas da Família (CF), responsáveis por 133 unidades (55,42%), seguidas por 103 Centros Municipais de Saúde (CMS) (42,92%) e 4 unidades estruturadas sob outras configurações assistenciais (1,67%), tais como Centros de Saúde Escola e Policlínicas.

A análise longitudinal revelou que o contingente anual de pacientes ativos expandiu-se de 1.381.324 usuários-ano (6,57% do total acumulado de interações anuais no septênio) em 2018 para 4.211.563 (20,02%) em 2024. A distribuição detalhada dessas características estruturais e temporais encontra-se consolidada na Tabela 1.

No tocante à modalidade de acesso à porta de entrada (n=35.841.656 episódios classificados estritamente como demanda agendada ou espontânea), observa-se maior volume de atendimentos agendados (22.687.718; 63,29%) em comparação aos decorrentes de demanda espontânea (13.153.938; 36,70%).

Tabela 1. Características gerais do ecossistema de saúde e evolução temporal dos pacientes na APS-Rio (2018–2024)

Dimensão Analítica	Indicador Assistencial / Estrutural	Frequência (N)	Proporção (%)
Volume Global da Base	Registros de Atendimento Acumulados	130.185.206	–
	Pacientes Únicos Atendidos no Período	6.414.289	–
Infraestrutura de Rede	Clínicas da Família (CF)	133	55,42
	Centros Municipais de Saúde (CMS)	103	42,92
	Outras Unidades (CSE, Policlínicas, CAPS)	4	1,67
	Total de Estabelecimentos Mapeados	240	100,00
Evolução Temporal	Pacientes Únicos Atendidos em 2018	1.381.324	6,57
(Contagem por Ano)	Pacientes Únicos Atendidos em 2019	2.108.334	10,02
	Pacientes Únicos Atendidos em 2020	2.434.872	11,57
	Pacientes Únicos Atendidos em 2021	3.006.074	14,29
	Pacientes Únicos Atendidos em 2022	3.851.397	18,31
	Pacientes Únicos Atendidos em 2023	4.045.601	19,23
	Pacientes Únicos Atendidos em 2024	4.211.563	20,02

*As proporções anuais refletem a participação do volume anual de usuários únicos sobre o somatório acumulado do fluxo no septênio.

Fonte: Elaborado pelos autores (2026).

Os demais atendimentos da base geral (n=94.343.550) corresponderam a registros codificados por protocolos assistenciais e programas específicos. Na divisão por macroprofissão, sobressai a atuação dos enfermeiros e médicos, que concentram a maior parcela da produção registrada, seguidos pelos demais profissionais de saúde, profissionais de saúde bucal e equipe administrativa. Entre os atendimentos classificados sob a modalidade de acesso de entrada (demanda agendada e espontânea; n=35.841.656), identificou-se que os quatro principais motivos de busca responderam por cerca de um quinto de todas as procuras às unidades. Prevaleceram o manejo da hipertensão arterial, a realização de exames dos olhos e da visão, o exame médico geral e a renovação de receitas (Tabela 2).

No mapeamento de dados clínicos, a aferição de altura apresentou incompletude em 92,5% dos registros; a frequência cardíaca mostrou-se ausente em 98,4% dos atendimentos e a medição da circunferência abdominal foi omitida em 99,5% dos episódios. Adicionalmente, verificou-se ausência de preenchimento de 70% na variável *motivo_atendimento* (90.736.722 registros nulos) e de 73% em *desfecho_atendimento* (95.046.250 nulos).

Os dados de cadastro (com base em 10.683.280 registros) apresentaram consistência em variáveis demográficas basais, apontando uma distribuição de 55,6% de usuárias do sexo feminino e 44,4% do sexo masculino. Quanto à raça/cor autoatribuída, identificou-se predominância de indivíduos pardos (5.385.932; 50,4%), brancos (3.641.256; 34,1%) e pretos (1.235.218; 11,6%). Contudo, campos socioeconômicos cruciais exibiram ausência de dados (*missings*), a exemplo da escolaridade (35,82% de nulos), situação profissional (57,59% de nulos) e renda familiar declarada (69,84% de nulos). Quanto à cobertura de saúde complementar, a base confirmou a dependência direta do SUS por parte da população adscrita: 7.724.559 indivíduos declararam não possuir plano de saúde privado (72,3% do cadastro total e 92,0% dos respondentes válidos).

4 DISCUSSÃO

Os resultados da análise exploratória da base APS-Rio revelam a magnitude da produção assistencial do município e evidenciam um paradoxo informacional com implicações diretas para o desenvolvimento de soluções baseadas em IA para a porta de entrada. O volume acumulado superior a 130 milhões de atendimentos reflete a robustez e a maturidade de um sistema de prontuário eletrônico unificado (Vitacare) que opera de forma capilarizada. Esse cenário de informatização em larga escala espelha o sucesso e as barreiras históricas da consolidação de Sistemas de Informação em Saúde no cenário nacional, no qual a transição tecnológica permitiu reestruturar dados clínicos e administrativos antes fragmentados, ao gerar grandes repositórios de dados institucionais prontos para análise [4].

No entanto, a discrepância entre a realidade empírica operacional das unidades de saúde e a estrutura dos dados registrados no sistema impõe uma revisão profunda na estratégia de engenharia de atributos (*feature engineering*) para futuros modelos preditivos. Um ponto crítico reside no fato de que, entre os atendimentos com codificação estrita da modalidade de acesso, a maioria figura sob o rótulo de “Agendada”, contraposta aos tipificados como

Tabela 2. Caracterização dos atendimentos, profissionais e motivos de consulta na APS-Rio (2018–2024)

Dimensão Analítica	Indicador de Fluxo e Demanda	Frequência (N)	Proporção (%)
Tipo de Atendimento*	Agendada	22.687.718	63,29
	Demanda Espontânea	13.153.938	36,70
	Subtotal de Modalidades de Acesso	35.841.656	100,00
Produção por profissional *(Macroprofissão)*	Enfermeiro	52.653.861	40,45
	Médico	29.372.080	22,56
	Outros Profissionais da Saúde	29.160.770	22,40
	Não Classificado	12.927.322	9,93
	Saúde Bucal	5.869.540	4,51
	Administrativo	201.633	0,15
Principais Subtipos *(Procedimentos)*	Visita Domiciliar**	40.410.955	31,04
	Atendimento SOAP	36.600.691	28,11
	Atendimento Técnico de Enfermagem	22.676.658	17,42
	Consulta de Enfermagem	12.573.775	9,66
	Atendimento de Saúde Bucal	4.577.991	3,52
Motivos de Busca *(agendadas e espontâneas; n=35.841.656)*	Hipertensão	3.034.594	8,46
	Exame dos olhos e da visão	2.260.012	6,30
	Exame médico geral	1.306.451	3,64
	Renovação de Receita	1.054.728	2,94

*Proporção referente exclusivamente aos atendimentos codificados como modalidade de acesso de entrada (agendada vs. espontânea). Excluem-se desse subgrupo os atendimentos codificados por protocolos clínicos específicos/linhas de cuidado. **Visita domiciliar corresponde a uma das atribuições do ACS.

Fonte: Elaborado pelos autores (2026).

"Demanda Espontânea". Contudo, essa distinção administrativa pode não refletir fielmente a rotina assistencial, visto que atendimentos por queixas agudas no mesmo dia frequentemente acabam registrados como "Agendada". Como a dinâmica empírica desse preenchimento na recepção pode variar e gerar vieses de classificação, o marcador administrativo bruto do tipo de atendimento não se mostra um indicador isolado suficiente para discriminar o fluxo programático das urgências. Consequentemente, em substituição à dependência exclusiva do rótulo administrativo, o modelo de IA deverá ser estruturado para inferir a real natureza e gravidade da demanda a partir do mapeamento de dados textuais (motivo do atendimento) e do histórico assistencial do paciente.

Essa desconexão entre a jornada real realizada pelo paciente na unidade e os caminhos formais mapeados rigidamente pelos sistemas administrativos de prontuário eletrônico pode ser compreendida sob a ótica da Mineração de Processos (*Process Mining*); investigações de itinerários assistenciais na APS demonstram que gargalos e desvios de fluxo ocorrem frequentemente devido à rigidez burocrática dos softwares, o que pode levar a uma incoerência entre o registro administrativo e a real e imediata demanda clínica por atendimento [6].

A alta densidade de atendimentos registrados sob a responsabilidade da equipe de Enfermagem e demais categorias (que inclui os ACS) corrobora a arquitetura assistencial adotada no Rio de Janeiro. Conforme preconizado pela Carteira de Serviços da APS-Rio [7], a recepção, o acolhimento inicial e o acompanhamento territorial são organizados por microáreas, posicionando o ACS como o primeiro contato do usuário com a rede. O expressivo volume de registros dessas categorias reflete a dinâmica multiprofissional da ESF, que combina grande volume de consultas clínicas com ações territoriais contínuas e inclui o lançamento massivo de Visitas Domiciliares realizadas pelas equipes, além do cadastramento e atualização frequente de dados pelas equipes de campo. Portanto, o alto volume assistencial do ACS documentado na literatura do Vitacare reflete o papel desse profissional como o principal sensor de dados sociodemográficos e epidemiológicos no território.

Um achado que merece maior aprofundamento para o desenho metodológico de um modelo de IA voltado à classificação de risco na chegada foi a elevada taxa de incompletude registrada nas variáveis de sinais vitais e medidas antropométricas. Em cenários hospitalares ou

de pronto atendimento, os modelos tradicionais de triagem (como o protocolo de Manchester) dependem fundamentalmente de métricas fisiológicas para inferir gravidade. Na APS do Rio de Janeiro, a escassez dessas métricas indica que a aferição estruturada de sinais vitais pode ter sido realizada na prática clínica, porém sem o registro correspondente nos campos estruturados do Vitacare.

Consequentemente, resta inviabilizada a construção de classificadores de risco baseados em algoritmos clássicos de urgência biológica. Para contornar essa restrição, a equipe de pesquisa se propôs a fazer revisão do banco junto à instituição ou o modelo de IA em desenvolvimento deverá adotar uma abordagem alternativa, ancorando sua capacidade preditiva em variáveis de contexto e histórico. A literatura de Saúde Digital fundamenta que o uso de técnicas computacionais avançadas, como redes neurais artificiais e o algoritmo *AdaBoost*, apresenta elevada acurácia e capacidade adaptativa para apoiar tomadas de decisão complexas e políticas públicas na APS a partir de dados secundários e perfis populacionais [1]. Os dados demonstram que, em contraposição aos *missings* clínicos, a base oferece trilhas robustas através do histórico longitudinal de utilização de serviços pelo paciente (número de consultas prévias, padrão de absenteísmo), da presença de condições crônicas mapeadas (como hipertensão e diabetes) e do mapeamento semântico dos motivos de consulta textuais.

A elevada proporção de dados ausentes nas variáveis socioeconômicas estruturadas no cadastro dos usuários (renda e escolaridade) também desafia a literatura internacional, que aponta os determinantes sociais de saúde como fortes preditores de vulnerabilidade. Não obstante, os resultados revelam que variáveis indiretas ligadas às condições ambientais e habitacionais (como esgotamento sanitário, destino do lixo, acesso à rede elétrica e recebimento de benefícios sociais) possuem preenchimento superior e uniforme, emergindo como substitutas (*proxies*) robustas para mensurar a vulnerabilidade social dos indivíduos na ausência de dados de renda.

Por fim, a integração longitudinal das bases Vitacare, sistemas de internações hospitalares e de mortalidade poderá estabelecer uma base sólida para a definição de rótulos (*labels*) preditivos robustos e eticamente alinhados ao SUS. A capacidade de rastrear se um indivíduo que buscou a porta de entrada da APS evoluiu para uma internação hospitalar sensível à atenção primária ou para o óbito em uma janela temporal restrita permitirá treinar o modelo de IA não apenas para replicar a decisão burocrática do fluxo atual, mas para prever o risco real de desfechos graves em saúde.

Esse processo será capaz de transformar o classificador automatizado em uma ferramenta de equidade, capaz de sinalizar precocemente indivíduos de alto risco clínico ou social negligenciados pelos processos administrativos convencionais da recepção.

5 CONCLUSÃO

A análise exploratória e a preparação da base de dados integrada da APS do município do Rio de Janeiro evidenciam o expressivo potencial dos registros eletrônicos de saúde rotineiros para subsidiar a inovação tecnológica no SUS. O mapeamento longitudinal de mais de 130 milhões de atendimentos e 10,6 milhões de cadastros confirmou a maturidade e a capilaridade do sistema Vitacare, ao estabelecer um repositório massivo que reflete com fidelidade a densidade da produção assistencial de uma das maiores e mais consolidadas redes municipais de atenção primária do país. Todavia, os achados revelam um importante paradoxo informacional, caracterizado por divergências entre a parametrização administrativa dos dados brutos e a realidade clínico-operacional vivenciada no cotidiano de acolhimento das unidades de saúde.

A imprecisão no registro nominal da modalidade de acesso decorrente da inserção de queixas agudas em agendas dinâmicas do mesmo dia no software, somada à reduzida completude de parâmetros fisiológicos estruturados na recepção, impõe limitações para a aplicação direta de protocolos analógicos tradicionais de classificação de risco baseados estritamente em critérios biomédicos lineares. Diante disso, o cenário desenhado propõe uma reavaliação contínua e um refinamento desta base de dados pela equipe de pesquisa, o que permite compreender a fundo as nuances do registro local. Paralelamente, delineia-se uma mudança de paradigma na engenharia de atributos para o desenvolvimento de soluções de IA voltadas à APS. Demonstra-se que o poder preditivo para gerenciar o acesso e estimar a urgência presumida na chegada pode ser alcançado por meio de modelagens alternativas, estruturadas a partir de variáveis contextuais, indicadores substitutos (*proxies*) de vulnerabilidade social e habitacional, histórico longitudinal de utilização dos serviços e do processamento de linguagem natural aplicado aos motivos de consulta textuais.

Por fim, o processo contínuo de qualificação e reavaliação do banco subsidiará, de forma segura, a consolidação da integração unificada e pseudonimizada entre os microdados da atenção primária e os Sistemas Nacionais de Internações Hospitalares (SIH/SUS) e de Mortalidade (SIM). Essa vinculação longitudinal de bases de dados viabiliza a extração de rótulos preditivos de desfechos clinicamente graves, ao permitir que os modelos de aprendizado de máquina em desenvolvimento sejam treinados de forma contextualizada à realidade local, por prever o risco real de desfechos desfavoráveis e internações sensíveis

à atenção primária. Conclui-se que a transição planejada para um sistema de triagem automatizada baseado em IA apresenta-se como uma estratégia promissora para apoiar as equipes de saúde, conferir maior transparência aos fluxos assistenciais e atuar como um instrumento indutor de equidade na porta de entrada do SUS.

Contribuição dos autores. Todos os autores acessaram e verificaram os dados, e todos os autores aprovaram a versão final para submissão e tiveram a responsabilidade final pela decisão de submeter para publicação. Concepção e delineamento do estudo: EASR, AFS, RV, DSR. Coleta: RV, DSR. Análise e interpretação dos dados: EASR, RV, DSR. Redação e revisão crítica do manuscrito: EASR, AFS, RV, MLC, DSR. Aprovação da versão final: EASR, AFS, RV, MLC, DSR.

Financiamento. O projeto é financiado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Disponibilidade de dados e materiais. Dados desidentificados dos participantes e materiais adicionais do estudo não estão disponíveis publicamente. Os dados poderão ser disponibilizados pelo autor correspondente mediante solicitação razoável.

Conflitos de interesse. Declaramos não haver conflitos de interesse.

Declaração de uso de IA. Durante a preparação deste trabalho, os autores utilizaram Notebook LM para auxiliar na edição de texto e aprimoramento da linguagem. Após o uso desta ferramenta, os autores revisaram e editaram o conteúdo conforme necessário e assumem total responsabilidade pelo conteúdo da publicação.

REFERÊNCIAS

1. Salerno AAP, Vianna RPT, Moraes RM. Técnica boosting: apoio à decisão na Atenção Primária à Saúde. J Health Inform. 2017 Jul-Sep;9(3):87-94: Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/498
2. Mendes EV. A construção social da atenção primária à saúde. Brasília: Conselho Nacional de Secretários de Saúde - CONASS; 2015. Disponível em: https://www.conass.org.br/biblioteca/pdf/A-CONSTR-SOC-ATEN-PRIM-SAUDE.pdf
3. Soranz D, Pinto LF, Penna GO. Themes and Reform of Primary Health Care (RCAPS) in the city of Rio de Janeiro, Brazil. Cien Saude Colet. 2016;21(5):1327-38. doi: https://doi.org/10.1590/1413-81232015215.01022016
4. Lacerda TC, Hammes JF, Fantonelli M, Dalmarco EM, Wazlawick RS. e-SUS APS strategy: Case of success on Primary Care informatization in Brazil. J Health Inform. 2020 Oct-Dec;12(4):138-43. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/754
5. Rio de Janeiro. Secretaria Municipal de Saúde. Carteira de serviços da atenção primária: abrangência do cuidado [Internet]. 3. ed. Rio de Janeiro: SMS; 2021 [citado 2026 out 18]. Disponível em: https://subpav.org/download/impressos/Livro_CarteiraDeServicosAPS_2021_20211229.pdf
6. Xavier EC, Sutile VM, Miranda CM, Carvalho DR. Análise de itinerário dos usuários para atendimento de serviços básicos de saúde, por meio de Mineração de Processos. J Health Inform. 2021 Oct-Dec;13(4):120-7. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/887
7. Rio de Janeiro. Secretaria Municipal de Saúde. Carteira de serviços da atenção primária: diretrizes para a abrangência do cuidado [Internet]. 3. ed. Rio de Janeiro: Secretaria Municipal de Saúde; 2021 [citado 2026 out 19]. Disponível em: https://subpav.org/download/impressos/Livro_CarteiraDeServicosAPS_2021_20211229.pdf

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde