

Capítulo

7

Soberania em Bioinformática Clínica Baseada em IA para o SUS na Próxima Década

Marcelo A. C. Fernandes

Abstract

The next decade will be characterized by the increasing integration of genomic, clinical, and socioeconomic data into healthcare practice, driving advances in precision medicine. However, this transformation takes place in a context of strong dependence on international infrastructures, databases, and analytical tools, limiting technological sovereignty and the adaptation of solutions to the Brazilian population. This chapter proposes, as a national grand challenge, the development of an integrated artificial intelligence-based clinical bioinformatics infrastructure tailored to the Brazilian Unified Health System (SUS). The proposed framework is structured around a layered architecture that integrates heterogeneous data sources, governance and interoperability mechanisms, advanced AI models, and large-scale computational infrastructure, enabling applications such as clinical decision support and genomic surveillance. In addition, the chapter discusses current gaps, key dimensions of the challenge, evaluation metrics, and a roadmap for the 2026-2036 period. By fostering the integration of science, technology, and healthcare systems, the proposal aims to reduce external dependencies, expand equitable access to advanced technologies, and position Brazil as a leader in clinical bioinformatics and digital health.

Resumo

A próxima década será marcada pela crescente integração de dados genômicos, clínicos e socioeconômicos na prática assistencial, impulsionando avanços na medicina de

precisão. No entanto, essa transformação ocorre em um contexto de forte dependência de infraestruturas, bases de dados e ferramentas internacionais, limitando a soberania tecnológica e a adaptação das soluções à realidade da população brasileira. Este capítulo propõe como grande desafio nacional a construção de uma infraestrutura integrada de bioinformática clínica baseada em Inteligência Artificial, orientada ao Sistema Único de Saúde (SUS). A proposta é estruturada a partir de uma arquitetura em camadas que integra dados heterogêneos, mecanismos de governança e interoperabilidade, modelos avançados de IA e infraestrutura computacional em larga escala, com aplicação direta em suporte à decisão clínica e vigilância em saúde. Além disso, são discutidas as principais lacunas atuais, as dimensões do desafio, métricas de avaliação e um roadmap para o período de 2026 a 2036. Ao promover a integração entre ciência, tecnologia e sistema de saúde, a proposta visa reduzir dependências externas, ampliar a equidade no acesso a tecnologias avançadas e posicionar o Brasil como referência em bioinformática clínica e saúde digital.

7.1 Introdução

A próxima década será marcada por uma transformação profunda na saúde, impulsionada pela integração de dados biomédicos em larga escala, em um cenário de crescente dependência tecnológica e necessidade de soberania em dados e infraestrutura. Avanços recentes em sequenciamento de nova geração, terapias avançadas baseadas em ácidos nucleicos e na medicina de precisão têm ampliado significativamente a capacidade de compreensão dos mecanismos biológicos subjacentes às doenças, abrindo novas possibilidades para diagnóstico, prognóstico e tratamento individualizado. Nesse contexto, a bioinformática clínica emerge como um componente central para a tradução desses dados em conhecimento acionável, sendo fortemente dependente de métodos computacionais avançados, em especial aqueles baseados em Inteligência Artificial (IA) [Ghebrehiwet et al. 2024, Alsaedi et al. 2025].

Apesar desse cenário promissor, a capacidade de transformar dados biológicos em decisões clínicas ainda é altamente desigual entre países e regiões. Na prática, a interpretação de variantes genéticas, a identificação de alvos terapêuticos e o monitoramento de patógenos dependem, em grande medida, de infraestruturas computacionais robustas, bases de dados consolidadas e pipelines validados, frequentemente concentrados em poucos países. Essa concentração tecnológica cria um cenário de dependência estrutural para países em desenvolvimento, limitando a autonomia científica, a soberania em dados e a capacidade de resposta a desafios emergentes em saúde pública. No caso brasileiro, esse problema é agravado pela elevada diversidade genética da população, resultante de um processo histórico de miscigenação, que não é adequadamente representada nas principais bases de dados internacionais, comprometendo a acurácia das interpretações clínicas e a efetividade de abordagens baseadas em medicina de precisão [Gao et al. 2023, Nunes et al. 2025].

No Brasil, embora existam iniciativas relevantes em genômica, bioinformática e saúde digital, ainda não há uma infraestrutura nacional integrada que permita a operação coordenada de bioinformática clínica em escala populacional. Programas estratégicos como o Genomas Brasil, conduzido pelo Ministério da Saúde, representam avanços importantes na geração de dados genômicos e na estruturação de capacidades nacionais,

porém ainda enfrentam desafios relacionados à integração sistêmica, interoperabilidade e uso efetivo desses dados em larga escala. A ausência de integração entre dados genômicos, registros clínicos e determinantes sociais, aliada à fragmentação institucional e tecnológica, dificulta a construção de soluções robustas e sustentáveis. Além disso, a incorporação de técnicas avançadas de IA, como modelos baseados em Transformers e abordagens generativas, ainda ocorre de forma pontual e desarticulada, sem a consolidação de um ecossistema nacional [de Oliveira and Lopes-Cendes 2025].

No cenário internacional, observa-se o avanço de iniciativas estruturantes voltadas à integração de dados genômicos e clínicos em larga escala, como o UK Biobank no Reino Unido [Bycroft et al. 2018], o programa All of Us nos Estados Unidos [of Us Research Program Investigators 2019], e o European Health Data Space Regulation (EHDS) na União Europeia [European Commission 2022]. Essas iniciativas estabelecem infraestruturas nacionais integradas que combinam dados biomédicos, plataformas computacionais avançadas e diretrizes robustas de governança, viabilizando aplicações em medicina de precisão e saúde pública. Em contraste, o Brasil ainda não dispõe de uma infraestrutura nacional plenamente integrada e operacional nesse nível, o que reforça a necessidade de uma abordagem estratégica orientada à soberania tecnológica e à adaptação às especificidades da população brasileira.

Diante desse cenário, torna-se imperativo estabelecer um novo patamar para a computação aplicada à saúde no Brasil, no qual a bioinformática clínica deixe de ser um conjunto de iniciativas isoladas e passe a constituir uma infraestrutura estratégica para o Sistema Único de Saúde (SUS). Este capítulo propõe, portanto, a formulação de um grande desafio nacional voltado à construção de uma infraestrutura integrada de bioinformática clínica baseada em IA, com foco na soberania tecnológica, na equidade em saúde e no fortalecimento do ecossistema de inovação em saúde digital para a próxima década [Khan et al. 2024].

7.2 Arquitetura da Infraestrutura Nacional Proposta

Para viabilizar a construção de uma infraestrutura nacional de bioinformática clínica baseada em IA alinhada aos princípios de soberania tecnológica e integração ao SUS, propõe-se uma arquitetura em camadas, projetada para suportar a integração de dados heterogêneos, o processamento em larga escala e a geração de conhecimento clínico acionável. Essa arquitetura organiza-se em cinco camadas principais, interdependentes e complementares: (i) camada de dados, (ii) camada de integração e governança, (iii) camada de IA, (iv) camada computacional e (v) camada de aplicação clínica, conforme ilustrado na Figura 1.1.

Camada de Dados: responsável pela coleta, armazenamento e organização de dados biomédicos em larga escala, incluindo dados genômicos, transcriptômicos, proteômicos, registros clínicos eletrônicos e determinantes sociais de saúde. Essa camada deve contemplar mecanismos de padronização, curadoria e versionamento, garantindo qualidade e rastreabilidade dos dados.

Camada de Integração e Governança: responsável pela interoperabilidade entre diferentes fontes de dados e sistemas institucionais. Inclui a definição de padrões nacionais de interoperabilidade, mecanismos de anonimização e segurança, além da con-



Figura 1.1. Arquitetura simplificada da infraestrutura nacional proposta para bioinformática clínica baseada em IA no SUS.

formidade com a Lei Geral de Proteção de Dados (LGPD). Essa camada também estabelece políticas de governança, controle de acesso e auditoria, fundamentais para garantir confiança e uso ético dos dados.

Camada de IA: dedicada ao desenvolvimento e aplicação de modelos avançados de IA para análise de dados biomédicos complexos e multimodais. Inclui modelos baseados em aprendizado profundo, Transformers, modelos generativos e abordagens de aprendizado federado, além de técnicas de IA Explicável (XAI) para garantir transparência, interpretabilidade e suporte à decisão clínica.

Camada Computacional: composta por infraestruturas de computação de alto desempenho (HPC), ambientes em nuvem e arquiteturas distribuídas e federadas. Essa camada é responsável por viabilizar o processamento eficiente de grandes volumes de dados e o treinamento de modelos de IA em escala nacional, garantindo escalabilidade, resiliência e eficiência energética.

Camada de Aplicação Clínica: responsável pela integração dos resultados analíticos aos fluxos assistenciais do SUS. Inclui sistemas de suporte à decisão clínica, ferramentas de vigilância genômica, plataformas de medicina de precisão e interfaces para profissionais de saúde. Essa camada traduz o conhecimento gerado em ações concretas no contexto assistencial.

A integração dessas camadas permite a construção de um ecossistema nacional capaz de transformar dados biomédicos em conhecimento clínico de alto valor, reduzindo a dependência de plataformas internacionais e promovendo a soberania tecnológica. Além disso, essa arquitetura favorece a reprodutibilidade, a padronização de pipelines e a validação em ambientes clínicos reais, elementos essenciais para a adoção segura e eficaz de

soluções baseadas em IA no SUS.

7.3 Lacunas e Desafios Atuais

Apesar dos avanços científicos e tecnológicos recentes, a bioinformática clínica no Brasil ainda enfrenta lacunas estruturais que limitam sua efetividade e escalabilidade. Entre os principais desafios, destaca-se a fragmentação dos dados, com bases genômicas, registros clínicos e informações socioeconômicas distribuídos em diferentes instituições, frequentemente sem padrões de interoperabilidade, o que dificulta sua integração e uso conjunto. Soma-se a isso a forte dependência de bases de dados e ferramentas internacionais para interpretação de variantes genéticas e suporte à decisão clínica, uma vez que repositórios e pipelines amplamente utilizados são majoritariamente desenvolvidos fora do país. Essa dependência compromete a soberania de dados, limita a adaptação a contextos populacionais específicos e afeta a sustentabilidade dessas soluções no cenário nacional, além de reduzir a capacidade de resposta rápida a emergências sanitárias, como surtos de doenças infecciosas [de Oliveira and Lopes-Cendes 2025].

Do ponto de vista computacional e metodológico, observa-se a ausência de uma infraestrutura nacional coordenada de alto desempenho voltada especificamente para aplicações em bioinformática clínica. Embora existam centros de computação de alto desempenho no país, sua utilização para aplicações clínicas ainda é limitada e pouco integrada ao sistema de saúde. Além disso, há desafios relacionados à reprodutibilidade e à padronização de pipelines, o que compromete a confiabilidade dos resultados e dificulta a validação em diferentes contextos institucionais. No âmbito da IA, a aplicação de modelos avançados em bioinformática clínica ainda ocorre de forma incipiente e pouco sistematizada, com baixa integração aos fluxos clínicos reais e carência de mecanismos robustos de validação, explicabilidade e auditoria em ambientes regulados. Adicionalmente, a dependência de bases de dados predominantemente derivadas de outras populações pode introduzir vieses significativos nos modelos de aprendizado de máquina, reduzindo sua capacidade de generalização e comprometendo a acurácia e a equidade das predições no contexto da população brasileira. A ausência de diretrizes nacionais claras para o uso de IA em bioinformática clínica reforça esse cenário, constituindo uma barreira relevante para sua adoção segura e eficaz em larga escala [Hasanzadeh et al. 2025].

Adicionalmente, existem desafios regulatórios e éticos relevantes, especialmente no que diz respeito à proteção de dados sensíveis, à conformidade com a Lei Geral de Proteção de Dados (LGPD) e à necessidade de transparência e auditabilidade dos sistemas utilizados. A inexistência de frameworks consolidados para certificação de modelos e pipelines em bioinformática clínica dificulta a sua incorporação em ambientes assistenciais e limita a confiança de profissionais de saúde e gestores. Destaca-se, ainda a necessidade de maior articulação entre academia, indústria e governo para o desenvolvimento de soluções sustentáveis e escaláveis. A ausência de um ecossistema integrado que conecte pesquisa, desenvolvimento tecnológico e aplicação clínica compromete o avanço da bioinformática clínica como um elemento estruturante da saúde digital no Brasil. Essas lacunas, em conjunto, evidenciam a necessidade de um esforço coordenado e de longo prazo, caracterizando um grande desafio para a computação aplicada à saúde na próxima década [Khan et al. 2024].

7.4 Formulação do Grande Desafio

Diante das lacunas estruturais identificadas, propõe-se como grande desafio para a próxima década a construção de uma base tecnológica, científica e institucional capaz de sustentar a bioinformática clínica como um componente estratégico do sistema de saúde brasileiro. Esse desafio não se restringe ao desenvolvimento de ferramentas isoladas ou à adoção pontual de tecnologias emergentes, mas envolve a consolidação de uma infraestrutura integrada, orientada por dados, baseada em IA e alinhada às necessidades do SUS e do Complexo Econômico-Industrial da Saúde (CEIS). Nesse contexto, o grande desafio pode ser formulado da seguinte forma: *Como projetar e implementar, até 2036, uma infraestrutura nacional de bioinformática clínica baseada em IA, capaz de integrar dados genômicos, clínicos e socioeconômicos em escala populacional, reduzir a dependência de plataformas internacionais e apoiar, de forma sustentável, o diagnóstico, a decisão terapêutica e o desenvolvimento de terapias avançadas no âmbito do SUS?*

A formulação desse desafio envolve múltiplas dimensões interdependentes, incluindo a integração de dados heterogêneos com padrões de interoperabilidade e segurança, o desenvolvimento de métodos avançados de IA capazes de lidar com dados biológicos complexos e multimodais, e a construção de uma infraestrutura computacional nacional escalável e orientada à soberania de dados. Adicionalmente, requer a definição de mecanismos de validação, certificação e auditoria de modelos e pipelines, bem como a articulação entre academia, indústria e governo para garantir sustentabilidade e impacto. Dessa forma, o desafio transcende aspectos puramente tecnológicos, configurando-se como um problema sistêmico de natureza científica, econômica, regulatória e social, cujo enfrentamento tem potencial de posicionar o Brasil em um novo patamar de autonomia e inovação em saúde, contribuindo para a redução de desigualdades e para o fortalecimento do SUS na próxima década.

7.5 Dimensões do Desafio

Para responder às lacunas identificadas na Seção 1.3 e operacionalizar a arquitetura proposta na Seção 1.2, a construção de uma infraestrutura nacional de bioinformática clínica baseada em IA configura-se como um desafio multifacetado, que exige a articulação de dimensões interdependentes. Nesse contexto, o desafio pode ser estruturado em quatro eixos principais: tecnológico, clínico, industrial e regulatório.

Dimensão Tecnológica: envolve o desenvolvimento de infraestruturas computacionais escaláveis e seguras, capazes de suportar o processamento de grandes volumes de dados genômicos e clínicos. Inclui a consolidação de ambientes de HPC, arquiteturas distribuídas e plataformas federadas de dados. Também abrange a definição de padrões de interoperabilidade, fundamentais para integração de sistemas, e o desenvolvimento de métodos avançados de IA para análise de dados biológicos complexos e multimodais, considerando requisitos de interpretabilidade, robustez e reprodutibilidade. Adicionalmente, destaca-se a necessidade de padronização e versionamento de pipelines, garantindo reprodutibilidade científica.

Dimensão Clínica: refere-se à capacidade de transformar dados biológicos em informação útil para a prática assistencial. Envolve a interpretação de variantes genéticas considerando a diversidade da população brasileira, a integração da bioinformática aos

fluxos do SUS e o desenvolvimento de sistemas de suporte à decisão clínica. Inclui ainda a vigilância genômica contínua, especialmente para doenças infecciosas, e a validação de modelos em ambientes reais, com avaliação de desempenho, segurança e impacto clínico.

Dimensão Industrial: está relacionada ao fortalecimento do Complexo Econômico-Industrial da Saúde (CEIS) e à promoção da inovação tecnológica. Engloba a redução da dependência de plataformas estrangeiras, o estímulo ao desenvolvimento de soluções nacionais, a criação de ambientes favoráveis à inovação e a ampliação da interação entre academia, indústria e governo. Inclui também a formação de recursos humanos qualificados, essencial para sustentar o crescimento do ecossistema de bioinformática clínica.

Dimensão Regulatória e Ética: abrange os desafios associados ao uso seguro e responsável de dados e modelos em saúde. Inclui a conformidade com a Lei Geral de Proteção de Dados (LGPD), a definição de frameworks regulatórios para validação, certificação e auditoria de modelos de IA, e a garantia de transparência e explicabilidade. Também envolve a promoção da equidade no acesso às soluções desenvolvidas, evitando a ampliação de desigualdades no sistema de saúde.

7.6 Problemas Científicos e Computacionais em Aberto

Embora a infraestrutura proposta tenha uma dimensão estratégica, seu núcleo é composto por problemas científicos e computacionais ainda não resolvidos. O primeiro desafio refere-se à integração multimodal de dados genômicos, clínicos, laboratoriais, epidemiológicos e socioeconômicos, os quais apresentam diferentes escalas temporais, níveis de qualidade, formatos, padrões semânticos e graus de completude. Não se trata apenas de interoperabilidade sintática, mas da construção de representações computacionais capazes de preservar relações biológicas, clínicas e contextuais relevantes para a tomada de decisão.

Um segundo problema aberto envolve a generalização de modelos de IA para populações sub-representadas. Modelos treinados majoritariamente em bases internacionais podem apresentar desempenho reduzido quando aplicados à população brasileira, marcada por elevada diversidade genética, regional e socioeconômica. Assim, torna-se necessário desenvolver métodos de aprendizado capazes de incorporar diversidade populacional, estimar incerteza, detectar vieses e garantir desempenho robusto em diferentes subgrupos.

Outro desafio central é o desenvolvimento de estratégias de aprendizado federado e distribuído orientadas à soberania de dados. Nesse cenário, instituições de saúde, laboratórios e centros de pesquisa poderiam colaborar no treinamento e validação de modelos sem transferir dados sensíveis para uma base centralizada. Entretanto, permanecem abertos problemas relacionados à heterogeneidade estatística entre instituições, segurança contra vazamento de informação, rastreabilidade das atualizações dos modelos e balanceamento entre privacidade, desempenho e custo computacional.

Adicionalmente, a incorporação de modelos generativos, Transformers e sistemas de apoio à decisão em bioinformática clínica exige novos métodos de validação, explicabilidade e auditoria. Em aplicações clínicas, não basta que o modelo apresente alta acurácia; é necessário que suas recomendações sejam rastreáveis, interpretáveis, clini-

amente justificáveis e a validadas em fluxos reais de cuidado. Por fim, a operação em escala nacional demanda avanços em computação de alto desempenho, orquestração de cargas, qualidade de serviço, eficiência energética e execução distribuída em ambientes de nuvem, borda e névoa computacional.

7.7 Casos de Uso Prioritários

A infraestrutura nacional proposta pode ser materializada a partir de casos de uso prioritários, capazes de demonstrar sua relevância clínica, computacional e estratégica para o SUS. Um primeiro exemplo envolve a interpretação de variantes raras em pacientes atendidos pela rede pública de saúde. Nesse fluxo, dados de sequenciamento poderiam ser processados por pipelines padronizados, anotados com bases nacionais e internacionais, avaliados por modelos de IA e submetidos a mecanismos de explicabilidade e curadoria clínica. O objetivo seria reduzir o tempo entre a geração do dado genômico e a emissão de um laudo clinicamente útil, ampliando a capacidade de diagnóstico em doenças raras e reduzindo a dependência de plataformas externas de interpretação.

Outro caso de uso relevante está associado à oncologia de precisão. Nesse contexto, dados genômicos tumorais, informações clínicas, histórico terapêutico e evidências extraídas da literatura científica poderiam ser integrados para apoiar a priorização de terapias, a identificação de alternativas terapêuticas, a seleção de ensaios clínicos e a definição de estratégias de acompanhamento. A soberania tecnológica, nesse caso, envolve a capacidade de construir bases nacionais de evidências, desenvolver modelos adaptados ao perfil da população brasileira e disponibilizar pipelines auditáveis para suporte à decisão clínica.

A vigilância genômica de patógenos constitui um terceiro exemplo estratégico. A infraestrutura proposta permitiria integrar sequências genômicas, metadados epidemiológicos, localização geográfica e indicadores assistenciais para detectar, classificar e monitorar variantes virais, bacterianas ou parasitárias de interesse em saúde pública. Essa capacidade seria especialmente importante em situações de emergência sanitária, nas quais a rapidez na análise dos dados e na emissão de alertas pode influenciar diretamente a resposta do sistema de saúde. Esse caso evidencia a necessidade de baixa latência, escalabilidade, interoperabilidade e mecanismos de notificação voltados a gestores, laboratórios e serviços de saúde.

Em conjunto, esses casos de uso demonstram que a infraestrutura proposta não se limita à criação de uma base tecnológica abstrata. Ela se traduz em aplicações concretas, com potencial de apoiar o diagnóstico, orientar decisões terapêuticas, fortalecer a vigilância em saúde e ampliar a capacidade do SUS de incorporar a bioinformática clínica e a IA em escala populacional.

7.8 Métricas de Avaliação do Progresso

A avaliação do progresso no enfrentamento do grande desafio proposto requer métricas claras, mensuráveis e multidimensionais, capazes de capturar avanços tecnológicos, impacto clínico, fortalecimento industrial e evolução da capacidade nacional ao longo do tempo. Para esse fim, propõe-se um conjunto de indicadores organizados em quatro categorias principais: técnica, clínica, estratégica e de formação de recursos humanos,

conforme apresentado na Tabela 1.1.

Tabela 1.1. Métricas de avaliação do progresso do grande desafio.

Categoria	Métrica	Descrição
Técnica	Tempo médio de interpretação de variantes	Tempo necessário para classificar variantes genéticas desde a obtenção dos dados até o laudo clínico.
Técnica	Volume de dados processados	Quantidade de genomas, exomas ou dados biomoleculares processados anualmente.
Técnica	Taxa de reprodutibilidade	Proporção de análises consistentes entre diferentes ambientes e instituições.
Técnica	Pipelines padronizados e certificados	Número de fluxos de análise validados segundo critérios nacionais.
Técnica	Índice de interoperabilidade	Grau de integração entre sistemas e bases de dados.
Clínica	Tempo para decisão terapêutica	Redução no tempo para definição de condutas clínicas baseadas em dados genômicos.
Clínica	Acurácia clínica	Desempenho dos modelos na identificação de variantes ou alvos terapêuticos.
Clínica	Cobertura populacional	Proporção da população atendida pelo SUS coberta pela infraestrutura.
Clínica	Redução de diagnósticos tardios	Diminuição de diagnósticos em estágios avançados.
Clínica	Resposta a eventos epidemiológicos	Tempo de identificação e análise de variantes em emergências sanitárias.
Estratégica	Dependência de plataformas internacionais	Proporção de uso de soluções nacionais em relação a estrangeiras.
Estratégica	Instituições integradas	Número de centros conectados à infraestrutura nacional.
Estratégica	Terapias avançadas suportadas	Número de iniciativas terapêuticas apoiadas pela infraestrutura.
Estratégica	Investimento em P&D	Volume de recursos destinados à área.
Estratégica	Produção científica e tecnológica	Número de publicações, patentes e transferências de tecnologia.
Formação	Profissionais capacitados	Número de especialistas formados em bioinformática e IA aplicada à saúde.
Formação	Programas de formação	Quantidade de cursos e programas estabelecidos.
Formação	Inserção no mercado	Taxa de absorção de profissionais no setor.

7.9 Roadmap 2026–2036

O enfrentamento do grande desafio proposto requer uma abordagem estruturada e progressiva, organizada em fases que permitam a evolução coordenada das capacidades tecnológicas, institucionais e regulatórias ao longo da próxima década. Para esse fim, propõe-se um roadmap dividido em três fases principais: estruturação, expansão e consolidação, que articulam ações estratégicas voltadas à construção, integração e maturidade da infraestrutura nacional de bioinformática clínica baseada em Inteligência Artificial, conforme apresentado na Tabela 1.2.

Tabela 1.2. Roadmap proposto para 2026–2036.

Fase	Período	Principais Ações
Estruturação e Convergência	2026–2028	Mapeamento das capacidades nacionais em bioinformática, genômica, IA e infraestrutura computacional; definição de padrões de interoperabilidade; implantação de ambientes piloto de dados federados; desenvolvimento de benchmarks nacionais; estabelecimento de diretrizes iniciais de governança de dados, LGPD e uso ético de IA.
Expansão e Integração	2028–2031	Ampliação da rede nacional de dados e computação com participação multi-institucional; integração da bioinformática clínica aos fluxos do SUS; consolidação de plataformas HPC para aplicações biomédicas; desenvolvimento e validação de modelos de IA em ambientes clínicos reais; implementação de protocolos de certificação e auditoria.
Consolidação e Autossuficiência Parcial	2031–2036	Operação plena da infraestrutura nacional com cobertura ampliada; redução da dependência de plataformas internacionais; suporte ao desenvolvimento de terapias avançadas nacionais; estabelecimento de mecanismos contínuos de avaliação e atualização; posicionamento do Brasil como referência regional em bioinformática clínica e saúde digital.

7.10 Conexão com os Grandes Desafios da Computação no Brasil

O grande desafio proposto está fortemente alinhado com os Grandes Desafios da Computação no Brasil, conforme discutidos no IV Seminário da Sociedade Brasileira de Computação (SBC), ao integrar múltiplas áreas estratégicas da computação em um contexto aplicado à saúde. Em particular, destaca-se sua conexão com **IA e Ciência de Dados**, ao demandar o desenvolvimento de modelos capazes de processar dados biológicos complexos e multimodais, bem como com **Sistemas Distribuídos e Computação em Larga Escala**, dada a necessidade de infraestruturas federadas e processamento de grandes volumes de dados biomédicos. O desafio também envolve avanços em **Cibersegurança e Privacidade de Dados**, especialmente no tratamento de informações sensíveis de saúde em conformidade com a LGPD, além de aspectos de **Computação Sustentável**, considerando o custo computacional e energético associado ao uso intensivo de IA.

Além da dimensão tecnológica, o desafio dialoga com temas centrais relacionados à **Ética, Transparência e Governança em Sistemas Computacionais**, ao exigir modelos explicáveis, auditáveis e confiáveis em contextos clínicos críticos. Destaca-se ainda sua contribuição para a **redução de desigualdades sociais**, ao propor a ampliação do acesso a tecnologias avançadas de saúde em diferentes regiões do país, fortalecendo o SUS. Dessa forma, o desafio não apenas articula áreas prioritárias da computação, mas também reforça seu papel estratégico na transformação do sistema de saúde brasileiro, posicionando a computação como elemento central para o desenvolvimento científico, tecnológico e social na próxima década.

7.11 Impacto Social e Econômico

A construção de uma infraestrutura nacional de bioinformática clínica baseada em IA tem potencial para gerar impactos estruturais na sociedade brasileira, especialmente no que se refere à equidade em saúde e ao fortalecimento do SUS. A ampliação do acesso a tec-

nologias avançadas de diagnóstico e suporte à decisão clínica, atualmente concentradas em centros de referência, pode contribuir para a democratização da medicina de precisão em diferentes regiões do país, incluindo áreas remotas. Isso possibilita a redução de desigualdades regionais, a diminuição de diagnósticos tardios e a adoção de intervenções mais precoces e personalizadas, resultando em melhor qualidade de vida para os pacientes e maior eficiência no uso de recursos públicos. Além disso, a capacidade de monitoramento genômico em tempo quase real fortalece a resposta a emergências sanitárias, contribuindo para um sistema de saúde mais resiliente e orientado por dados.

No âmbito econômico, essa infraestrutura pode impulsionar o CEIS, promovendo a inovação e reduzindo a dependência de tecnologias estrangeiras. O desenvolvimento de soluções nacionais baseadas em dados biomédicos cria oportunidades para startups, empresas de base tecnológica e parcerias público-privadas, aumentando a competitividade do país no cenário internacional. Adicionalmente, a formação de recursos humanos altamente qualificados fortalece a economia do conhecimento e sustenta o crescimento do ecossistema de bioinformática clínica. Destaca-se ainda o impacto na bioeconomia, com potencial de acelerar o desenvolvimento de terapias avançadas, e na governança em saúde, ao viabilizar decisões mais informadas por meio da integração de dados e análises avançadas. Dessa forma, o grande desafio proposto posiciona a bioinformática clínica como um elemento central para o desenvolvimento sustentável, científico e tecnológico da saúde no Brasil na próxima década.

Sobre os Autores

Marcelo Augusto Costa Fernandes Marcelo Augusto Costa Fernandes é Professor Associado da Universidade Federal do Rio Grande do Norte (UFRN), atuando no Departamento de Engenharia da Computação e Automação, sendo também docente permanente dos Programas de Pós-Graduação em Engenharia Elétrica e Computação (PPgEEC) e em Bioinformática (PPgBioInfo). É Diretor do Núcleo de Inteligência Artificial e Ciência de Dados (nIACD) do Instituto Metrôpole Digital e líder do InovAI Lab, ambos vinculados ao IMD/UFRN. É Bolsista de Produtividade em Pesquisa do CNPq. Entre 2015 e 2016, atuou como pesquisador visitante no King's College London, em Londres, Reino Unido. No período de 2019 a 2021, foi pesquisador visitante na John A. Paulson School of Engineering and Applied Sciences, da Universidade de Harvard, em Cambridge, Estados Unidos. Em 2022, realizou estágio como pesquisador visitante no Centro de Pesquisa em Mente, Cérebro e Comportamento, da Universidade de Granada, em Granada, Espanha.

Sua atuação científica concentra-se no desenvolvimento de soluções baseadas em Inteligência Artificial para problemas complexos, com aplicações em saúde, bioinformática, sistemas inteligentes e computação de alto desempenho. Possui ampla experiência na coordenação de projetos de pesquisa e desenvolvimento em larga escala, envolvendo a integração de dados biomédicos, aprendizado de máquina e infraestrutura computacional, com ênfase em genômica, classificação de sequências biológicas, interações fármaco-alvo e saúde digital. Atualmente, é coordenador do grupo Inteligência Artificial Aplicada à Saúde Materno-Infantil (IASMIN) e da rede SOFIA Soluções em Observação, Formação, Inovação e Aplicação em Saúde Materno-Infantil.

ORCID: <https://orcid.org/0000-0001-7536-2506>

Lattes: <http://lattes.cnpq.br/3475337353676349>

Referências

- [Alsaedi et al. 2025] Alsaedi, S., Ogasawara, M., Alarawi, M., Gao, X., and Gojobori, T. (2025). Ai-powered precision medicine: utilizing genetic risk factor optimization to revolutionize healthcare. *NAR Genomics and Bioinformatics*, 7(2):lqaf038.
- [Bycroft et al. 2018] Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O’onnell, J., et al. (2018). The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209.
- [de Oliveira and Lopes-Cendes 2025] de Oliveira, T. C. and Lopes-Cendes, I. (2025). Population molecular genetics in brazil: From genomic databases and research to the implementation of precision medicine. *Journal of Community Genetics*, 16(4):409–420.
- [European Commission 2022] European Commission (2022). Proposal for a regulation on the european health data space. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en.
- [Gao et al. 2023] Gao, Y., Sharma, T., and Cui, Y. (2023). Addressing the challenge of biomedical data inequality: An artificial intelligence perspective. *Annual Review of Biomedical Data Science*, 6(Volume 6, 2023):153–171.
- [Ghebrehiwet et al. 2024] Ghebrehiwet, I., Zaki, N., Damseh, R., and Mohamad, M. S. (2024). Revolutionizing personalized medicine with generative ai: a systematic review. *Artificial Intelligence Review*, 57(5):128.
- [Hasanzadeh et al. 2025] Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., and White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *NPJ Digital Medicine*, 8(1):154.
- [Khan et al. 2024] Khan, A. M., Işık, Esra Büşra, and Tan, T. W. (2024). A global initiative on addressing bioinformatics’ grand challenges. *Briefings in Bioinformatics*, 25(4):bbae278.
- [Nunes et al. 2025] Nunes, K., e Silva, M. A. C., Rodrigues, M. R., Lemes, R. B., Pezo-Valderrama, P., Kimura, L., de Sena, L. S., Krieger, J. E., Varela, M. C., de Azevedo, L. O., Camargo, L. M. A., Ferreira, R. G. M., Krieger, H., Bortolini, M. C., Mill, J. G., Sacuena, P., ao F. Guerreiro, J., de Souza, C. M. B., Veronese, F. V., Vianna, F. S. L., Comas, D., Pereira, A. C., Pereira, L. V., and Hünemeier, T. (2025). Admixture’s impact on brazilian population evolution and health. *Science*, 388(6748):eadl3564.
- [of Us Research Program Investigators 2019] of Us Research Program Investigators, A. (2019). The ”all of us” research program. *New England Journal of Medicine*, 381(7):668–676.