

Investigação de Padrões Epidemiológicos em Oncologia Pediátrica a partir de Integração e Visualização de Dados

Felipe de Assis Ribeiro^{1,2}, Marialva Sinigaglia^{2,3}, Alessandro Vargas de Ávila²,
Isabel Cristina Siqueira da Silva¹

¹Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)
Porto Alegre, RS – Brasil

²Bioinformática – Instituto do Câncer Infantil
Porto Alegre, RS – Brasil

³Instituto Nacional de Ciência e Tecnologia em Biologia do Câncer Infantil
e Oncologia Pediátrica - INCT BioOncoPed
Porto Alegre, RS – Brasil

{felipeasr, isabel.siqueira}@ufcspa.edu.br,
avargasdeavila@yahoo.com.br, msinigaglia@ici.org

Abstract. *This article describes a study that aimed to investigate epidemiological patterns in the DATASUS and RHC databases, based on the extraction and processing of data related to childhood cancer. From the collected, processed and integrated data, a web system was developed consisting of three dashboards, which present different types of interactive visual representations of pediatric oncology data. The system is available for public access, since the information collected from the aforementioned databases is already anonymized, and aims to assist, mainly, in strategic decision-making related to the treatment and prevention of childhood cancer.*

Resumo. *Este artigo descreve um estudo com o objetivo de investigar padrões epidemiológicos, nas bases de dados do DATASUS e RHC, a partir da extração e processamento de dados relacionados ao câncer infantojuvenil. A partir dos dados coletados, tratados e integrados, foi desenvolvido um sistema web composto por três dashboards, os quais apresentam diferentes tipos de representações visuais interativas de dados de oncologia pediátrica. O sistema está disponibilizado para acesso público, uma vez que as informações coletadas das bases de dados citadas já estão anonimizadas, e visa auxiliar, principalmente, a tomada de decisão estratégica relacionada ao tratamento e prevenção do câncer infantil.*

1. Introdução

No Brasil, o câncer infantil é a primeira causa de morte por doença em crianças de 1 a 19 anos, superando as doenças infecciosas [INCA 2023]. O Instituto Nacional do Câncer (INCA) estimou 704 mil novos casos de câncer pediátrico no Brasil para cada ano do triênio 2023-2025, com estimativa de sobrevida de 64%. Esta estimativa apontou, ainda, uma disparidade significativa entre as regiões do país, com destaque para as regiões Sul e Sudeste, que concentram cerca de 70% da incidência.

Atualmente, existem diferentes bases de dados essenciais para a compreensão da incidência de câncer na população brasileira em geral, tanto públicas quanto privadas. Dentre as públicas, destaca-se o painel oncológico do Departamento de Informática do Sistema Único de Saúde (DATASUS), que reúne dados de diferentes sistemas, como, por exemplo, o sistema de informação ambulatorial (SIA), o sistema de informação hospitalar (SIH), o sistema de informação do câncer (SISCAN), o sistema de informação de morbimortalidade (SIM) e o cadastro nacional de estabelecimentos de saúde (CNES)[Atty et al. 2020]. Além do DATASUS, os registros hospitalares de câncer (RHC) e os registros de câncer de base populacional (RCBP) são outras importantes fontes de dados para estudos epidemiológicos da doença [INCA 2023].

Os tumores infantis diferem dos tumores de adultos quanto ao aspecto morfológico, comportamento clínico e localização primária. Assim, é necessário extrair, das bases de dados oncológicos, apenas as informações relacionadas à incidência de câncer em crianças e adolescentes a fim de pesquisá-las e investigar padrões epidemiológicos. No entanto, dados sobre oncologia pediátrica apresentam obstáculos consideráveis na construção de uma epidemiologia abrangente em função da natureza heterogênea dos registros de dados e às complexidades que envolvem sua coleta [Linnet et al. 1999, Bernardi et al. 2022, Costantini et al. 2023].

O registro dos dados, por exemplo, nem sempre é realizado por profissionais que compreendem as especificidades destes e, por vezes, sintomas de câncer infantil podem ser confundidos com outras doenças comuns na infância quando uma investigação mais aprofundada não é realizada. A ausência de feedbacks sobre o uso desses dados na construção de políticas públicas também é um fator que pode comprometer a percepção da importância da coleta e do registro de dados oncológicos, principalmente os pediátricos. Como consequência, os dados podem apresentar inconsistências, as quais, por sua vez, podem levar a diagnósticos incorretos e tomadas de decisões equivocadas, comprometendo a integridade dos dados e tomada de decisão [Khan and Qin 2018, Batko and Ślęzak 2022].

Neste contexto, a ciência de dados aplicada à saúde engloba a coleta, o armazenamento e a análise de grandes volumes de dados, visando auxiliar na organização, no gerenciamento e na recuperação da informação, bem como na descoberta de conhecimento [Jensen et al. 2012, Obermeyer and Emanuel 2016, Gazzarata et al. 2024]. A relevância dessa tendência se torna ainda mais evidente à medida que potencializa discussões relacionadas ao tratamento de dados de saúde e à tomada de decisão embasada nestes [Rodriguez-Galindo et al. 2013, Mohammadzadeh et al. 2017, Khan and Qin 2018, Osterman et al. 2020].

Para que a análise de dados seja efetiva e possa conduzir à extração de informações abrangente, técnicas de visualização de dados podem ser aplicadas a partir do desenvolvimento de representações visuais interativas que auxiliam no entendimento dos diferentes relacionamentos entre estes. Na área da saúde, a visualização de dados colabora para encurtar o tempo decorrido entre a análise dos dados e a implementação de medidas práticas. A interação com as representações visuais em um dashboard auxilia profissionais da saúde na caracterização, na segmentação e na correlação de dados, apoiando a decisão e reduzindo o esforço cognitivo para analisar as informações e descobrir conhecimento [Chan et al. 2018, Park et al. 2022].

Considerando o exposto, este artigo apresenta um estudo que teve, por objetivo, o mapeamento dos dados de oncologia pediátrica presentes nos sistemas do DATASUS e no RHC. A partir deste mapeamento, foi desenvolvido um sistema web composto por três dashboards, os quais apresentam diferentes tipos de representações visuais interativas dos dados mapeados, extraídos e integrados. O sistema está disponibilizado para acesso público, uma vez que as informações coletadas das bases de dados citadas já estão anonimizadas, e visa auxiliar, principalmente, a tomada de decisão estratégica relacionada ao tratamento e prevenção do câncer infantil.

Os dashboards apresentados foram definidos junto a uma equipe multidisciplinar de profissionais que atuam em pesquisas relacionadas ao câncer infantojuvenil. Esta equipe fez parte deste estudo desde sua proposição, contribuindo de forma significativa para a identificação de necessidades específicas na extração de dados das bases oncológicas.

O texto está organizado como segue. Além desta seção introdutória, a seção 2 apresenta os materiais e métodos empregados neste estudo. A seção 3 apresenta os resultados enquanto a seção 4 realiza a discussão dos resultados e traz as considerações finais com a indicação de trabalhos futuros.

2. Materiais e Métodos

Este estudo começou com definição de uma equipe multidisciplinar de profissionais com interesse em investigar padrões epidemiológicos sobre o câncer infantojuvenil. Esta equipe é formada por uma médica hematologista pediátrica, um médico oncologista pediátrico e superintendente de uma instituição focada na prevenção e no tratamento do câncer infantojuvenil, um médico pediatra, que atua em uma secretaria municipal de saúde, uma bióloga e líder de projetos na área da bioinformática, uma cientista da computação, que trabalha com visualização de dados, e dois informata biomédicos.

A equipe realizou, então, a análise das principais bases de dados públicas e disponíveis no Brasil que permitissem a extração de dados relacionados à oncologia pediátrica. Estes dados precisavam ser representativos e relevantes de modo a contribuir para a tomada de decisão por parte de profissionais e gestores da área da saúde em relação à prevenção e ao tratamento do câncer infantojuvenil. Em paralelo, tinha-se a intenção de disponibilizar estas informações para facilitar o entendimento do quadro de oncologia pediátrica no país.

Na sequência, os dados foram analisados, processados e padronizados de modo a permitir sua análise e definição das principais informações que poderiam ser extraídas a partir destes. A partir de então, definiu-se as tecnologias, que seriam empregadas para a construção do sistema web proposto e seus dashboards, visando permitir a análise dos dados oncológicos de maneira mais intuitiva e integrada.

Estas etapas estão detalhadas a seguir.

2.1. Escolha das Bases de Dados

O primeiro passo dessa etapa envolveu o levantamento das bases de dados existentes, com foco naquelas que contêm informações oncológicas relevantes. Após o levantamento, foi realizada uma triagem das bases de dados para selecionar aquelas que ofereciam dados

específicos sobre o câncer infantil, com a intenção de fornecer uma visão clara e precisa da situação da oncologia pediátrica no Brasil. A escolha dessas bases considerou, além da qualidade e especificidade dos dados, a abrangência e confiabilidade das fontes.

As bases de dados escolhidas para este estudo foram as seguintes:

- Painel oncológico do DATASUS: Esta base reúne dados provenientes dos seguintes sistemas:
 - SIA: Contém registros detalhados sobre atendimentos ambulatoriais relacionados à oncologia, fornecendo dados sobre consultas e tratamentos realizados fora do ambiente hospitalar;
 - SIH: Apresenta informações sobre internações hospitalares, incluindo dados sobre procedimentos oncológicos realizados durante a internação dos pacientes;
 - SISCAN: Reúne dados sobre o rastreamento e o diagnóstico precoce do câncer, sendo uma fonte essencial para analisar a detecção precoce da doença.
- SIM do DATASUS: Utilizado para analisar as taxas de óbito e a sobrevivência de pacientes pediátricos com câncer, fornecendo dados fundamentais para o entendimento do impacto da doença.
- CNES do DATASUS: Contém informações sobre hospitais e centros de saúde especializados no tratamento oncológico, permitindo mapear os locais de tratamento disponíveis.
- RHC: Trata-se de um banco de dados altamente especializado, mantido pelo Instituto Nacional de Câncer (INCA), que oferece registros detalhados sobre os pacientes atendidos em unidades de oncologia hospitalar. A análise dessa base é crucial para a avaliação da qualidade do atendimento e da evolução do tratamento oncológico.

Tais bases foram selecionadas, dentre outras, em função da completude dos dados disponibilizados.

2.2. Processamento e Padronização dos Dados

Após a coleta automatizada dos dados das bases selecionadas, estes passaram por um processo de pré-processamento para garantir sua integridade e qualidade. Esse processo foi essencial para corrigir inconsistências, remover duplicidades e preparar os dados para análises futuras.

A verificação da integridade dos dados foi realizada por meio de diversas etapas, incluindo:

- Detecção de valores ausentes ou inconsistentes, garantindo que os registros possuíssem informações completas sempre que possível. Nos casos em que a ausência de dados comprometia a análise, optou-se pela remoção desses registros;
- Identificação e eliminação de registros duplicados, evitando que informações fossem contabilizadas mais de uma vez, o que poderia distorcer os resultados;
- Análise detalhada das bases de dados, que se mostrou um desafio, pois estas não possuem um identificador único comum entre elas. Dessa forma, não foi possível realizar uma integração automática dos diferentes conjuntos de dados. Como alternativa, cada base foi analisada separadamente, garantindo que os padrões internos

fossem respeitados. Essa análise individual permitiu a correção de erros internos e a padronização de informações dentro de cada conjunto de dados.

Como os dados provinham de fontes distintas, foi necessário um processo de normalização e padronização para melhorar a qualidade das análises, mesmo que as bases não pudessem ser integradas. Esse processo incluiu:

- Padronização de nomenclaturas para evitar variações nos nomes de categorias, diagnósticos e outras informações relevantes.
- Organização e reestruturação dos códigos de identificação de doenças conforme a Classificação Internacional de Doenças (CID-10) e Classificação Internacional de Doenças para Oncologia (CID-O), garantindo que os registros dentro de cada base seguissem um padrão.
- Ajuste de formatos de datas, convertendo-as para um padrão unificado, o que facilitou a análise temporal dentro de cada conjunto de dados.
- Tratamento de caracteres especiais e remoção de inconsistências textuais, tornando os registros mais homogêneos e evitando problemas durante a análise.

Em relação às tecnologias empregadas, para o pré-processamento dos dados, foi utilizada a biblioteca Ydata-profiling. Esta permite gerar relatórios detalhados sobre a qualidade dos dados, identificando rapidamente padrões, valores ausentes e inconsistências.

Dado o volume e a complexidade das bases, a linguagem de programação Python foi utilizada para automatizar o processamento dos dados. Esta automatização permite a atualização contínua dos dashboards a partir do monitoramento e da extração automática de dados das bases de dados.

2.3. Visualização de Dados

Após a definição das bases de dados, que seriam utilizadas para extrair informações sobre a oncologia pediátrica, e da análise destas, passou-se ao planejamento dos dashboards, das representações visuais que iriam compor os mesmos e das formas de interação com estas junto à equipe deste projeto.

Foram definidos três principais dashboards:

- Um dashboard que apresenta os dados do painel de oncologia do DATASUS (SIA, SIH e SISCAN) integrados ao CNES do DATASUS;
- Um dashboard que apresenta os dados RHC integrados ao CNES do DATASUS;
- Um dashboard que apresenta os dados do SIM do DATASUS integrados ao CNES do DATASUS.

Os dashboards foram distribuídos desta forma em relação à atualização dos dados das bases relacionadas:

- O dashboard com os dados do painel oncológico do DATASUS abrange o período de 2013 a 2024. Esses dados são atualizados a cada três meses, o que garante a relevância e atualidade das informações apresentadas.
- O dashboard baseado nos dados do RHC apresenta informações do período de 2013 a 2021. Esses dados, por serem mais antigos, são considerados mais consolidados, mas a sua atualização ocorre a cada dois anos, o que implica em um desfasamento temporal.

- O painel com dados do SIM exibe dados do período de 2013 a 2023. No entanto, há um ano de defasagem nas atualizações dos dados do SIM, pois as informações demográficas e sobre mortalidade são mais difíceis de serem obtidas e consolidadas rapidamente.

De acordo com a natureza e características dos dados selecionados para cada dashboard, elencou-se técnicas de visualização que fossem mais adequadas para auxiliar tanto a tomada de decisão quanto a geração de insights com base nestes. Assim, as representações visuais selecionadas foram:

- Gráficos do tipo rosca (*donut*): Indicados para mostrar a relação entre porcentagens ou proporções de dados categóricos, pois é um tipo de visualização que auxilia a comparar categorias e como elas se relacionam com o todo;
- Gráficos de barra: Indicados para representar a comparação de valores entre diferentes categorias ou para exibir distribuições;
- Gráficos *treemap* e *sunburst*: Indicados para representar dados hierárquicos, principalmente quando é preciso representar a dimensionalidade dos dados;
- Gráficos coropléticos: Indicados para representar dados espaciais e quantitativos em um mapa.

Os dashboards e as representações visuais foram criadas com as bibliotecas como Plotly, Dash, Streamlit e Seaborn.

A próxima seção apresenta os resultados deste estudo.

3. Resultados

Após a análise das características das bases de dados selecionadas para este estudo e das definições dos dashboards que iriam apresentá-las, deu-se início à implementação do sistema web para visualização dos dados oncológicos extraídos. No entanto, visando a automatização e a eficiência na extração de dados, foi necessário desenvolver um módulo específico para coletar, padronizar e integrar os dados extraídos das bases de dados DATASUS e RHC. A Figura 1 apresenta o pipeline que resume o fluxo de entrada e saída de dados do sistema desenvolvido.

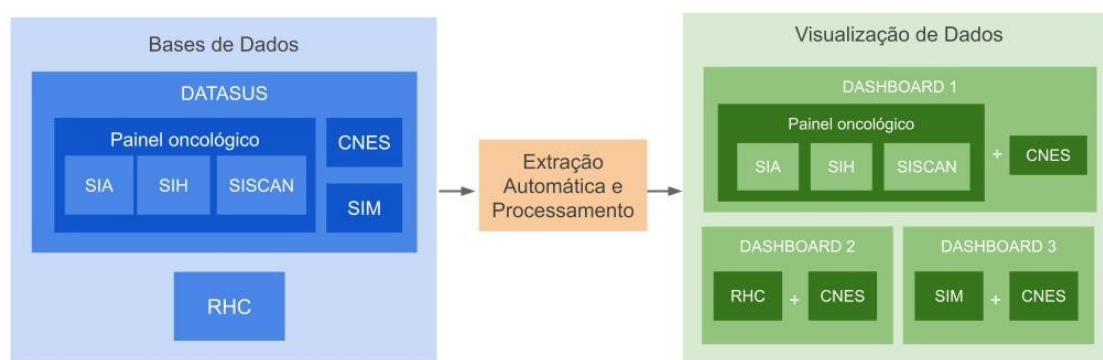


Figura 1. Pipeline da extração, processamento, integração e visualização dos dados de oncologia pediátrica

O módulo desenvolvido monitora e realiza o download automatizado das bases através de seus servidores FTP sempre que novos dados são disponibilizados ou quando

dados existentes são alterados. Garante-se, assim, que os dados sejam atualizados de forma contínua para análise. Essa automação elimina a necessidade de intervenção manual, uma vez que também realiza a descompactação dos arquivos e a conversão destes para outros formatos.

Com a extração, processamento e integração dos dados de forma automatizada, pôde-se estudar a melhor forma de apresentá-los nos dashboards, visando fornecer, aos profissionais de saúde, gestores públicos e pesquisadores, uma visualização clara e detalhada dos dados mais relevantes sobre o tratamento oncológico infantil no Brasil. As representações visuais presentes nos dashboards procuram auxiliar na análise das seguintes informações:

- A quantidade total de diagnósticos registrados em oncologia pediátrica, permitindo a análise de tendências e padrões ao longo do tempo;
- O número de pacientes que iniciaram tratamento oncológico, fornecendo uma visão mais precisa sobre o início do tratamento e suas possíveis lacunas;
- O tempo médio entre o diagnóstico e o início do tratamento, o que é fundamental para avaliar a agilidade no atendimento e o impacto do tempo no prognóstico dos pacientes;
- O diagnóstico mais frequente em pacientes pediátricos, com destaque para os tipos de câncer mais comuns em crianças e adolescentes;
- Os tipos mais comuns de tratamentos iniciais administrados aos pacientes, permitindo uma análise sobre as opções terapêuticas mais frequentemente adotadas;
- Os locais de saúde habilitados para tratamento de oncologia pediátrica, fornecendo uma visão sobre a distribuição geográfica e a capacidade de atendimento;
- Dados sobre mortalidade infantil relacionada à oncologia pediátrica, o que permite avaliar o impacto do câncer em crianças e adolescentes e as possíveis lacunas no tratamento.

As Figuras 2, 3 e 4 apresentam uma parte de cada um dos dashboards desenvolvidos.

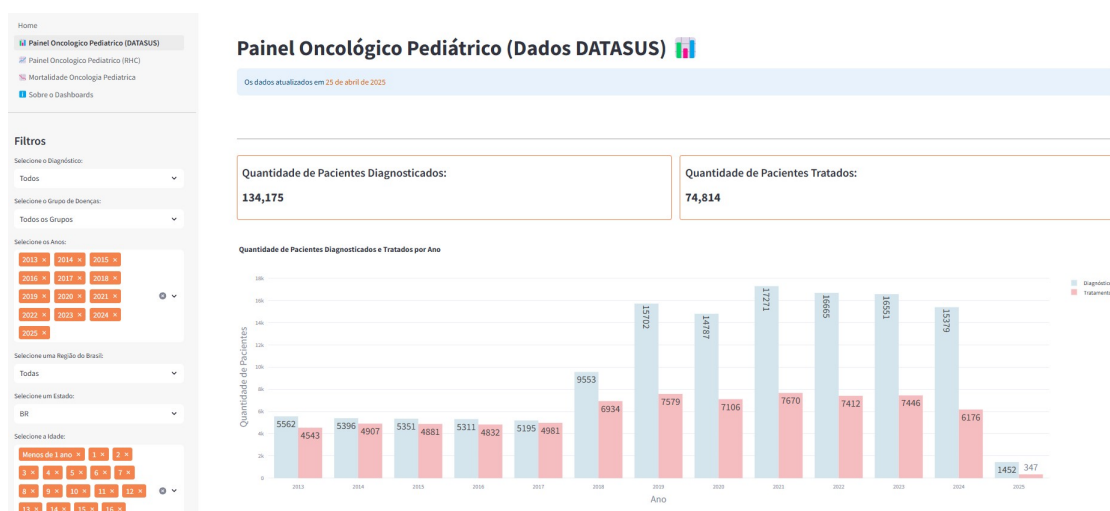


Figura 2. Parte inicial do dashboard exibindo dados do painel de oncologia do DATASUS (SIA, SIH e SISCAN) integrados ao CNES do DATASUS

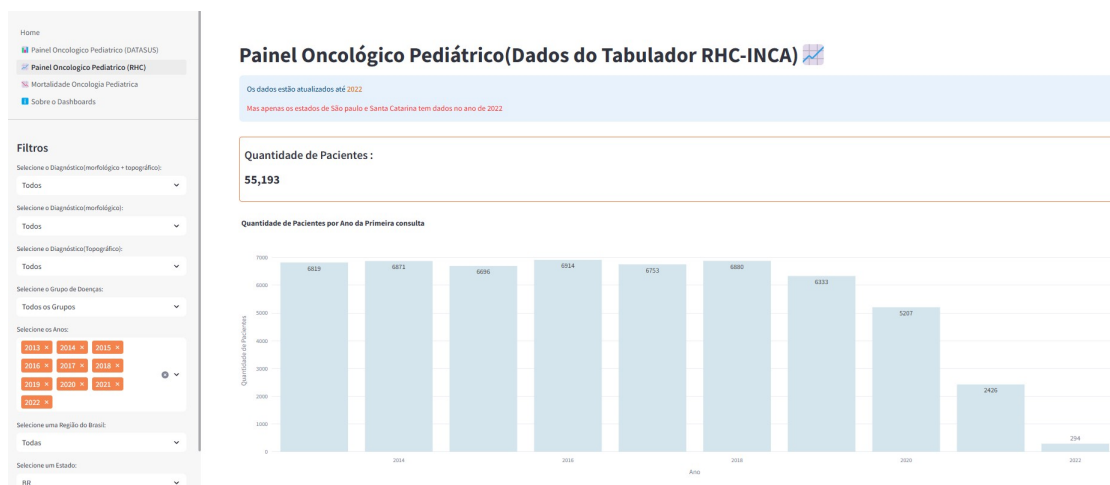


Figura 3. Parte inicial do dashboard exibindo dados do RHC integrados ao CNES do DATASUS

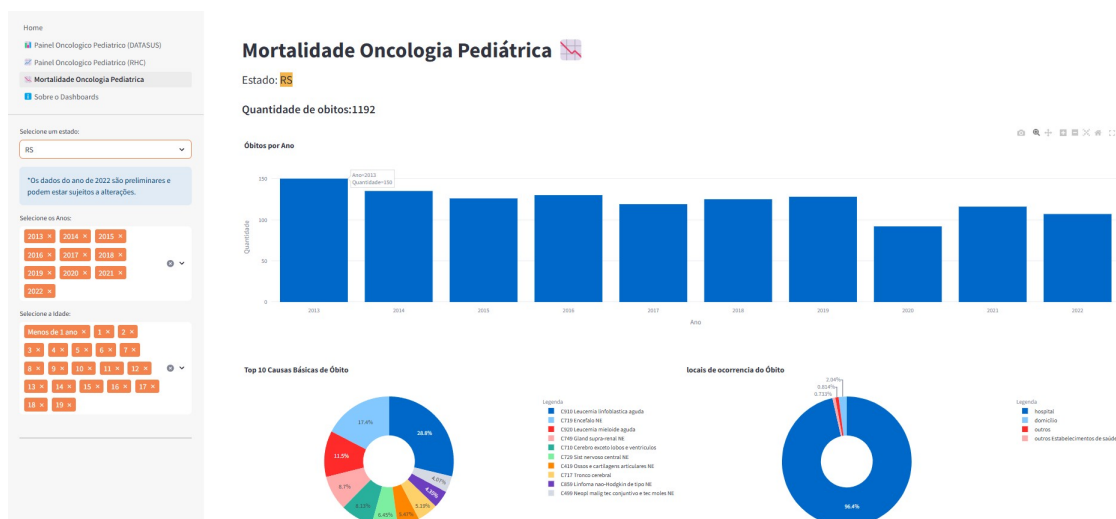


Figura 4. Parte inicial do dashboard exibindo dados do SIM do DATASUS integrados ao CNES do DATASUS

As Figuras 5 e 6 apresentam exemplos de outros tipos de visualização que também estão presentes nos dashboards (representações *treemap* e *sunburst*). Todas estas permitem interação com o usuário, destacando as informações dos gráficos de acordo com a necessidade do usuário.

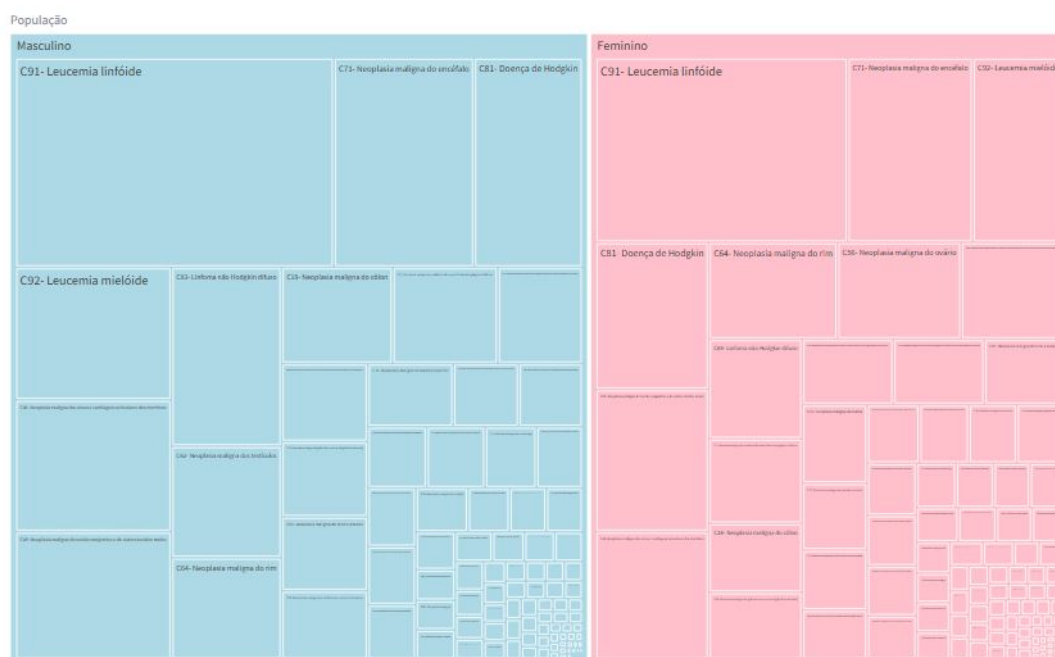


Figura 5. Gráfico Treemap

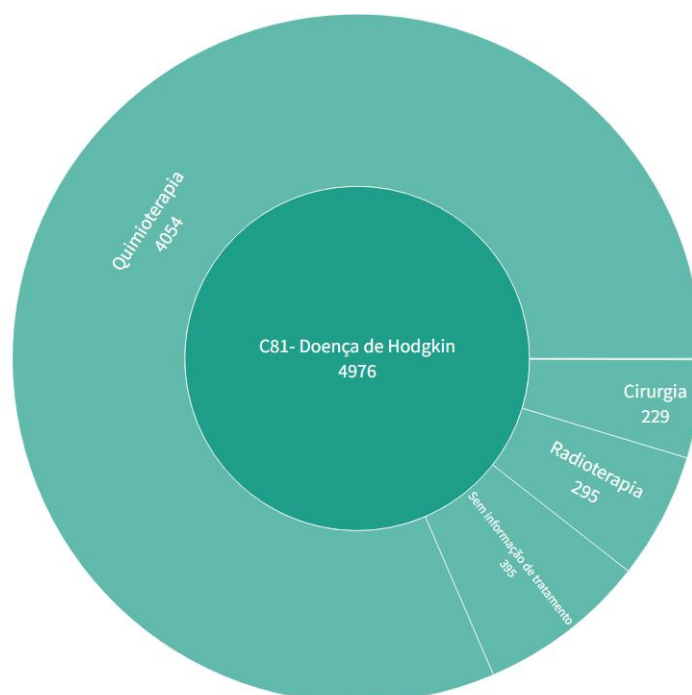


Figura 6. Gráfico Sunburst

Além das representações visuais em forma de diferentes gráficos interativos, foi implementado um sistema de filtros, permitindo que os usuários explorem os dados de forma personalizada. Os filtros possibilitam ajustes conforme variáveis como idade, tipo de câncer, estado e hospital responsável pelo atendimento oncopediátrico, garantindo que diferentes perfis de usuários possam extrair informações relevantes para suas necessidades

específicas. Os dashboards do Painel OncoPed podem ser acessados por meio do link: <https://dashoncologico.streamlit.app>.

A próxima seção apresenta discussões, sobre esses resultados, e as considerações finais.

4. Discussão e Considerações Finais

Neste estudo, as bases de dados foram, especificamente, selecionados pela equipe que colaborou com este estudo visando proporcionar uma visão abrangente e detalhada da oncologia pediátrica no Brasil. Algumas bases de dados públicas não entraram neste estudo em função dos dados estarem incompletos ou inconsistentes. Desta forma, garante-se maior qualidade dos dados apresentados visando, assim, auxiliar profissionais de saúde, gestores e pesquisadores na análise de tendências e investigação de padrões epidemiológicos de dados oncológicos pediátricos com maior eficiência.

Para a exibição dos dados, extraídos e integrados, nos dashboards, a escolha das técnicas de visualização de dados e das formas de interação com estas, através de filtros, foi cuidadosamente realizada e testada pela equipe. Buscou-se, assim, fornecer representações visuais acessíveis e efetivas para auxiliar a tomada de decisão, desde a formulação de novas políticas públicas até a resolução de problemas já existentes, como a falta de acesso ao tratamento oncológico em determinadas regiões.

A equipe que participou deste estudo, após diferentes testes com o sistema, constatou que o uso de tecnologias de integração e análise de dados de saúde não apenas ajudou a superar as limitações das bases de dados fragmentadas, mas também abriu a possibilidade de criar soluções mais eficientes e adaptáveis. Os profissionais destacaram, ainda, que a construção de softwares, que sejam capazes de integrar dados de diferentes fontes e garantir a qualidade e a confiabilidade das informações, é essencial para a formulação de políticas públicas eficazes e para a implementação de práticas clínicas mais adequadas às necessidades da população.

Nota-se que o avanço da ciência de dados na saúde, combinado com uma infraestrutura interoperável, tem sido um importante fator para a transformação da forma como a saúde pública é gerida. A integração entre bases de dados tem permitido uma comunicação mais eficiente entre diferentes sistemas de informação hospitalares, possibilitando decisões baseadas em dados reais e, consequentemente, melhorando a eficácia dos tratamentos, a gestão de recursos e, em última instância, a qualidade de vida dos pacientes. Quando esta integração é apresentada através do emprego de técnicas de visualização de dados, possibilita-se que profissionais da saúde analisem padrões, tendências e anomalias em grandes volumes de dados, gerando insights e auxiliando a identificação de problemas e oportunidades de melhoria.

Como continuidade deste estudo, destaca-se o aprimoramento das representações visuais e das formas de interação com estas, visando deixá-las mais intuitivas e com menos sobrecarga cognitiva. Além disso, pretende-se expandir as bases de dados incluindo dados de instituições privadas. Para tanto, realizou-se um convênio com a secretaria municipal de saúde local, submetendo este projeto aos respectivos comitês de ética e pesquisa envolvidos. O projeto já foi aprovado em ambos perante o compromisso de anonimizar os dados e disponibilizar o sistema para acesso público, bem como seu código-fonte.

5. Agradecimentos

Este trabalho contou com apoio financeiro da Fundação de Amparo à pesquisa do Estado do RS (FAPERGS), do Centro de Inteligência Artificial Aplicada À Saúde (CIARS) (projeto n.º 22/2551-0000390-7) e do Instituto do Câncer Infantil do Rio Grande do Sul (ICI).

Referências

- Atty, A. T. d. M., Jardim, B. C., Dias, M. B. K., Migowski, A., and Tomazelli, J. G. (2020). Painel-oncologia: uma ferramenta de gestão. *Rev. Bras. Cancerol.*, 66(2):e-04827.
- Batko, K. and Ślęzak, A. (2022). The use of big data analytics in healthcare. *Journal of Big Data*, 9(1):3.
- Bernardi, F. A., Alves, D., Crepaldi, N. Y., Yamada, D. B., Lima, V. C., and Lopes Rijo, R. P. C. (2022). Data quality in health research: a systematic literature review. *medRxiv*.
- Chan, T., Sebok-Syer, S., Thoma, B., Wise, A., Sherbino, J., and Pusic, M. (2018). Learning analytics in medical education assessment: The past, the present, and the future. *AEM Education and Training*, 2(2):178–187.
- Costantini, A., Michels, F., Ruhl, J., Hill, S., Kohler, B., and Negoita, S. (2023). The trajectory of pediatric cancer data and collection in the united states. *Journal of Registry Management*, 50(3):82–84.
- Gazzarata, R., Almeida, J., Lindsköld, L., Cangiolli, G., Gaeta, E., Fico, G., and Chronaki, C. E. (2024). HL7 fast healthcare interoperability resources (hl7 fhir) in digital healthcare ecosystems for chronic disease management: Scoping review. *International Journal of Medical Informatics*, 189:105507.
- INCA (2023). *Estimativa 2023: incidência de câncer no Brasil*. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância, Rio de Janeiro.
- Jensen, P. B., Jensen, L. J., and Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6):395–405.
- Khan, S. S. and Qin, G. (2018). A review on healthcare data analytics: Recent advances and future directions. *Journal of King Saud University-Computer and Information Sciences*.
- Linnet, M. S., Ries, L. A. G., Smith, M. A., Tarone, R. E., and Devesa, S. S. (1999). Cancer surveillance series: Recent trends in childhood cancer incidence and mortality in the united states. *JNCI: Journal of the National Cancer Institute*, 91(12):1051–1058.
- Mohammadzadeh, Z., Ghazisaeedi, M., Nahvijou, A., Rostam Niakan Kalhori, S., Davoodi, S., and Zendehdel, K. (2017). Systematic review of hospital based cancer registries (hbcrs): Necessary tool to improve quality of care in cancer patients. *Asian Pacific Journal of Cancer Prevention*, 18(8):2027–2033.
- Obermeyer, Z. and Emanuel, E. J. (2016). Predicting the future - big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13):1216–1219.

- Osterman, T. J., Terry, M., and Miller, R. S. (2020). Improving cancer data interoperability: The promise of the minimal common oncology data elements (mcode) initiative. *JCO Clinical Cancer Informatics*, (4):993–1001. PMID: 33136433.
- Park, S., Bekemeier, B., Flaxman, A., and Schultz, M. (2022). Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. *Informatics for Health and Social Care*, 47(2):175–193.
- Rodriguez-Galindo, C., Friedrich, P., Morrissey, L., and Frazier, L. (2013). Global challenges in pediatric oncology. *Current Opinion in Pediatrics*, 25(1):3–15.