



Projeto Orientado em Computação I

TÉCNICAS DE CIÊNCIAS DOS DADOS PARA AVALIAÇÃO DO PERFIL ONCOLÓGICO DO BRASIL

DINIZ, G.M.S¹; VELOSO, A.A²
1- Acadêmico da Universidade Federal de Minas Gerais
2- Professor do Departamento de Ciência da Computação da UFMG
Email: gabrielmsd1994@gmail.com

DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

INTRODUÇÃO

Segundo uma previsão realizada pelo International Union Against Cancer (UICC), em 2020 haverá mais de 15 milhões de casos de câncer no mundo. Hoje a doença já é responsável por mais de 12% de todas as causas de óbito no planeta e, devido à sua gravidade, necessita constantemente de novos investimentos em pesquisa para compreender a multiplicidade de causas que determinam os riscos de se adquirir a doença.

Um dos maiores aliados no tratamento das neoplasias é a detecção precoce, ou seja, o diagnóstico pré-sintomático da doença. Assim, o rastreamento de pessoas com maiores riscos é imprescindível para o diagnóstico prévio e, com isso, melhor prognóstico.

Para realização desta técnica, conhecida por screening, há, de fato, uma extensa quantidade de dados relacionados aos pacientes com câncer no Brasil. No entanto, a análise minuciosa destas informações possibilitaria a definição de um perfil epidemiológico do câncer que contribuiria muito no diagnóstico precoce da doença.

OBJETIVO

Obter uma compreensão mais apurada dos fatores de riscos e do perfil oncológico do Brasil mediante análise e visualização de dados de acordo com estilo de vida, idade, hábitos de saúde e informações genéticas dos casos clínicos disponibilizados pelo Oncoguia.

METODOLOGIA

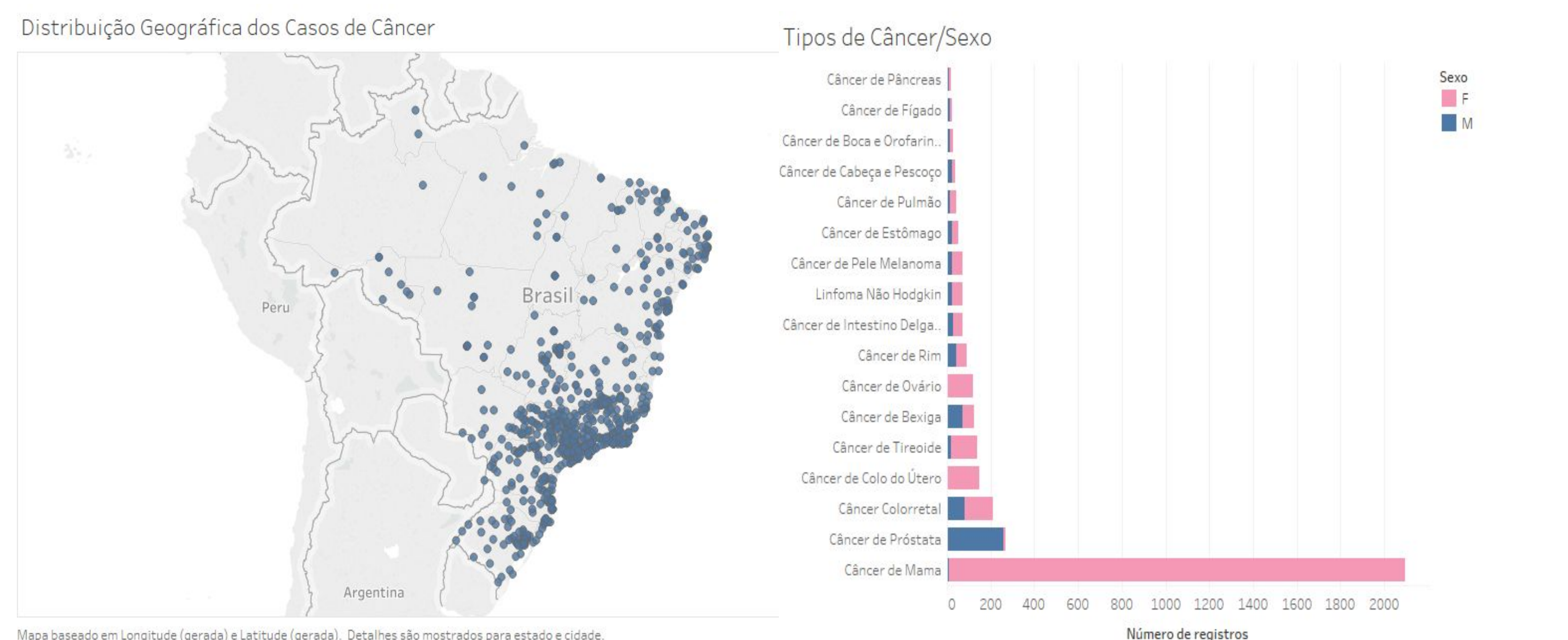
O primeiro passo foi a aquisição dos dados do Instituto Oncoguia, plataforma online que dispõe de uma ferramenta de coleta de dados de casos de câncer no Brasil.

Posteriormente, foi realizada uma preparação e o tratamento dos dados coletados para a etapa seguinte, que compreende a análise dos dados. Nessa etapa foi realizado um estudo dos variados tipos de neoplasias a fim de selecionar os parâmetros da base de dados obtidas a serem estudados.

Por fim, após a preparação e a seleção dos dados, foram utilizadas ferramentas de "Business Intelligence" e aplicada uma técnica de redução de dimensionalidade de dados de alta dimensão por meio do algoritmo t-SNE com o objetivo de realizar uma visualização inteligente dos dados.

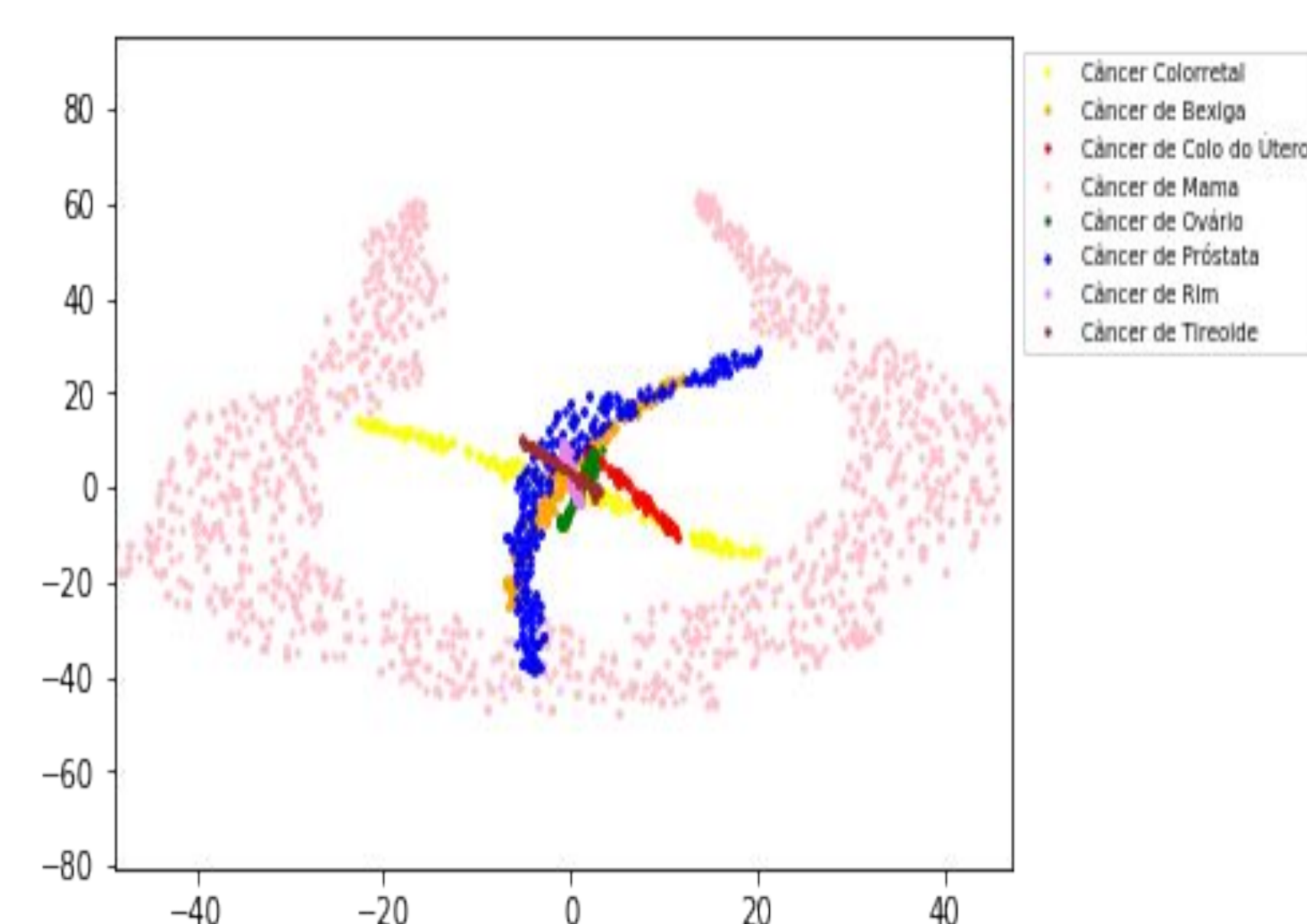
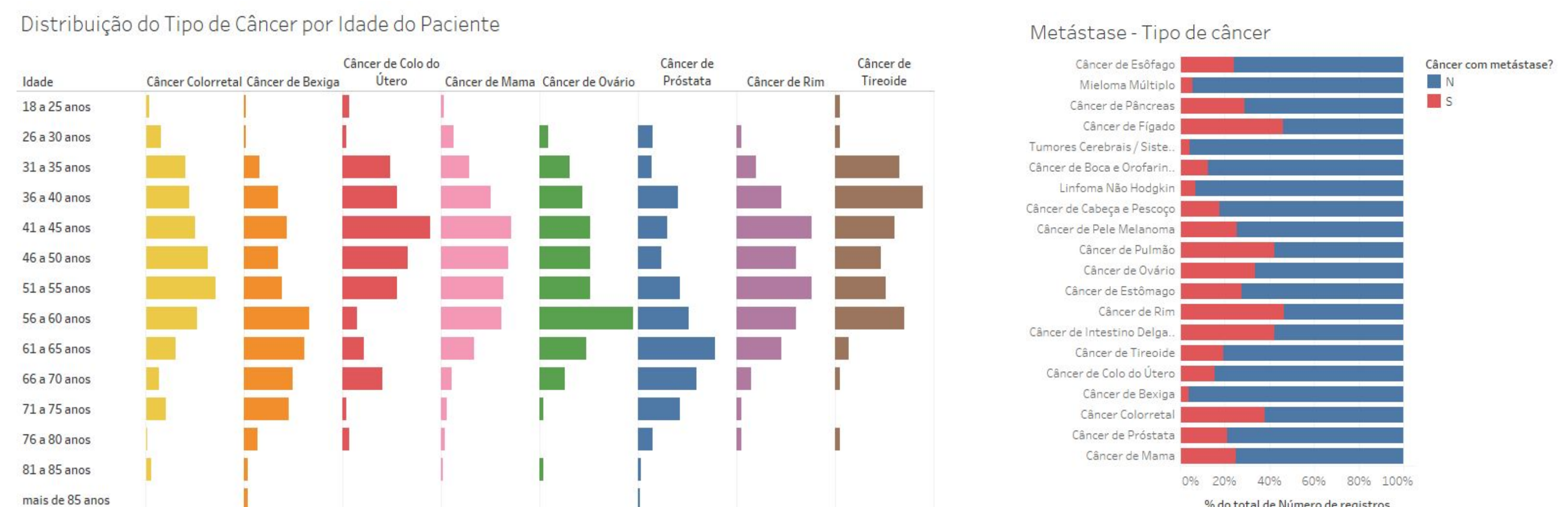
RESULTADOS

Foram elaboradas visualizações com a finalidade de ter um entendimento do público do Oncoguia. Dentre as visualizações, destacam-se as referentes à distribuição geográfica e à distribuição de gênero para os tipos de câncer mais comuns nos dados obtidos.



A metástase ocorre quando as células cancerígenas se propagam para partes distintas do corpo além de seu sítio primário. Esse processo indica que a doença se encontra em estágio avançado. Portanto, tornou-se relevante a visualização dos tipos de câncer que mais se encontram nesse estágio avançado dentre os dados obtidos.

Além disso, a técnica de rastreamento leva em consideração a idade dos pacientes para estabelecer um perfil de pessoas mais predispostas à doença. Por esse motivo foi realizado uma pirâmide etária para os tipos de câncer com maiores ocorrências no Oncoguia.



O t-SNE é uma das técnicas de redução de dimensionalidade para dados categóricos com a intenção de classificar e distanciar os dados.

Foi elaborado um t-SNE dos tipos de câncer com relação à dados como gênero, idade, condições socio-econômicas, forma de tratamento e metástase dos casos clínicos.

CONCLUSÃO

As análises e visualizações dos dados realizadas nessa pesquisa são fortes arcabouços teóricos para se estabelecer um perfil oncológico mais detalhado no Brasil, que poderá amparar a técnica de screening a obter diagnósticos mais precoces da doença.

Sobretudo os resultados obtidos poderão servir de base para ações políticas com a finalidade de elaborar uma maior prevenção do câncer e, cientificamente, servir de estudos sobre a epidemiologia da doença.

REFERÊNCIAS

1. INSTITUTO ONCOGUIA. Disponível em: <http://www.oncoguia.org.br>
2. BENGIO, Y; GOODFELLOW, I; COURVILLE, Aaron. Deep Learning. 1ª edição. MIT, 2016.
3. Caderno de Atenção Primária, nº 29. Rastreamento. Brasília: Ministério da Saúde, 2010.
4. WORLD HEALTH ORGANIZATION. Prevention. Geneva, 2007. Cancer control: knowledge into action: WHO guide for effective programmes.
5. INSTITUTO NACIONAL DE CÂNCER. A situação do câncer no Brasil. Brasília: Ministério da saúde, 2006.
6. INSTITUTO NACIONAL DE CÂNCER. A detecção precoce. Disponível em: http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_controle_cancer_mama/deteccao_precoce
7. <https://lvdmaaten.github.io/tsne/>