



# Aplicação de mineração de dados para extração de conhecimento sobre ocorrência de fístulas após palatoplastia

*Application of data mining to extract knowledge about the occurrence of fistulas after palatoplasty*

PATRICK PEDREIRA SILVA<sup>1\*</sup>  
ELVIO GILBERTO DA SILVA<sup>2</sup>  
VINICIUS SANTOS  
ANDRADE<sup>1</sup>  
TELMA VIDOTTO DE SOUSA  
BROSCO<sup>3</sup>  
GABRIELA APARECIDA  
PREARO<sup>4</sup>  
MARIA INÊS  
PEGORARO-KROOK<sup>1,4</sup>  
JENIFFER DE CASSIA  
RILLO DUTKA<sup>1,4</sup>

## ■ RESUMO

**Introdução:** As técnicas de mineração de dados ampliam o acesso a informações importantes para o processo de tomada de decisão durante os cuidados com a saúde. O objetivo do estudo propõe a utilização de técnicas de mineração de dados para identificar variáveis (protocolos de tratamento cirúrgico, características do paciente, intercorrências pós-cirúrgicas) associadas à ocorrência de fístulas após palatoplastia primária em pacientes com fissura transforame incisivo unilateral (F'TIU). **Método:** Um conjunto de dados de 222 pacientes com F'TIU sem síndromes, operados por quatro cirurgiões com as técnicas de palatoplastia primária de Furlow ou von Langenbeck, foi analisado para este estudo. Dois modelos para detecção do resultado da cirurgia foram induzidos usando técnicas de mineração de dados (Árvore de Decisão e Apriori). **Resultados:** Cinco regras foram selecionadas de uma árvore de decisão apontando para algumas variáveis como preditivas de fístulas associadas à palatoplastia primária: infecção, tosse, hipernasalidade, cirurgião. A análise do modelo indica que ele classifica corretamente 95,9% das ocorrências entre ausência e presença de fístulas. O segundo modelo indica que a ausência de intercorrências pós-cirúrgicas (infecção e febre) e resultado de fala normal (hipernasalidade ausente, sem sugestivo de disfunção velofaríngea) estão relacionados à ausência de fístulas. Em relação aos procedimentos cirúrgicos, o uso da técnica de Furlow e retalho de Vomer foram mais frequentes nos pacientes com fístulas. **Conclusão:** Técnicas de mineração de dados, conforme aplicadas no presente estudo, apontaram para infecção e tosse, presença de hipernasalidade, cirurgião e técnica cirúrgica como preditores de fístulas relacionadas à palatoplastia primária.

**Descritores:** Mineração de dados; Saúde; Fissura palatina; Fístula bucal; Algoritmos.

## ■ ABSTRACT

**Introduction:** Data mining techniques expand the access to information important to the decision-making process during healthcare. The objective the study proposes using data mining techniques to identify variables (surgical treatment protocols, patient characteristics, post-surgical complications) associated with the occurrence of fistulas after primary palatoplasty in patients with unilateral cleft lip and palate (UCLP). **Methods:** A dataset of 222 patients with UCLP, without syndromes, operated by four

Instituição: Hospital de Reabilitação de Anomalias Craniofaciais da Universidade de São Paulo, Bauru, SP, Brasil.

Artigo submetido: 3/11/2021.  
Artigo aceito: 13/9/2022.

Conflitos de interesse: não há.

DOI: 10.5935/2177-1235.2023RBCP657-PT

<sup>1</sup> Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Programa de Pós-Graduação em Ciências da Reabilitação, Bauru, SP, Brasil.

<sup>2</sup> Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Programa de Pós-Doutorado, Bauru, SP, Brasil.

<sup>3</sup> Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo, Departamento de Cirurgia Plástica, Bauru, SP, Brasil.

<sup>4</sup> Faculdade de Odontologia de Bauru, Universidade de São Paulo, Programa de Pós-Graduação em Fonoaudiologia, Bauru, SP, Brasil.

surgeons with the Furlow or the von Langenbeck primary palatoplasty techniques, was analyzed for this study. Two models for detecting the result of the surgery were induced using data mining techniques (Decision Tree and Apriori). **Results:** Five rules were selected from a decision tree pointing to some variables as predictors of fistulas associated with primary palatoplasty: infection, cough, hypernasality, surgeon. The analysis of the model indicates that it correctly classifies 95.9% of the instances between the absence and presence of fistulas. The second model indicates that the absence of postoperative complications (infection and fever) and normal speech results (absence of hypernasality, absence of velopharyngeal dysfunction) are related to the absence of fistulas. Regarding surgical procedures, the use of the Furlow technique and Vomer flap were more frequent in patients with fistulas. **Conclusion:** Data mining techniques, as applied in the current study, pointed to infection and cough, presence of hypernasality, surgeon, and surgical technique as predictors of fistulas associated with primary palatoplasty.

**Keywords:** Data mining; Health; Cleft palate; Oral fistula; Algorithms.

## INTRODUÇÃO

Um dos principais objetivos da cirurgia primária do palato na fissura labiopalatina (FLP) é a reconstrução bem-sucedida da cinta muscular dos elevadores, de forma a propiciar um mecanismo velofaríngeo funcional para produção adequada da fala<sup>1</sup>. A presença de fístula oronasal é uma das complicações mais significativas após o reparo cirúrgico do palato, já que que suas implicações podem interferir na qualidade de vida do indivíduo. A incidência de fístulas oronasais residuais é um dos fatores indicativos do sucesso do reparo cirúrgico primário do palato<sup>2-4</sup>.

A fístula, conforme reportado por Brosco<sup>4,5</sup>, é uma falha na cicatrização ou ruptura do reparo cirúrgico primário do palato que pode ocorrer em qualquer local ao longo da linha de fechamento da fenda. A literatura apresenta dados conflitantes com relação à ocorrência de fístula<sup>6-9</sup>, por exemplo, Salimi et al.<sup>10</sup> relatam incidência variando entre 0 e 78%. Para prevenir e minimizar estas complicações cirúrgicas, é importante compreender quais variáveis (protocolos de tratamento cirúrgico, intercorrências pós-cirúrgicas, resultados de fala após a cirurgia e características dos pacientes) estão associadas à ocorrência de fístulas.

A chamada “era da informação” caracteriza-se pela crescente expansão no volume de dados gerados e armazenados, fenômeno que também se reflete na área de saúde em geral, o que amplia possibilidade de obtenção de informações importantes no apoio ao processo decisório<sup>11</sup>. Os dados dos pacientes, bem como os resultados das cirurgias, ficam disponibilizados em seus prontuários, podendo ser usados como elementos para estudos clínicos.

Porém, muitas vezes, o volume de dados gerados é tão grande que dificulta sua utilização e análise manual, demandando processos mais sofisticados como, por exemplo, os processos automatizados, para a

manipulação de tais dados. É exatamente neste contexto de superabundância de dados que surgiu a mineração de dados, como um processo sistemático, interativo e iterativo, de preparação e extração automática de conhecimento a partir de bases de dados<sup>11,12</sup>.

Na mineração de dados a indução de hipóteses se dá a partir de um conjunto de dados observados, como, por exemplo, dados sobre pacientes. Cada paciente é denominado objeto e sobre cada objeto são armazenados diversos atributos (nome, identificação, sexo, idade, sintomas, etc.), que correspondem aos diversos dados desse paciente. Em uma das tarefas típicas da mineração busca-se aprender formas de se prever um dos atributos (esse atributo específico do qual se deseja fazer a predição é denominado de classe ou, simplesmente, atributo alvo ou atributo de saída). Os demais atributos utilizados para se fazer a predição do atributo alvo são chamados de preditores ou atributos de entrada.

A partir de um conjunto de dados busca-se criar um modelo ou hipótese (representados por um algoritmo ou conjunto de regras) capaz de relacionar um ou mais atributos (preditores) ao atributo alvo (classe). Por meio de um viés indutivo, cada modelo identificado a partir da mineração de dados utiliza uma representação para descrever a hipótese induzida a partir do conjunto de dados.

## OBJETIVO

Este trabalho tem por objetivo utilizar técnicas de mineração de dados para a extração automática de conhecimento sobre variáveis (protocolos de tratamento cirúrgico, intercorrências pós-cirúrgicas, resultados de fala após a cirurgia e características dos pacientes) associadas à ocorrência de fístulas oronasais em pacientes com fissura transforame incisivo unilateral (FTIU).

## MÉTODOS

A investigação trata de uma pesquisa descritiva, quantitativa, experimental e aplicada, aprovada pelo Comitê de Ética em Pesquisa do Hospital de Reabilitação de Anomalias Craniofaciais, da Universidade de São Paulo (parecer 1.753.467), realizada nessa instituição em setembro de 2021. A amostra refere-se a um subconjunto de prontuários de pacientes com fissura labiopalatina participantes de um estudo clínico randomizado (ECR) com FTIU<sup>13</sup>.

Dados sobre a ocorrência de fistulas foram obtidos para o total de 466 pacientes (bebês). Estes pacientes foram randomizados (usando-se um *script* – código escrito em linguagem de programação, desenvolvido na Universidade da Flórida) para receber diferentes protocolos de tratamento cirúrgico incluindo: 1) queiloplastia primária entre 3 e 6 meses de idade com a técnica de Millard (M) ou Spina (S); 2) palatoplastia precoce (9 a 12 meses) ou tardia (>12 meses); 3) palatoplastia primária com a técnica de von Langenbeck (VL) ou de Furlow (F); e 4) para um de quatro possíveis cirurgões (C1, C2, C3, C4).

As informações sobre a ocorrência de fístula após a palatoplastia primária foram de interesse para o presente estudo. Para determinação das classes na mineração de dados, usou-se a classificação de Spina<sup>14</sup> agrupando-se os pacientes em dois grupos: SUCESSO (pacientes sem fístula ou com fístula na região pré-forame incisivo); INSUCESSO (pacientes com fístula na região pós-forame incisivo ou fístulas transforame). O forame incisivo demarca os limites dos palatos primário (parte central do lábio superior e pré-maxila) e secundário (palato duro e mole).

A Tabela 1 apresenta as informações identificadas nos prontuários dos pacientes para este estudo. A definição das variáveis de interesse está relacionada diretamente a alguns fatores que incluem: 1) Protocolos de tratamento cirúrgico (técnica cirúrgica na queiloplastia e na palatoplastia, cirurgião, uso de modificações na cirurgia como incisão relaxante e retalho de vômer, duração da palatoplastia em minutos); 2) Características do paciente (idade na palatoplastia, tempo da palatoplastia); 3) Intercorrências pós-cirúrgicas (se ocorreu infecção na palatoplastia - no local - ou em outro local após a palatoplastia primária; se houve vômito ou tosse no pós-operatório da palatoplastia); 4) Resultados de fala após a cirurgia (se houve diagnóstico sintomático da disfunção velofaríngea, presença de hipernasalidade - registrada em conversa espontânea ou dirigida); resultados dos testes de emissão de ar nasal, de hipernasalidade e de hiponasalidade (observados durante repetição de 10 vocábulos); 5) Resultado da cirurgia quanto à ocorrência de fístula (SUCESSO ou INSUCESSO). As variáveis de interesse estão listadas na coluna “nome do atributo”.

As variáveis associadas aos protocolos de tratamento cirúrgico e às características do paciente podem ajudar a indicar se há maior propensão ao desenvolvimento de fístulas (ainda antes do procedimento cirúrgico), já as variáveis relacionadas às intercorrências pós-cirúrgicas e aos resultados de fala podem ser indicativas de fístulas clinicamente relevantes após a cirurgia.

No gerenciamento da FLP o SUCESSO do tratamento ocorre na ausência de fístula e ausência de alterações de fala. Para o presente estudo, a ocorrência de fístula em região posterior ao forame incisivo e presença de disfunção velofaríngea foram interpretadas como indicativas de INSUCESSO do tratamento. A pergunta norteadora da mineração de dados envolveu a verificação de quais fatores (protocolos de tratamento cirúrgico, características do paciente, intercorrências pós-cirúrgicas e resultados de fala após a cirurgia) estariam associados à ocorrência ou não das fístulas. Buscou-se, portanto, identificar com este estudo se algumas das

**Tabela 1.** Definição das variáveis (atributos) de interesse para este estudo.

Variáveis (Nome do Atributo)	Categorias (Valores)
Técnica cirúrgica na queiloplastia	Millard, Spina
Tempo da palatoplastia	Precoce (9-12m), Tardio (>12m)
Idade na palatoplastia	Meses (m)
Técnica cirúrgica na palatoplastia	Furlow, von Langenbeck
Cirurgião	C1, C2, C3, C4
Incisão relaxante	Sem incisão, unilateral, bilateral
Retalho de Vômer	sim, não
Duração da palatoplastia	Minutos
Infecção na palatoplastia em outro local	Não houve; no local da cirurgia,
Vômito no pós-operatório da palatoplastia	sim, não
Tosse no pós-operatório da palatoplastia	sim, não
Febre	sim, não
Sugestivo de disfunção velofaríngea	sim, não
Hipernasalidade	sim, não
Teste de emissão de ar	[1-10]
Teste de hipernasalidade	[1-10]
Teste de hiponasalidade	[1-10]
*Ocorrência de fístula	SUCESSO, INSUCESSO

m=meses; C=cirurgião; \*atributo alvo (classe)

variáveis analisadas podem ser usadas como preditoras da ocorrência de fístula no palato ou como indicativos de fístulas clinicamente relevantes pós-palatoplastia.

Para o cômputo dos resultados do experimento, foi utilizado o algoritmo C4.5 (J48) que gera árvores de decisão para encontrar a relação entre as características consideradas e os resultados das cirurgias, bem como o “algoritmo Apriori” (de associação) para a geração de regras. A utilização de árvores de decisão permite que as variáveis ou os atributos possam ser categóricos (qualitativos) ou numéricos (quantitativos), podendo ser usados simultaneamente pelo modelo (o que se mostrou adequado considerando os diferentes tipos de variáveis da base de dados usada na investigação). Já o algoritmo Apriori lida somente com variáveis qualitativas. Ambos os modelos induzem uma hipótese por meio de um modelo representado por regras (“se...então”).

Nesta análise as variáveis de interesse foram tratadas como os atributos no *software* WEKA. Considerando uma tarefa de mineração típica, o experimento foi dividido em quatro etapas: pré-processamento dos dados, extração de características, classificação e descrição de resultados. O procedimento foi realizado considerando-se como resultado primário a ocorrência de fístula após a palatoplastia. O pré-processamento foi realizado de maneira semiautomática. Os dados dos prontuários disponibilizados no formato de arquivo “.XLS” (planilha de Excel®) foram convertidos para o formato “.ARFF” (usado pelo WEKA) por meio do *software* Excel2ArffConverter. Antes da conversão os atributos foram identificados conforme descrito na Tabela 1.

## RESULTADOS

Somente os pacientes com dados completos foram selecionados para análise, considerando os parâmetros descritos na Tabela 1. Após descartar os pacientes com dados incompletos para qualquer uma das variáveis, um total de 222 pacientes foram selecionadas para análise. Devido à possibilidade de enviesamento da base, optou-se por não se estimar os valores ausentes<sup>7</sup>. As informações sobre ocorrência de algum tipo de fístula foram identificadas nos prontuários de 222 (47,6%) dos 466 pacientes estudados e dados destes pacientes foram os minerados para o presente artigo.

No grupo de 222 pacientes considerados para este estudo 98 (44,1%) indivíduos pertenciam ao sexo feminino e 124 (55,9%) ao sexo masculino. A idade média de realização da palatoplastia primária foi de 12,8 meses ( $\sigma=3,2$ ). Neste grupo 114 (51,3%) recebeu o procedimento de Millard na queiloplastia primária, enquanto 108 (48,7%) recebeu o procedimento de Spina.

Um total de 112 pacientes (50,4%) recebeu a técnica de Furlow na palatoplastia primária enquanto 110 (49,6%) receberam von Langenbeck. Da amostra pacientes, 182 (81,9%) pertenciam ao grupo SUCESSO e 40 (18,1%) ao grupo INSUCESSO.

Por meio da construção de uma árvore de decisão, 37 regras foram geradas a partir do conjunto completo dos dados referentes aos pacientes. Entretanto, neste artigo optou-se por exibir apenas as 5 regras com maior valor para a métrica de cobertura de cada resultado final da cirurgia (SUCESSO ou INSUCESSO). A métrica de cobertura é a proporção entre os dados classificados corretamente e o total de dados da amostra daquela classe. Também foram consideradas informações sobre a métrica de precisão da regra (probabilidade do resultado final condicionado aos atributos, isto é, a capacidade do modelo de evitar falsos positivos).

A média de precisão das regras associadas ao SUCESSO cirúrgico é de 97,26% ( $\sigma=4,59$ ). As cinco regras juntas apresentam uma cobertura de cerca de 77,5%, ou seja, se aplicadas aos dados, conseguem detectar 77,5% dos casos de SUCESSO. Já para a classe de INSUCESSO, a média de precisão das regras associadas é de 84,32% ( $\sigma=9,40$ ). A cobertura das cinco regras é de 62,5%, isto é, a quantidade de casos de INSUCESSO que as regras conseguem detectar se aplicadas à base de dados.

A regra com maior cobertura e precisão para predição de um bom resultado indica que os principais fatores envolvidos são: infecção (“ausência”), testes de hipernasalidade (“ $\leq 6$ ”) e hiponasalidade (“ $> 9$ ”) e a técnica cirúrgica (“von Langenbeck”). A interpretação desta regra indica, portanto, que pacientes submetidos ao procedimento “von Langenbeck”, sem infecção e com resultados de teste de hiponasalidade com valores maiores ou iguais a 9 e teste de hipernasalidade com valores menores ou iguais a 6, têm maior probabilidade de obterem SUCESSO como resultado final da cirurgia. Já para o INSUCESSO, segundo as duas regras com maior precisão e cobertura, os fatores envolvidos são relacionados às intercorrências pós-cirúrgicas e resultados de fala e incluem: infecção (“ausência ou em outro local”), testes de hipernasalidade (“maior que 6”), emissão de ar (“maior que 9”) e febre (“sim”). As regras são exibidas na Tabela 2.

Quando se analisa o desempenho global do modelo (árvore de decisão gerada) com relação à sua capacidade preditiva, observa-se que ele classifica corretamente 95,9% dos pacientes e incorretamente apenas 4,1%. Considerando cada categoria individualmente, o modelo consegue acertar 90,0% dos casos em que ocorre um resultado de INSUCESSO. Já para a outra classe, o modelo consegue acertar 97,3% dos casos em que ocorre um resultado de SUCESSO.

**Tabela 2.** Resultado da cirurgia.

Número	Regra	Resultado (classe)	Cobertura	Precisão
1	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=von Langenbeck” e “teste de hiponasalidade>9”	SUCESSO	77	100%
2	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=Furlow” e “amplitude da fissura=regular”	SUCESSO	33	96,9%
3	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=Furlow” e “amplitude da fissura=ampla” e “cirurgião=C3”	SUCESSO	19	89,4%
4	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=Furlow” e “amplitude da fissura=ampla” e “cirurgião=C2”	SUCESSO	7	100%
5	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=Furlow” e “amplitude da fissura=ampla” e “cirurgião=C1” e “incisão relaxante=sem”	SUCESSO	5	100%
6	Se “infecção=não houve” e “teste de hipernasalidade>6” e “teste de emissão de ar>9” e “febre=sim”	INSUCESSO	6	83,3%
7	Se “infecção=ocorreu em outro local”	INSUCESSO	6	83,3%
8	Se “infecção=não houve” e “teste de hipernasalidade>6” e “teste de emissão de ar>9” e “febre=não” e “incisão relaxante=bilateral” e “vômito=não” e “cirurgião=C3”	INSUCESSO	5	80,0%
9	Se “infecção=não houve” e “teste de hipernasalidade≤6” e “tosse=não” e “técnica cirúrgica=Furlow” e “amplitude da fissura=ampla” e “cirurgião=C4” e “teste de emissão de ar>2”	INSUCESSO	4	100%
10	Se “infecção=não houve” e “teste de hipernasalidade>6” e “teste de emissão de ar>9” e “febre=não” e “incisão relaxante=sem”	INSUCESSO	4	75,0%

As correlações encontradas usando o algoritmo Apriori foram obtidas usando as métricas de suporte (mínimo de 60%) e confiança (mínimo de 90%). O objetivo foi encontrar regras que fossem frequentes (alto valor de suporte) na base de dados e com alto grau de confiança (diretamente relacionada à validade da regra). Foram encontradas quatro regras com uma confiança média de 90,75% ( $\sigma=0,5$ ) e um suporte médio de 69,45% ( $\sigma=0,49$ ), que atendem os requisitos citados, conforme Tabela 3.

Considerando-se apenas os 40 pacientes do grupo de INSUCESSO, os resultados mostram as seis regras encontradas com suporte mínimo de 67,5% e confiança mínima de 100% (Tabela 3). As regras apresentam um suporte médio de 72,08% para este grupo. No grupo de SUCESSO o modelo indica ausência de intercorrências pós-cirúrgicas (tosse e infecção) e resultado de fala com hipernasalidade ausente. Pacientes do grupo de INSUCESSO também apresentaram ausência de tosse e de infecção.

A Tabela 4 traz um resumo da relação entre a duração da palatoplastia e o resultado quanto à ocorrência

de fístula. Observa-se que os tempos de cirurgia variam de 25 a 140 minutos.

Os algoritmos utilizados permitem uma exploração baseada em dados de relações não lineares e interações entre muitas variáveis, gerando modelos que são fáceis de interpretar. Entretanto, como fraqueza do método pode-se apontar o desbalanceamento entre os dois grupos (SUCESSO e INSUCESSO) e a utilização integral da amostra para a indução dos modelos, o que pode causar um super-ajuste aos dados, prejudicando a extrapolação dos achados (regras) para outras bases de dados.

## DISCUSSÃO

Especificamente com relação às ocorrências de fistulas, as regras encontradas com alto grau de precisão e cobertura podem evidenciar padrões úteis sobre quais variáveis, dentre protocolos de tratamento cirúrgico, características do paciente, resultados da fala após cirurgia e intercorrências pós-cirúrgicas, são determinantes para o sucesso ou insucesso da palatoplastia. A oportunidade de adotar a mineração sobre



**Tabela 3.** Regras com alto valor de suporte e confiança.

Características	Resultado	Suporte	Confiança
Ausência de tosse e infecção sem sugestivo de disfunção velofaríngea	SUCESSO	69,8%	91,0%
Ausência de tosse e infecção com hipernasalidade ausente	SUCESSO	69,8%	91,0%
Ausência de tosse e infecção sem sugestivo de disfunção velofaríngea com hipernasalidade ausente	SUCESSO	69,8%	91,0%
Ausência de tosse e infecção e sem febre	SUCESSO	68,4%	90,0%
Ausência de tosse	INSUCESSO	77,5%	100,0%
Ausência de infecção	INSUCESSO	77,5%	100,0%
Técnica cirúrgica de Furlow	INSUCESSO	72,5%	100,0%
Utilização de retalho de Vômer	INSUCESSO	70,0%	100,0%
Ausência de vômito	INSUCESSO	67,5%	100,0%
Ausência de tosse e de infecção	INSUCESSO	67,5%	100,0%

**Tabela 4.** Relação entre duração da palatoplastia e classes (SUCESSO e INSUCESSO).

Duração: Minutos	N	Média	Desvio padrão	Mínimo	Máximo
Duração da palatoplastia – Todos os grupos	222	65,62	24,43	25	140
Duração da palatoplastia (grupo SUCESSO)	182	62,57	22,89	25	125
Duração da palatoplastia (grupo INSUCESSO)	40	79,5	26,62	25	140

os dados de pacientes submetidos à palatoplastia pode oportunizar melhor entendimento das especificidades que podem ocorrer com o grupo de pacientes, ampliando, assim, o conhecimento do profissional na identificação das condutas a serem adotadas.

Neste estudo específico, a visibilidade dada a alguns fatores (Tabela 1) permite que profissionais de saúde, com a devida análise desse conjunto de descobertas, possam identificar padrões de associação de variáveis, as quais possam dar significado às ações diagnósticas e terapêuticas. Da mesma forma do que em outros estudos prévios, esta investigação optou pela combinação de diferentes tipos de tarefas de mineração de dados para a realização do experimento ou identificação de padrões<sup>15-19</sup>.

Apesar da disponibilidade inicial de dados referentes a 466 pacientes, optou-se pelo uso de 222 (considerando apenas aqueles completos). Isso pode ter limitado as regras obtidas, bem como não ter evidenciado outras associações dos fatores relativas aos resultados finais da palatoplastia. Tal decisão segue as diretrizes de outros trabalhos<sup>20</sup>. Para estudos futuros, toda a base poderá ser usada, pois alguns algoritmos podem lidar com dados faltantes<sup>11</sup>.

Outra limitação referente à base é o fato de as duas classes consideradas estarem desbalanceadas, entretanto, como elas refletem a situação real em que os resultados de SUCESSO são mais comuns que os INSUCESSOS, optou-se pela manutenção da proporção natural dos dados. Essa presença de classes

majoritárias com frequência muito maior que as outras classes minoritárias faz com que os algoritmos tenham uma tendência para responder bem para as classes majoritárias em detrimento das minoritárias. Como trabalhos futuros, o experimento poderá ser repetido usando técnicas de amostragem aleatória dos dados de forma a gerar conjuntos balanceados<sup>21</sup>.

O fato de toda a base ter sido usada para indução e teste do modelo pode gerar um viés de ajuste aos dados. Qualquer método de mineração está sujeito a gerar um modelo que se super-ajusta aos dados sobre os quais ele foi induzido (*overfitting*), mas não consegue generalizar o conhecimento aprendido, não obtendo um bom desempenho quando confrontado com dados de uma outra base. Entretanto, como o intuito deste experimento não é induzir um modelo visando automatizar o processo de classificação de cirurgias, mas sim extrair regras que possam ser avaliadas por humanos, evidenciando padrões úteis, optou-se por esta abordagem.

A análise da Tabela 2 indica que os resultados de SUCESSO estão associados às intercorrências pós-cirúrgicas como ausência de infecção e de tosse, além disso, os pacientes apresentaram teste de hipernasalidade abaixo ou igual a 6 (numa escala que vai até 10). No caso das fissuras amplas associadas à técnica cirúrgica de Furlow, além das intercorrências ressaltadas, há também a influência do fator cirúrgico para o resultado final.

No caso dos resultados de INSUCESSO, a presença de infecção parece ser um fator importante,

entretanto, não é decisivo. Devido à similaridade entre as regras 9, 3 e 4 (Tabela 2) o fator decisivo para obtenção de um resultado de INSUCESSO está atrelado ao cirurgião. Sob as mesmas condições, os cirurgiões C2 e C3 obtiveram resultados de SUCESSO, entretanto, o cirurgião C4 obteve SUCESSO em apenas em 50% das cirurgias, o que pode indicar a influência do fator cirurgião. No caso dos resultados de fala, valores de testes de hipernasalidade superiores a 6 são indicativos de um possível INSUCESSO.

Do mesmo modo que as regras da árvore de decisão, as regras apresentadas pelo modelo induzido pelo algoritmo Apriori deverão ser avaliadas por um profissional, de forma a validá-las frente à realidade. O algoritmo Apriori não lida com atributos quantitativos, somente com os categóricos, o que exige a exclusão de alguns atributos ou mesmo a transformação deles para dados não numéricos (processo de discretização); essa estratégia foi utilizada em alguns processamentos realizados neste trabalho. Assim, para evitar essa limitação em trabalhos futuros, outros algoritmos poderão ser experimentados, tais como AprioriTid, SETM, AprioriHybrid<sup>22</sup>.

A análise da Tabela 3 indica que, de um modo geral, a ausência de intercorrências pós-cirúrgicas (infecção e febre) e resultado de fala com hipernasalidade ausente, bem como pacientes sem sugestivo de disfunção velofaríngea, apresentam SUCESSO pós-palatoplastia primária. Com relação aos procedimentos cirúrgicos, há indícios de que a utilização da técnica de Furlow e de retalho de Vômer são frequentes no grupo de INSUCESSO. Já observações como ausência de tosse, vômito ou infecção, isoladamente, não podem ser usadas como parâmetros para descartar um possível INSUCESSO.

A análise da Tabela 4 mostra que uma palatoplastia do grupo dos pacientes que tiveram um resultado de INSUCESSO dura em média 79,5 minutos; já para o grupo de pacientes com resultados de SUCESSO, a média cai para 62,57 minutos. Há indícios, portanto, que cirurgias mais demoradas tendem a causar piores resultados.

Reconhece-se, finalmente, que este estudo oferece somente uma perspectiva pontual da realidade, por meio da análise de modelos induzidos por técnicas de mineração de dados na base considerada, já que revela apenas alguns fatores associados aos resultados da palatoplastia do ponto de vista dos algoritmos de mineração, havendo a necessidade de validação por profissionais da área de saúde.

## CONCLUSÃO

A análise dos dados revelou que a ausência de algumas intercorrências pós-cirúrgicas (febre,

tosse, infecção) em conjunto com resultados de fala após a cirurgia (hipernasalidade, sugestivo de disfunção velofaríngea) e com características associadas protocolos de tratamento cirúrgico (técnica, retalho de vômer, cirurgião) pode ajudar a prever o sucesso ou insucesso da palatoplastia.

## COLABORAÇÕES

- PPS** Análise e/ou interpretação dos dados, Análise estatística, Coleta de Dados, Investigação, Redação - Preparação do original.
- EGS** Análise e/ou interpretação dos dados, Coleta de Dados, Metodologia, Redação - Preparação do original.
- VSA** Aprovação final do manuscrito, Coleta de Dados, Redação - Preparação do original, Redação - Revisão e Edição
- TVSB** Aprovação final do manuscrito, Redação - Preparação do original, Redação - Revisão e Edição, Visualização
- GAP** Gerenciamento de Recursos, Redação - Preparação do original, Redação - Revisão e Edição.
- MIPK** Redação - Preparação do original, Redação - Revisão e Edição, Visualização.
- JCRD** Aprovação final do manuscrito, Gerenciamento do Projeto, Redação - Revisão e Edição, Supervisão.

## REFERÊNCIAS

- Dutka JCR, Pegoraro-Krook MI. Avaliação e tratamento das disfunções velofaríngeas. In: Marchesan IQ, Silva HJ, Tomé MC, eds. Tratado das Especialidades em Fonoaudiologia. São Paulo: Grupo Gen-Editora Roca; 2014. p. 363-8.
- de Agostino Biella Passos V, de Carvalho Carrara CF, da Silva Dalben G, Costa B, Gomide MR. Prevalence, cause, and location of palatal fistula in operated complete unilateral cleft lip and palate: retrospective study. *Cleft Palate Craniofac J*. 2014;51(2):158-64.
- Deshpande GS, Campbell A, Jagtap R, Restrepo C, Dobie H, Keenan HT, et al. Early Complications After Cleft Palate Repair. *J Craniofac Surg*. 2014;25(5):1614-8.
- Brosco TVS. Fistula de palato após reparo da fissura labiopalatina em um estudo clínico randomizado [Tese]. Bauru: Hospital de Reabilitação de Anomalias Craniofaciais, Universidade de São Paulo; 2017. 168 p.
- Brosco TVS, Prearo GA, Silva HLA, Dutka JCR. Brosco-Dutka classification system for palate fistulas. *Rev Bras Cir Plást*. 2021;36(2):164-72.
- Bykowski MR, Naran S, Winger DG, Losee JE. The Rate of Oronasal Fistula Following Primary Cleft Palate Surgery: A Meta-Analysis. *Cleft Palate Craniofac J*. 2015;52(4):e81-7.
- Hardwicke JT, Landini G, Richard BM. Fistula incidence after primary cleft palate repair: a systematic review of the literature. *Plast Reconstr Surg*. 2014;134(4):618e-27e.

8. Timbang MR, Gharb BB, Rampazzo A, Papay F, Zins J, Doumit G. A systematic review comparing Furlow double-opposing Z-plasty and straight-line intravelar veloplasty methods of cleft palate repair. *Plast Reconstr Surg.* 2014;134(5):1014-22.
9. Park MS, Seo HJ, Bae YC. Incidence of fistula after primary cleft palate repair: a 25-year assessment of one surgeon's experience. *Arch Plast Surg.* 2022;49(1):43-9.
10. Salimi N, Aleksejūnienė J, Yen EH, Loo AY. Fistula in Cleft Lip and Palate Patients-A Systematic Scoping Review. *Ann Plast Surg.* 2017;78(1):91-102.
11. Castro LN, Ferrari DG. Introdução à mineração de dados: Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva; 2017.
12. Netto AV, Berton L, Takahata AK. Ciência de Dados e a Inteligência Artificial na Área da Saúde. São Paulo: Editora dos Editores; 2021.
13. Dutka JCR. Estudo clínico randomizado - Projeto Florida (ECR-PF- ementa fase 2): função velofaríngea para a fala e estudo do crescimento da face e dos arcos dentários após a palatoplastia primária. Projeto em andamento e com aprovação Ética do CEP/CONEP desde 02/09/2016. Pesquisador Responsável: Jeniffer de Cássia Rillo Dutka. CAAE: 57727416.9.0000.5441. Instituição Proponente: Hospital de Reabilitação de Anomalias Craniofaciais da USP.
14. Spina V, Psillakis JM, Lapa FS, Ferreira MC. Classificação das fissuras lábio-palatais: sugestão de modificação. *Rev Hosp Clín Fac Med (São Paulo).* 1972;27:5-6.
15. West D, Mangiameli P, Rampal R, West V. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. *Eur J Oper Res.* 2005;162(2):532-51.
16. Steiner MTA, Soma NY, Shimizu T, Nievola JC, Steiner Neto PJ. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. *Gest Prod.* 2006;13(2):325-37.
17. Bodini Junior AC. Utilização de técnicas de data mining na detecção de outliers em auxílio à auditoria operacional com um estudo de caso com dados do sistema de informações hospitalares [Tese]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2009.
18. Kuretzki CH. Técnicas de mineração de dados aplicadas em bases de dados da saúde a partir de protocolos eletrônicos [Dissertação]. Curitiba: Universidade Federal do Paraná; 2009. 98 p.
19. Salarte Martínez GR, Castro Bermúdez YV. Modelo híbrido para el diagnóstico de enfermedades cardiovasculares basado en inteligencia artificial. *Tecnura.* 2012;16(33):35-52.
20. Carvalho DR, Moser AD, Silva VA, Dallagassa MR. Mineração de dados aplicada à fisioterapia. *Fisioter Mov.* 2012;(3):595-605.
21. Goldschmidt R, Passos E, Bezerra E. Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações. 2ª ed. Rio de Janeiro: Elsevier; 2015.
22. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* 1993;22(2):207-16.

\*Autor correspondente: **Patrick Pedreira Silva**

R. Silvio Marchione, 3-20, Vila Nova, Cidade Universitaria, Bauru, SP, Brasil.

CEP: 17012-900

E-mail: patrickpsilva@alumni.usp.br