

# Análise da via de parto a partir de seus preditores obstétricos: uma abordagem baseada em mineração de dados

Marcelo Santos Júnior<sup>1</sup>, Juliano S. Gaspar<sup>2</sup>, Zilma S. Reis<sup>2</sup>, Cristiane Neri Nobre<sup>1</sup>

<sup>1</sup> Pontifícia Universidade Católica de Minas Gerais

<sup>2</sup>Faculdade de Medicina da UFMG

{marrsantosjunior, julianogaspar, zilma.medicina}@gmail.com,

nobre@pucminas.br

**Abstract.** *Brazil has one of the highest cesarean rates in the world, with approximately 55% of the total number of deliveries, while the World Health Organization (WHO) expects it to be 10% to 15%. This work presents the characterization of pregnant women submitted to childbirth in maternity-school, Hospital das Clínicas (HC) of Federal University of Minas Gerais (UFMG), using data mining techniques, considering the database provided by the Sismater<sup>©</sup> system. Through a decision tree generated with the algorithm C4.5, with precision and recall of 95.0% and 97.4% respectively for vaginal and 93.6% and 88.1% for cesarean sections, it was possible to observe the main characteristics of women who give birth: trial of labour in the time of admission, fetal presentation, gestational age, twinning, and number of previous cesarean sections.*

**Resumo.** *O Brasil é detentor de uma das maiores taxas de cesárea do mundo, com aproximadamente 55% do número total de partos, enquanto o esperado pela Organização Mundial de Saúde (OMS) é de 10% a 15%. O presente trabalho apresenta a caracterização das gestantes submetidas a parto, na Maternidade do Hospital das Clínicas da Universidade Federal de Minas Gerais, a partir da utilização de técnicas de mineração de dados, considerando-se uma base de dados disponibilizada pelo sistema hospitalar Sismater<sup>©</sup>. Através de uma árvore de decisão gerada com o algoritmo C4.5, com precisão e sensibilidade de, respectivamente, 95,0% e 97,4% para vaginal e 93,6% e 88,1% para cesáreas, foi possível observar as principais características de mulheres submetidas a parto: trabalho de parto na admissão obstétrica, estática fetal, idade gestacional, gemelaridade e quantidade de cesáreas prévias.*

## 1. Introdução

As altas taxas de cesarianas já são uma realidade mundial e têm se transformado em um grave problema de saúde pública. Desde 1985, a comunidade médica internacional sugere que a taxa ideal de cesárea seja entre 10% e 15%. Entretanto, os procedimentos de cesárea vêm se tornando cada vez mais frequente, seja em países desenvolvidos ou em desenvolvimento (Organização Mundial de Saúde, 2015).

O Brasil está entre os países que possuem as maiores taxas, de aproximadamente 55%, algo muito acima do limite de 15% esperado pela Organização Mundial de Saúde (OMS) (Betrán et al., 2016). Quando realizadas por motivos médicos, as cesáreas podem efetivamente reduzir a mortalidade e a morbidade materna e perinatal. Assim, o

grande problema não são, necessariamente, as altas taxas, mas, sim, a quantidade dos procedimentos realizados em casos que, muitas vezes, não seria necessário. Assim como qualquer cirurgia, uma cesárea acarreta riscos imediatos e a longo prazo.

Atualmente existe muita disponibilidade de dados em saúde, porém ainda com pouca aplicabilidade na área. Uma alternativa para tratar essa complexidade é a utilização de técnicas de inteligência artificial para extração de conhecimento a partir de base de dados existentes. Essas técnicas permitem a criação de modelos preditivos que evoluem a partir da inclusão de novos dados (Koh e Tan, 2005).

O Sistema Sismater<sup>©</sup> nasceu diante dos desafios enfrentados no processo de informatização dos registros clínicos de saúde associados ao nascimento. Uma equipe multidisciplinar, composta por profissionais da saúde e da ciência da computação, idealizou e implantou o sistema na Maternidade do Hospital das Clínicas da Universidade Federal de Minas Gerais (HC-UFMG) (Gaspar et al., 2013). O projeto foi criado com o objetivo de gerar indicadores de qualidade da assistência materno-neonatal na maternidade, a partir da coleta e registro dos dados clínicos. Dessa forma, a larga utilização do sistema pelos profissionais da saúde possibilitou o desenvolvimento de um grande banco de dados sobre o cuidado obstétrico hospitalar, contemplando informações epidemiológicas, administrativas e clínicas.

Considera-se importante o entendimento do perfil de parturientes e fatores associados ao tipo de parto. Esta informação pode fundamentar decisões de gestão, decisões clínicas e políticas públicas que visem a redução das taxas de cesárea. Certamente a utilização de novas tecnologias baseadas em inteligência artificial pode criar um cenário positivo para esta situação. Neste alinhamento, as perguntas que sintetizam este trabalho de pesquisa são: Qual o perfil das parturientes da maternidade deste Hospital em questão? quais fatores associados ao tipo de parto? Como técnicas de inteligência artificial poderiam apoiar a traçar este perfil?

As árvores de decisão surgem como uma importante ferramenta para a medicina em situações onde a tomada de decisão deve ser realizada de forma eficaz e confiável. Na medicina o processo de decisão possui um papel muito importante, especialmente no ramo da medicina diagnóstica. Sendo assim, técnicas de inteligência artificial e extração de regras através de árvores de decisão, podem trazer grandes benefícios na resolução de problemas médicos (Podgorelec et al., 2002).

Desta forma, este trabalho tem como objetivo caracterizar, por meio de regras, o perfil das mulheres atendidas na Maternidade do Hospital das Clínicas da UFMG. Ao utilizar o modelo de classificação por árvore de decisão, é possível traçar este perfil através das regras geradas pela árvore.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico utilizado. Na Seção seguinte, são apresentados os trabalhos relacionados que foram encontrados durante a pesquisa. A Seção 4 traz a metodologia utilizada para a realização deste trabalho. Na Seção 5 são apresentados e discutidos os resultados obtidos e, por fim, a Seção 6 exhibe as considerações finais e apresenta as sugestões para pesquisas futuras.

## **2. Referencial Teórico**

### **2.1. Parto vaginal X cesariana**

As cesarianas são um procedimento cirúrgico desenvolvido originalmente para salvar as vidas da criança e da mãe em situações nas quais ocorrem complicações durante a gestação ou no próprio parto. Como todo procedimento cirúrgico, a cesárea não é isenta de riscos, de modo que está associada, no Brasil e no mundo, à maior taxa de mortalidade materna e infantil, quando comparada aos partos realizados por via vaginal (Barbosa et al., 2003).

Qualquer intervenção médica cirúrgica, éticamente, deve ser realizada baseada no balanço entre os possíveis riscos e benefícios apresentados. Entretanto, no Brasil e em outros países, a cesárea tem sido utilizada de maneira indiscriminada, sem benefícios para as mulheres e recém-nascidos (Shearer, 1993).

Desde 1985, a comunidade médica internacional recomenda que a taxa de realização de cesárea esteja entre 10% e 15%. No entanto, a quantidade de procedimentos de cesarianas tem crescido muito, de modo que as altas taxas já são uma realidade mundial, tornando-se um grave problema de saúde pública.

O problema é ainda mais grave no Brasil, onde as cesarianas ultrapassam 55% do total de partos realizados no país. Ao analisar somente os números da saúde suplementar, isto é, mulheres que realizaram parto na rede privada, essas taxas superam 80% do número total de partos realizados (Betrán et al., 2016). Os países com maior taxa de cesárea por região do mundo são: Brasil (55,6%) e República Dominicana (56,4%) na América Latina e Caribe, Egito (51,8%) na África, Irã e Turquia na Ásia (47,9% e 47,5%, respectivamente), Itália (38,1%) na Europa, Estados Unidos (32,8 %) na América do Norte e Nova Zelândia (33,4%) na Oceania (Betrán et al., 2016).

As taxas de cesariana são mais elevadas em situações de alto risco, isto é, em condições nas quais a realização de um parto vaginal poderia trazer grandes riscos para o bebê e para a mãe. Sendo assim, o grande foco está na diminuição da taxa de cesáreas em casos de baixo risco (Armstrong et al., 2016).

O médico e o tipo de hospital também são fatores associados ao parto cesárea. Há uma grande tendência ao aumento de partos operatórios em hospitais de ensino e privados. Do mesmo modo, parece existir uma relação entre os hábitos dos médicos e essa indicação, sobretudo o agendamento de cesáreas eletivas, nas quais as mulheres decidem pela realização do procedimento (Haddad e Cececatti, 2011).

### **2.2. Modelos de classificação de cesarianas**

No decorrer das últimas três décadas, novas informações surgiram sobre os benefícios e riscos da cesárea, pois houve uma melhora significativa em relação aos cuidados obstétricos, e grandes avanços nos métodos de avaliação das evidências para que novas recomendações sejam formuladas. Diante disso, a comunidade médica internacional pediu que as recomendações, formuladas em 1985, fossem revistas (Organização Mundial de Saúde, 2015). Porém, definir a taxa ideal de cesarianas em termos populacionais, ou seja, cesáreas com indicações médicas que, ao mesmo tempo evitem cirurgias desnecessárias, ainda é um grande desafio.

Segundo a declaração da OMS, a falta de um sistema de classificação confiável e internacionalmente aceito, capaz de fornecer dados de forma padronizada, é um dos fatores que dificultam a definição da taxa ideal de cesáreas. Sistemas de classificação são amplamente usados na medicina para transformar dados brutos em informação útil. Esses sistemas são baseados na identificação de diferentes conceitos e regras que, quando arranjados, podem resultar em grupos específicos ou categorias que compartilham características entre si (Robson et al., 2013).

Diante das necessidades de se ter um sistema de classificação de cesáreas universal, capaz de prover uma forma de monitoramento dessas taxas de maneira eficaz, é sugerido o uso do classificador de Robson para este propósito de acordo com os resultados da revisão sistemática realizada por Torloni et al. (2011) e orientações da Organização Mundial de Saúde (2015).

O classificador de Robson é um sistema de classificação de cesáreas, baseado nas características maternas, que permite comparar as taxas de cesáreas entre diferentes hospitais, cidades, países ou regiões, de maneira que os dados possam ser utilizados para promover ações relevantes. Ele classifica todas as gestantes em 1 entre 10 grupos mutuamente exclusivos e totalmente inclusivos, conforme mostra a Tabela 1 (Robson, 2001). Os grupos são criados levando em consideração cinco características obstétricas: (1) Paridade (nulípara ou múltipara com e sem cesárea anterior); (2) Início do parto (espontâneo, induzido ou cesárea antes do início do trabalho de parto); (3) Idade gestacional (pré-termo ou termo); (4) Apresentação/situação fetal (cefálica, pélvica ou transversa; e (5) Número de fetos (único ou múltiplo).

**Tabela 1. Grupos da classificação de Robson e sua evitabilidade para cesárea**

Grupo	Características clínicas de cada grupo
Grupo 1	Nulíparas, gestação única, apresentação cefálica, $\geq 37$ semanas, trabalho de parto espontâneo;
Grupo 2	Nulíparas, gestação única, apresentação cefálica, $\geq 37$ semanas, indução ou cesárea antes do trabalho de parto;
Grupo 3	Múltiparas (sem cesárea anterior), gestação única, apresentação cefálica, $\geq 37$ semanas, trabalho de parto espontâneo;
Grupo 4	Múltiparas (sem cesárea anterior), gestação única, apresentação cefálica, $\geq 37$ semanas, indução ou cesárea antes do trabalho de parto;
Grupo 5	Com cesárea anterior, gestação única, apresentação cefálica, $\geq 37$ semanas;
Grupo 6	Todas apresentações pélvicas em nulíparas;
Grupo 7	Todas apresentações pélvicas em múltiparas (incluindo cesárea anterior);
Grupo 8	Todas gestações múltiplas (incluindo cesárea anterior);
Grupo 9	Todas posições anormais (incluindo cesárea anterior);
Grupo 10	Todas gestações únicas, apresentação cefálica, $\leq 36$ semanas (incluindo cesárea anterior).

Gaspar et al. (2017) realizaram um estudo no qual dividiram os grupos de Robson em três novos grupos baseados no grau de evitabilidade de cesáreas: pouco evitável, evitável, muito evitável. No primeiro estão contidas as mulheres presentes nos grupos 6 e 9 de Robson. No segundo estão contidas as dos grupos 2, 5, 7, 8 e 10. Já no terceiro, as dos grupos 1, 3 e 4.

### 3. Trabalhos relacionados

O estudo realizado por Gharehchopogh et al. (2012) explora a utilização da mineração de dados clínicos para determinar quando uma cesárea deve ser realizada. Foi gerado um modelo baseado em árvore de decisão, que obteve resultados de predições corretas de mais de 86,25%. As principais características utilizadas para a geração do modelo foram: idade materna, pressão sanguínea e problema de coração materno. O estudo encontrou

uma importante relação entre os problemas cardíacos, pressão sanguínea e o tipo de parto, e, como trabalhos futuros, propõem a inserção de novos atributos graves.

Robu e Holban (2015) utilizaram de modelos baseados em vários algoritmos de classificação e geração de regras: *Naive Bayes*, *J48*, *k-Nearest Neighbour*, *Random Forest*, *Support Vector Machines*, *AdaBoost*, *LogitBoost*, *JRipp*, *REPTree*, *SimpleCart*. O principal objetivo do trabalho era criar um modelo capaz de estimar o *Apgar score*<sup>1</sup>, baseado nos dados maternos. Utilizou-se de dados como: idade gestacional, estática fetal, quantidade de partos anteriores e episiotomia. O melhor modelo foi obtido utilizando-se o algoritmo *LogitBoost*, alcançando sensibilidade de 80,24% de um total de 2086 instâncias.

Sims et al. (2000) objetivaram determinar se o uso de algoritmos baseados em árvore de decisão poderiam ser usados para a previsão e determinação de parto cesáreo. O modelo foi criado considerando-se 78 variáveis, de uma divisão de 50% para treino e os outros 50% para validação. Os autores conseguiram bons resultados quando comparados à regressão logística, possuindo a vantagem de serem bem legíveis aos médicos, pelo formato de apresentação da árvore. As características relevantes no modelo de árvore de decisão foram: estática fetal, idade materna e sofrimento fetal.

Sodsee (2014) propôs uma modificação do algoritmo K-NN, com o objetivo de prever os riscos sobre realização de cesarianas devido à desproporção cefalopélvica, isto é, quando a pelve materna não permite a passagem da cabeça fetal. O algoritmo desenvolvido, intitulado CPD-NN, identificou as seguintes características relevantes: idade materna, quantidade de partos anteriores e índice de massa corporal.

Métodos estatísticos são muito utilizados na medicina quando o objetivo é realizar predições e classificações, como objetiva este trabalho. Dessa forma, muitos trabalhos encontrados utilizaram-se da regressão logística em seus modelos. Mardy et al. (2016) desenvolveram um modelo para a predição de partos vaginais após cesáreas prévias em mulheres em trabalho de parto durante um período de prematuridade, isto é, idade gestacional menor que trinta e sete semanas. Utilizando-se da regressão logística, foi desenvolvido um modelo preditivo, com uma divisão dos dados escolhidos de forma aleatória com 70% para treinamento e 30% para validação. As características relevantes para o modelo foram: hipertensão, diabetes, quantidade de partos vaginais anteriores e admissão obstétrica.

#### **4. Metodologia**

Esta seção apresenta a metodologia de pesquisa realizada buscando alcançar os objetivos propostos neste trabalho. Realizou-se o pré-processamento da base de dados, onde foram realizadas seleção de características, instâncias e codificação de atributos. Após o pré-processamento, gerou-se o modelo de classificação utilizando o algoritmo de árvore de decisão C4.5. A utilização do método de árvore de decisão foi definida pelo fato de ser um trabalho no campo da saúde, de modo a facilitar a visualização dos resultados e causas geradas pelas regras da árvore. Dessa forma, seria possível entender quais as características das mulheres que realizaram parto a fim de encontrar uma relação de causa e efeito. Finalizando a geração do modelo, os resultados e métricas foram analisados.

---

<sup>1</sup>Consiste na avaliação por um pediatra de 5 sinais vitais objetivos do recém-nascido, atribuindo-se a cada um dos sinais uma pontuação de 0 a 2

#### 4.1. Descrição do Sistema SISMater<sup>©</sup>

Diante dos desafios da informatização dos registros clínicos associados ao nascimento, uma equipe multidisciplinar da escola de medicina do HC-UFMG idealizou e implantou o sistema Sismater<sup>©</sup>. Esse software foi desenvolvido objetivando a coleta e registro dos dados clínicos, e geração de indicadores de qualidade da assistência materno-neonatal na Maternidade do Hospital das Clínicas. A utilização do sistema pelos profissionais de saúde possibilitou a geração de um banco de dados sobre o cuidado obstétrico hospitalar, atualmente com mais de 300 variáveis, contemplando informações epidemiológicas, administrativas e clínicas, possibilitando a emissão de diversos tipos de relatórios e alertas.

Este estudo faz parte de um projeto da Faculdade de Medicina da UFMG que visa a análise e desenvolvimento de indicadores obstétricos e neonatais, utilizando-se de dados anonimizados dos registros eletrônicos da maternidade. Tais indicadores são utilizados para a melhoria do cuidado da saúde materno-infantil, de modo que a direção do hospital possa tomar providências necessárias de acordo com o que é indicado. Assim, este projeto, no qual este estudo faz parte, foi submetido ao Conselho de Ética em Pesquisa da Instituição e do Hospital, e tem aprovação do Comitê de Ética em Pesquisa.

O Sistema possui dados de várias fases da história clínica da mulher, e do recém nascido, desde dados colhidos no pré-natal, até os dados da alta materna e do bebê. Sendo assim, os dados podem ser categorizados em três grupos, baseados no período em que foram coletados: pré-parto, durante o parto e pós-parto.

#### 4.2. Pré-processamento da base de dados

Inicialmente a base de dados possuía um total de 595 atributos e 7599 instâncias. Numa análise inicial, percebeu-se que havia muitos dados faltantes e em determinados atributos o valor se repetia em praticamente todas as instâncias, como alguns fatores de risco gestacional, algumas complicações maternas como eclâmpsia e *nearmiss*<sup>2</sup>, se houve gravidez ectópica<sup>3</sup>, e complicações no parto como cabeça derradeira<sup>4</sup> e prolapso de cordão<sup>5</sup>. Nestes casos estes atributos foram desconsiderados, uma vez que não seriam relevantes para um processo de aprendizado de máquina.

O objetivo deste trabalho foi gerar um modelo capaz de identificar quais características maternas têm influenciado na definição de qual via de parto escolher. Diante disto, os dados a serem considerados devem ser do período pré-parto. Desta forma, pelo fato de que a base também possui dados que foram colhidos em período posterior ao parto, ou seja, consequência do parto, os mesmos foram desconsiderados.

A base possui alguns atributos de campo aberto, como bairro, cidade e endereço. Tais campos são preenchidos manualmente pelo usuário, de modo que foi necessário normalizar esses dados para que fosse possível extrair informação dos mesmos. Desta forma, foi desenvolvido um *script* utilizando a linguagem de programação *Python*, através da *Application Programming Interface* (API) CorreiosAPI<sup>6</sup>, que retorna o endereço completo

<sup>2</sup>[http://www.who.int/reproductivehealth/topics/maternal\\_perinatal/nmconcept/en/](http://www.who.int/reproductivehealth/topics/maternal_perinatal/nmconcept/en/)

<sup>3</sup>Gestação em que o óvulo fertilizado é implantado fora do útero

<sup>4</sup>Nome dado quando o corpo do bebê que estava pélvico sai e a cabeça fica presa

<sup>5</sup>Quando, durante o parto, o cordão fica comprimido pela cabeça do bebê o que diminui ou impede o normal fornecimento de sangue e oxigênio

<sup>6</sup>Disponível em: <https://correiosapi.apphb.com/>

através de uma consulta HTTP GET passando o CEP como parâmetro na URL. Também foram desenvolvidos *script* em *Python* para calcular idade através da data de nascimento, e transformar todas as variáveis numéricas para nominal.

O Sismater<sup>©</sup> também possui dados de mulheres que foram submetidas a internações anteparto e pós-parto, isto é, internações que não geraram parto. Desta forma, estas instâncias foram desconsideradas, pelo fato de que elas não teriam a variável de classe, tipo de parto, preenchida. Além disso, todas as instâncias que também não possuíam a classe preenchida foram desconsideradas.

Há no sistema um grupo de trinta e quatro variáveis chamado “Fatores de Risco Gestacional”. Estas variáveis são todas dicotômicas, assumindo *TRUE* ou *FALSE*, e, em grande parte delas, o preenchimento era de mais de 90% para *FALSE*. Entretanto, é uma variável que poderia ter um peso muito grande na classe, devido ao fato de representarem os riscos gestacionais. Desta forma, criou-se uma variável contendo a quantidade de *TRUE* que a instância teve nesse grupo de trinta e quatro variáveis. Além disso, o sistema já possuía uma variável que representa gestantes de alto risco, que era preenchida automaticamente caso a gestante possuísse pelo menos um fator de risco. Sendo assim, as trinta e quatro variáveis foram agrupadas em apenas duas, que poderiam trazer mais informação do que todas as outras.

Para a geração do modelo e pré-processamento utilizou-se o software WEKA (Hall et al., 2009), que possui uma coleção de algoritmos de aprendizado de máquina para serem utilizados em tarefas de mineração de dados. Ele oferece ferramentas para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização de dados.

Para a seleção de instâncias, inicialmente utilizou-se o método *Wrapper* (Olvera-López et al., 2010). Através do *RemoveMisclassified*<sup>7</sup>, filtro disponibilizado pelo software WEKA, foi possível conhecer todas as instâncias que não se encaixavam no modelo. Por ser um filtro de *Wrapper*, ele filtra as instâncias baseadas no algoritmo de classificação escolhido. Basicamente ele gera um modelo com o conjunto de dados completo, após isto, analisando a matriz de confusão, ele remove todas as instâncias que o modelo classificou de forma errada. Desta forma, ao executar o filtro foi possível analisar, junto a profissionais da medicina, o perfil das 778 gestantes que foram removidas.

Das 778 gestantes removidas pelo filtro *RemoveMisclassified*, 683 realizaram parto vaginal. Este fato vai de acordo com os objetivos deste trabalho, diante das dificuldades de se encontrar uma regra exata para a escolha da via de parto.

Nesta análise, foi possível perceber que grande parte deste subgrupo era de mulheres que realizaram parto cesárea e se encaixavam nos grupos 1, 2 ou 4 do classificador de Robson. Gaspar et al. (2017) classificaram estes três grupos como sendo de alta evitabilidade para procedimentos de cesariana, de modo que o grupo de gestantes encontrado seriam cesarianas realizadas por motivos indevidos segundo a literatura médica. No restante das gestantes, identificou-se um subgrupo de mulheres que tiveram trabalho de parto induzido, mas realizaram cesariana. Estes dois subgrupos encontrados revelam 560 instâncias que possuem muitas características de uma mulher que deveria ter realizado parto vaginal, mas por circunstâncias realizou cesárea.

<sup>7</sup><https://weka.wikispaces.com/Removing+misclassified+instances+from+dataset>

De acordo com as indicações de cesáreas já conhecidas na literatura, e baseando-se na classificação de Robson, foram identificadas 327 mulheres, dos subgrupos encontrados, que realizaram cesárea mesmo que fosse altamente evitável. Estas mulheres pertencem aos grupos 1, 2 ou 4 do classificador de Robson, possuindo características como: gestação única, estática fetal cefálica, idade gestacional maior ou igual a 37 semanas, e cesáreas prévias. Algumas delas já tiveram outros partos vaginais anteriores, enquanto outras não tiveram, mas foram admitidas com trabalho de parto espontâneo.

Utilizando-se de um método de seleção de instâncias *Filter*, foi possível remover um número menor de instâncias em relação ao filtro *RemoveMisclassified* disponibilizado pelo WEKA, totalizando 560 remoções.

Assim, após as remoções e todo o processamento da base, restaram 27 atributos e 5797 instâncias, uma redução de 95,4% e 26,4% no valor total e atributos e instâncias da base de dados, respectivamente.

Após o pré-processamento dos dados, a base obtida ficou pronta para a geração de um modelo de classificação baseado em árvore de decisão. O modelo foi criado utilizando-se o algoritmo J48<sup>8</sup>, implementação do C4.5 na linguagem de programação JAVA, disponível no software WEKA.

### 4.3. Validação e métricas de avaliação

A avaliação dos resultados foi realizada através das métricas de precisão ( $P = \frac{VP}{VP+FP}$ ), sensibilidade ( $S = \frac{VP}{VP+FN}$ ) e *F-measure* ( $F - measure = \frac{2*VP}{2*VP+FP+FN}$ ). Precisão mede a quantidade de instâncias corretamente classificadas em uma classe, dentre todas que foram classificadas naquela classe. Sensibilidade mede a quantidade de instâncias corretamente classificadas dentre todas as instâncias da classe. A *F-measure* representa a média harmônica da precisão e sensibilidade. VP é o valor de Verdadeiros Positivos, VN corresponde aos Verdadeiros Negativos, FN é a quantidade de Falsos Negativos e FP representa a quantidade de Falsos Positivos.

A amostragem dos dados de treinamento e teste foi realizada utilizando-se o método validação cruzada de 10 dobras.

## 5. Resultados e discussões

Esta seção apresenta a discussão dos resultados, contendo a análise das regras obtidas a partir da execução do algoritmo de árvore de decisão.

A Tabela 2 apresenta os resultados das métricas de avaliação da qualidade do modelo obtido por esta árvore de decisão.

**Tabela 2. Resultados obtidos pela árvore de decisão**

Classe	Precisão	Sensibilidade	<i>F-measure</i>
Vaginal	95%	97,4%	96,2%
Cesárea	93,6%	88,1%	90,8%

Analisando os dados, observa-se bons resultados, o que valida a boa qualidade do conjunto de dados pré-processados. A alta precisão obtida, demonstra que houve um

<sup>8</sup><http://weka.sourceforge.net/doc.stable/weka/classifiers/trees/J48.html>

baixo número de falsos positivos, já uma sensibilidade alta está ligada diretamente ao baixo número de falsos negativos. Percebe-se que a sensibilidade para a classe de cesárea está menor que as demais. Isto pode ser explicado justamente pela falta de critério válido para a realização deste tipo de procedimento, tornando difícil a tarefa de extrair um perfil de uma mulher que tenha suas características totalmente pertencentes a esse grupo de estudo.

Entretanto, percebe-se que a grande maioria das instâncias foram classificadas corretamente em suas classes originais, já que somente 104 instâncias da classe vaginal foram classificadas como cesárea, e 207 da classe cesárea foram classificadas como sendo vaginal.

Os principais atributos encontrados pela árvore que compõem as regras são apresentados na Tabela 3. A Tabela 4 mostra três regras principais geradas para a classificação cesárea e vaginal, bem como a quantidade de instâncias classificadas corretamente por cada regra.

**Tabela 3. Atributos identificados como importantes na caracterização do perfil da mulher que faz parto cesária e vaginal**

Atributos
Admissão obstétrica
Estática fetal
Quantidade de cesáreas prévias
Semanas da Idade Gestacional
Quantidade de Recém Nascidos

**Tabela 4. Principais regras geradas pela árvore de decisão**

<b>Regra 1:</b> SE Admissão obstétrica = Sem trabalho de parto ENTÃO <b>Cesárea</b> (1412 instâncias)
<b>Regra 2:</b> SE Admissão obstétrica = Trabalho de parto induzido ENTÃO <b>Vaginal</b> (1137 instâncias)
<b>Regra 3:</b> SE Admissão obstétrica = Com trabalho de parto E Estática fetal = Cefálica E Quantidade de cesáreas prévias $\leq 2$ ENTÃO <b>Vaginal</b> (3008 instâncias)
Quantidade de partos do tipo cesária e vaginal: 1605, 4192, respectivamente.

Observando-se os dados, percebe-se que o atributo que corresponde à admissão obstétrica mostrou-se como a mais relevante sendo classificados como cesárea as mães admitidas sem trabalho de parto, e como vaginal para as mães que tiveram parto induzido (regras 1 e 2, respectivamente). Somente estas regras foram capazes de classificar 43,97% da base de dados. Estas regras são facilmente explicadas, visto que mulheres admitidas no hospital não estando em trabalho de parto possuem alta chance de realização de parto cesárea (Mikolajczyk et al., 2016), sendo que essas admissões possivelmente foram previamente programadas.

A Regra 3 também foi muito relevante, onde se avaliou a admissão obstétrica em casos de mulheres admitidas com trabalho de parto, estática fetal pélvica, idade gestacional menor ou igual a 31 semanas e uma gestação com dois bebês ou menos. Esta regra sozinha foi capaz de classificar 48,33% da base de dados. Este resultado é interessante, pois mulheres com cesáreas prévias comumente são indicadas a operações de cesariana. Entretanto, a literatura mostra sucesso em casos de mulheres que realizam parto vaginal após duas cesáreas prévias, com taxa de sucesso de acima de 70% (Tahseen e Griffiths, 2009).

Outra característica muito importante é a estática fetal. Quando cefálica, gera grande chance de parto vaginal, conforme a Regra 3. Esta característica está presente no classificador dos 10 grupos de Robson, e é um fator importante na escolha do tipo de parto. Quando transversal ou oblíqua a cesárea é pouco evitável. Porém, em casos de apresentação cefálica em bebês prematuros, a cesárea é evitável (Gaspar et al., 2017).

O modelo obtido errou na classificação de 311 instâncias. Aproximadamente 30% destas eram da classe vaginal, e as 70% restantes da classe cesárea. Analisando as instâncias, foi possível observar que a grande maioria delas fazia parte dos grupos 2, 5, 7, 8 ou 10, do classificador de Robson, de modo que seriam casos de cesárea com nível médio de evitabilidade, de acordo com os estudos de Gaspar et al. (2017).

Das instâncias classificadas erroneamente, as mulheres que foram classificadas como parto vaginal apresentaram características de mulheres que normalmente realizam cesárea. Contrariamente ao modelo gerado, as mulheres deste subgrupo foram admitidas sem trabalho de parto em internações de caráter eletivo, isto é, uma internação marcada e que não é de urgência. Este subgrupo também contém mulheres que foram realizaram mais de duas cesáreas previamente.

As instâncias classificadas de maneira errada como cesárea, também apresentaram características contrárias à árvore encontrada. Mulheres com menos de duas cesáreas anteriores, com apresentação fetal cefálica realizaram parto cesárea; assim como as que tiveram apresentação cefálica pélvica e idade gestacional menor que 32 semanas com gestação única, ou seja, não gemelar (regras não mostradas).

## **6. Considerações finais e trabalhos futuros**

Neste trabalho é apresentado um panorama sobre a importância da relação entre a tecnologia no cuidado à saúde da mulher, de modo que é apresentado o estado atual da computação, mais especificamente da mineração de dados, aplicada aos estudos da perinatologia, de modo a auxiliar os profissionais e instituições de saúde no cuidado materno-infantil.

A partir dos resultados obtidos, foi possível observar que os fatores que mais influenciaram na decisão da via de parto foram a admissão obstétrica, estática fetal, idade gestacional e a quantidade de cesáreas prévias. Estas, são características já ditas na literatura, estando no contexto de indicações de cesarianas, estando, grande parte delas, presente no classificador de Robson. Observa-se nos trabalhos relacionados que são características comumente presentes na determinação da via de parto. Com o modelo de árvore gerado, torna-se possível tomar uma ação direta nos casos em que a cesárea é de fato evitável, diante das evidências apresentadas.

Dentre as instâncias que o classificador não conseguiu classificar de forma correta, foi possível enxergar casos de cesáreas realizadas em que fugiram do padrão previamente exercido. Os fatores que foram mais determinantes, e causaram estes erros no classificador, foram a quantidade de cesáreas prévias e o número de gestações anteriores. Outro atributo importante na decisão foi a admissão obstétrica, no caso que mulheres admitidas sem trabalho de parto realizaram parto vaginal, quebrando a Regra 1 da árvore gerada. Estas três características fizeram com que o classificador se comportasse de maneira errada, devido ao diferente perfil destas gestantes em relação às demais.

Sabe-se que o cuidado obstétrico pelo SUS é assegurado a todas as gestantes através de uma Política Pública, denominada Rede Cegonha, criada em 2011. Nesta estratégia, ações para ampliação e humanização da assistência à mulher vem sendo desenvolvidas em todo o território nacional e visam, sempre, a redução das taxas de morbimortalidade materno-infantil (Ministério da Saúde, 2011). Além disso, dois dos oito Objetivos de Desenvolvimento do Milênio, estabelecidos pela Organização das Nações Unidas, envolvem a saúde materno-infantil, sendo o quinto objetivo melhorar a saúde materna. Desta forma, este trabalho contribui com o objetivo descrito, de modo a entender o que tem levado as mulheres a realização de parto cesárea, a partir das principais características geradas pelo modelo de classificação baseado em árvore de decisão.

Como proposta de trabalhos futuros, sugere-se a análise dos dados de outros anos, não considerados no modelo criado. Além disso, sugere-se a realização do modelo em outros hospitais universitários, a fim de comparar as taxas e avaliar o comportamento do padrão encontrado, se é condizente com o encontrado. Outra proposta é a geração de outros modelos baseados em outros algoritmos de classificação, como redes neurais e análises de regressão, ou até mesmo modelos de clusterização, podendo encontrar sub-grupos dentre as gestantes que possuem características em comum que possam explicar a decisão da via de parto.

## Referências

- C. J. Sims, L. Meyn, R. Caruana, R. Rao, T. Mitchell, e M. Krohn. Predicting cesarean delivery with decision tree models. *American Journal of Obstetrics and Gynecology*, 183(5):1198–1206, nov 2000. doi: 10.1067/mob.2000.108891.
- V. Podgorelec, P. Kokol, B. Stiglic, e I. Rozman. Decision trees: An overview and their use in medicine. 26:445–63, 11 2002.
- G. P. Barbosa, K. Giffin, A. Angulo-Tuesta, A. de Souza Gama, D. Chor, E. D’Orsi, e A. C. G. V. dos Reis. Parto cesáreo: quem o deseja? em quais circunstâncias? *Cadernos de Saúde Pública*, 19(6):1611–1620, dec 2003. doi: 10.1590/s0102-311x2003000600006.
- H. Koh e G. Tan. Data mining applications in healthcare. *Journal of healthcare information management : JHIM*, 19:64–72, 02 2005.
- M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, e I. H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009. ISSN 1931-0145. doi: 10.1145/1656274.1656278.
- S. Tahseen e M. Griffiths. Vaginal birth after two caesarean sections (VBAC-2)-a systematic review with meta-analysis of success rate and adverse outcomes of VBAC-2 versus VBAC-1 and repeat (third) caesarean sections. *BJOG: An International Journal of Obstetrics & Gynaecology*, 117(1):5–19, sep 2009. doi: 10.1111/j.1471-0528.2009.02351.x.
- J. A. Olvera-López, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, e J. Kittler. A review of instance selection methods. *Artificial Intelligence Review*, 34(2):133–143, may 2010. doi: 10.1007/s10462-010-9165-y.
- S. E. M. T. Haddad e J. G. Cececatti. Estratégias dirigidas aos profissionais para a redução das cesáreas desnecessárias no Brasil. *Revista Brasileira de Ginecologia e Obstetrícia*, 33:252 – 262, 05 2011. ISSN 0100-7203.

- M. R. Torloni, A. P. Betran, J. P. Souza, M. Widmer, T. Allen, M. Gulmezoglu, e M. Meriardi. Classifications for Cesarean Section: A Systematic Review. *PLoS ONE*, 6(1): e14566, jan 2011. doi: 10.1371/journal.pone.0014566.
- F. S. Gharehchopogh, P. Mohammadi, e P. Hakimi. Article: Application of decision tree algorithm for data mining in healthcare operations: A case study. *International Journal of Computer Applications*, 52(6):21–26, August 2012.
- M. Robson, L. Hartigan, e M. Murphy. Methods of achieving and maintaining an appropriate caesarean section rate. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 27(2):297–308, apr 2013. doi: 10.1016/j.bpobgyn.2012.09.004.
- J. Gaspar, J. Chagas, R. Cruz-Correia, e Z. Reis. Health information system for obstetric and neonatal healthcare monitoring: Sismater. pages 1–7, Lisboa, PT, 01 2013. 8th Iberian Conference on Information Systems and Technologies (CISTI).
- R. Robu e S. Holban. The analysis and classification of birth data. *Acta Polytechnica Hungarica*, 12(4), 2015.
- J. C. Armstrong, K. B. Kozhimannil, P. McDermott, G. R. Saade, e S. K. Srinivas. Comparing variation in hospital rates of cesarean delivery among low-risk women using 3 different measures. *American Journal of Obstetrics and Gynecology*, 214(2):153–163, feb 2016. doi: 10.1016/j.ajog.2015.10.935.
- A. P. Betrán, J. Ye, A.-B. Moller, J. Zhang, A. M. Gülmezoglu, e M. R. Torloni. The increasing trend in caesarean section rates: Global, regional and national estimates: 1990–2014. *PLOS ONE*, 11(2):e0148343, feb 2016. doi: 10.1371/journal.pone.0148343.
- A. H. Mardy, C. V. Ananth, W. A. Grobman, e C. Gyamfi-Bannerman. A prediction model of vaginal birth after cesarean in the preterm period. *American Journal of Obstetrics and Gynecology*, 215(4):513.e1–513.e7, oct 2016. doi: 10.1016/j.ajog.2016.05.039.
- R. T. Mikolajczyk, J. Zhang, J. Grewal, L. C. Chan, A. Petersen, e M. M. Gross. Early versus late admission to labor affects labor progression and risk of cesarean section in nulliparous women. *Frontiers in Medicine*, 3, jun 2016. doi: 10.3389/fmed.2016.00026.
- J. Gaspar, Z. Reis, e J. Barra. Is a decision support system based on robson’s classification enough to reduce cesarean section rates? In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, jun 2017. doi: 10.1109/cbms.2017.143.
- B. Ministério da Saúde. SAÚDE. portaria nº 1.459, de 24 de junho de 2011: Institui, no âmbito do sistema Único de saúde - sus - a rede cegonha., jun 2011. URL [http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2011/prt1459\\_24\\_06\\_2011.html](http://bvsmms.saude.gov.br/bvs/saudelegis/gm/2011/prt1459_24_06_2011.html).
- Organização Mundial de Saúde. Declaração sobre taxas de cesáreas. [http://apps.who.int/iris/bitstream/10665/161442/3/WHO\\_RHR\\_15.02\\_por.pdf](http://apps.who.int/iris/bitstream/10665/161442/3/WHO_RHR_15.02_por.pdf), 2015. [Online; Acessado em 03 de Dezembro de 2017].
- M. Robson. Classification of caesarean sections. *Fetal and Maternal Medicine Review*, 12(01), jan 2001. doi: 10.1017/s0965539501000122.
- E. L. Shearer. Cesarean section: Medical benefits and costs. *Social Science & Medicine*, 37(10):1223–1231, nov 1993. doi: 10.1016/0277-9536(93)90334-z.
- S. Sodsee. Predicting caesarean section by applying nearest neighbor analysis. *Procedia Computer Science*, 31:5–14, 2014. doi: 10.1016/j.procs.2014.05.239.