

Classification of Newborns with Congenital Syndrome Associated with Zika Virus Infection Using Machine Learning

Érika G. de Assis, Luis E. Zárate, Cristiane N. Nobre

Pontifícia Universidade Católica de Minas Gerais
egassis@sga.pucminas.br, zarate@pucminas.br, nobre@pucminas.br

Abstract. Due to evidence that Zika virus (ZIKV) infection during pregnancy caused congenital brain anomalies, including microcephaly, in 2016 the WHO declared this disease a worldwide public health problem. The objective of this work is to identify the most important characteristics for the diagnosis of children with congenital syndrome due to ZIKV virus infection. We applied machine learning algorithms to RESP-Microcephaly, a database from the Brazilian Ministry of Health that records suspected cases of congenital abnormalities. At the end of the process, the most relevant characteristics were: weight, age of the pregnant woman, length, head circumference and region where the mother lives. This information is very significant as it is in agreement with the literature that associates these attributes with critical factors for the occurrence of congenital infection.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: Zika Vírus, Congenital Syndrome, Machine Learning

1. INTRODUÇÃO

Em 31 de março de 2016, a Organização Mundial de Saúde (OMS) anunciou a infecção pelo vírus Zika (ZIKV) como um problema emergencial de saúde pública mundial devido a associação deste arbovírus com a ocorrência da síndrome congênita do Zika. Neste contexto, as crianças nascidas de mulheres infectadas pelo ZIKV durante a gravidez apresentaram vários graus de comprometimento do sistema nervoso, como a microcefalia e outros distúrbios [Nunes et al. 2016].

As microcefalias congênitas são descritas como anomalias neurológicas, em que o encéfalo não se desenvolve adequadamente e o perímetro cefálico (PC) aferido no recém-nascido mostra-se inferior ao esperado para a idade gestacional e o sexo correspondente [Ribeiro et al. 2018]. Pode estar associada a vários fatores ambientais, genéticos, sofrimento fetal e infecções congênitas por STORCH¹. Além disto, algumas síndromes genéticas também causam microcefalia, tais como as síndromes de Seckel, Smith-Lemli-Optiz, Apert, Crouzon, Saethre-Chotzen, Pfeiffer e Carpenterx [Ribeiro et al. 2018]. Dentre os fatores genéticos, citam-se as alterações monogênicas (mendelianas), anomalias cromossômicas ou distúrbios multifatoriais (poligênicos), além de exposição intrauterina à radiação ionizante, abuso de álcool e/ou drogas, e fenilcetonúria materna (fenilalanina em excesso) [Ribeiro et al. 2018].

Segundo [Nunes et al. 2016], o ZIKV infecta e lesiona as células progenitoras neurais, o que pode explicar sua capacidade de prejudicar o desenvolvimento do cérebro fetal e causar microcefalia e outras lesões do neurodesenvolvimento.

¹O acrônimo é composto pelos patógenos mais frequentemente relacionados às infecções: bactéria *Treponema Pallidum* que causa a sífilis (S), o protozoário *Toxoplasma Gondii* que causa a toxoplasmose (TO) e os vírus da rubéola (R), citomegalovírus (C), vírus herpes simples (H) [Ribeiro et al. 2018].

Com estudos observacionais adicionais, um conjunto de anomalias congênitas foi identificado e ligado à infecção do ZIKV no útero. Esse conjunto de anormalidades foi denominado Síndrome Congênita do Zika (CZS) que inclui, além da microcefalia, desproporção craniofacial, irritabilidade, espasticidade, convulsões, dificuldades de alimentação, anormalidades oculares e perda auditiva, bem como calcificações, distúrbios corticais e dilatação dos ventrículos cerebrais fetais na neuroimagem, denominado ventriculomegalia [Nunes et al. 2016].

Os métodos de identificação dessas alterações são realizados através de exames de imagem, tais como: a ultrassonografia, a tomografia computadorizada e a ressonância magnética [Ribeiro et al. 2018]. Os recém-nascidos com suspeição de microcefalia são submetidos a exame físico e medição do perímetro cefálico, além de exames neurológicos e de imagem. A ultrassonografia transfontanela é a primeira opção indicada, e a tomografia, quando a moleira estiver fechada [Brasil et al. 2017].

O Ministério da Saúde, seguindo recomendação da Organização Mundial de Saúde (OMS), adotou como referência para as primeiras 24-48h de vida os parâmetros de InterGrowth-21st². Nessa nova tabela de referência, para uma criança que nasceu com 37 semanas de gestação, a medida do perímetro cefálico (PC) será 30,24 cm para meninas e 30,54 cm para meninos. Contudo, é necessário que seja consultada a tabela para cada idade e sexo, sendo que a medida deve ser aferida com a maior precisão possível, de preferência com duas casas decimais [Brasil et al. 2017].

Diante do exposto, o objetivo deste artigo é identificar as características mais relevantes para o diagnóstico de crianças com síndrome congênita devido a infecção pelo ZIKV. A intenção é aprimorar o diagnóstico precoce, aumentando a prevenção e, conseqüentemente, diminuindo as conseqüências tão devastadoras para mães e filhos. Para isso, serão utilizados algoritmos de aprendizado interpretáveis como Árvore de Decisão, que fornece regras de classificação, e o Random Forest, que fornece a relevância dos atributos.

2. TRABALHOS RELACIONADOS

Nunes et al. [2016] revisaram a literatura disponível sobre o surto do vírus Zika no Brasil e sua possível associação com os casos de microcefalia. Os dados epidemiológicos sugerem uma associação temporal entre o aumento do número de notificações de microcefalia no Brasil e o surto do Zika vírus, principalmente no Nordeste do Brasil. Agentes virais comprovadamente podem ser causadores de microcefalia congênita adquirida. Os autores concluíram que o desenvolvimento de técnicas de diagnóstico que confirmem uma associação causa-efeito e estudos sobre a fisiopatologia do comprometimento do sistema nervoso central devem ser priorizados.

Em [Ribeiro et al. 2018] houve um estudo descritivo dos casos de microcefalia no estado do Piauí, durante a epidemia do vírus Zika em 2015-2016, com dados dos recém-nascidos, adquiridos no Sistema de Informações sobre Nascidos Vivos (SINASC), do Registro de Eventos em Saúde Pública (RESP) e em prontuários. Mães e recém-nascidos foram verificados para dengue, chikungunya e Zika, além de STORCH. Os autores observaram que dos 75 casos de microcefalia, 34 foram relacionados a processo infeccioso congênito; nenhum apresentou resultado positivo para STORCH, dengue ou chikungunya; 1 foi Zika IgM reagente. Exames de imagem confirmaram que 34 nascidos vivos apresentavam calcificações, 23 atrofia cerebral, 14 lisencefalia⁴, 12 ventriculomegalia⁵ e 6 disgenesias.

²O projeto INTERGROWTH-21st desenvolveu uma curva de crescimento fetal padrão para uso internacional. O objetivo principal foi estudar o crescimento, saúde, nutrição e desenvolvimento neuromotor desde 14 semanas de gestação até os dois anos de idade. A curva padrão do INTERGROWTH-21st complementa a curva de crescimento de crianças da OMS, para ambos os sexos³.

⁴Lisencefalia significa “cérebro liso”, ou seja, a falta de sulcos e reentrâncias que observamos em um cérebro normal. Este defeito é causado por uma falha na migração dos neurônios durante a formação do embrião e traz, como conseqüência, graves anomalias neurológicas [Nunes et al. 2016].

⁵A ventriculomegalia é definida como a dilatação dos ventrículos cerebrais fetais [Nunes et al. 2016].

Herber et al. [2019] realizaram um estudo transversal onde foram incluídas todas as notificações de microcefalia congênita no estado do Rio Grande do Sul no período de 1º de dezembro de 2015 a 31 de dezembro de 2016. Dos 153.744 nascidos vivos registrados no período do estudo, 148 casos foram notificados, mas apenas 58 foram confirmados para microcefalia. Destes, 50% estavam associados a STORCH, 15,5% a SNC isolado e 10,3% a síndromes genéticas. A síndrome do ZIKV congênito (CZS) com fenótipo típico foi diagnosticada em três casos (5,2% de todos os casos confirmados de microcefalia ou 10,4% de todas as infecções congênicas). A conclusão do estudo é que no Rio Grande do Sul, não foi registrado nenhum surto de infecção por ZIKV, e as infecções congênicas foram a principal causa de microcefalia.

Em [Prata-Barbosa et al. 2019], os autores revisaram os dados dos primeiros 2,5 anos de uma coorte de crianças expostas ao Zika vírus durante a gravidez. Foram relatados em crianças com síndrome congênita do Zika, tanto a restrição de crescimento intrauterino quanto o baixo peso ao nascer. O déficit de crescimento pós-natal dessas crianças parece ser diretamente proporcional ao grau de comprometimento neurológico. As causas são multivariadas e possivelmente fatores nutricionais estão envolvidos. Os dados indicam que, em crianças com Zika congênita, quanto maior o grau de comprometimento neurológico, maior o impacto sobre o crescimento, associado ou não à microcefalia no nascimento.

Almeida et al. [2019] estudaram todos os casos de microcefalia congênita confirmados no estado do Piauí entre 2015 e 2016, sendo $n=100$. As taxas de incidência nos meses de outubro, novembro e dezembro de 2015 atingiram 4.46, 6.33 e 3.86/1000 dos nascidos vivos. Foram registrados 44 casos na capital do estado. Os autores concluíram que o Estado do Piauí, como outros na região Nordeste, enfrentou, entre 2015 e 2016, uma epidemia de microcefalia congênita, supostamente relacionada à infecção congênita pelo vírus Zika, mais intensa na capital. Os autores relatam ainda que os desafios incluem melhora do controle de vetores, pesquisa básica, ampliação de ferramentas de diagnóstico para exame pré-natal do vírus Zika, vacinas e cuidados de saúde para crianças afetadas.

3. MATERIAIS E MÉTODOS

3.1 Visão Geral do RESP

O RESP-Microcefalia é um formulário online desenvolvido pelo DATASUS-Brasil, instituído pelo Ministério da Saúde (MS) desde 19 de novembro de 2015, com a intenção de registrar os casos e óbitos suspeitos de alterações no crescimento e desenvolvimento relacionado à infecção pelo vírus Zika e outras etiologias infecciosas [Brasil et al. 2017].

A base de dados foi fornecida pelo MS e não apresenta variáveis que permitem a identificação dos indivíduos e suas famílias. Os casos registrados no RESP com infecção congênita devido ao Zika Vírus teve seu pico em 2016 com mais de 1600 registros. A partir de maio de 2016 nota-se um queda no número de casos, comportamento verificado nos anos subsequentes.

A base de dados do RESP apresenta 43 atributos, organizados em nove (9) categorias:

- (1) *Notificação*: Apresenta a classificação dos casos suspeitos de infecção congênita (recém nascido, criança, feto em risco, aborto espontâneo ou natimorto) e a data em que foi notificado;
- (2) *Dados da gestante*: idade, raça/cor e estado de residência (UF);
- (3) *Informações sobre o nascido vivo*: sexo, data de nascimento, peso (gramas) e comprimento (centímetros);
- (4) *Dados sobre a gestação e parto*: tipos de alterações congênicas, quando a alteração foi detectada (na gravidez ou após o parto), idade gestacional de detecção da microcefalia, tipo de gravidez, classificação do nascido vivo, perímetro cefálico e data da medição do perímetro cefálico. O tipo de gravidez que pode ser definido em pré-termo (com idade gestacional menor que 37 semanas

de gestação), a termo (idade gestacional entre 37 e 41 semanas de gestação), pós-termo (idade gestacional maior que 42 semanas);

- (5) *Dados clínicos epidemiológicos da mãe*: data do início dos sintomas, o tipo de sintomas (febre, erupções cutâneas, coceira, conjuntivite, dor de cabeça e acometimento neurológico), realização de exame de STORCH e resultado, resultado exame Zika, histórico de arbovírus e mal formação congênita;
- (6) *Informações sobre os exames de imagem*: ultrassonografia, ultrassonografia transfontanela, tomografia computadorizada e ressonância magnética;
- (7) *Dados sobre o estabelecimento de saúde*: município e estado;
- (8) *Dados sobre a evolução da doença*: óbito e data de óbito;
- (9) *Campos de acesso restrito ao gestor*: Classificação final do caso suspeito de alterações Congênicas e Critério de confirmação através de exames laboratoriais realizados (Zika, Dengue, Chikungunya, STORCH, outros e imagem).

3.2 Pré-processamento da base de dados

Inicialmente, a base de dados possuía 17451 instâncias. Como o interesse do trabalho é estudar características de crianças nascidas com alterações congênicas devido a infecção pelo vírus Zika, trabalhamos com a notificações de recém nascidos e crianças⁶. Também foram excluídos todos os casos com confirmação laboratorial para sífilis ou toxoplasmose. Ao final do processo de seleção tínhamos 10.510 instâncias, sendo 2.904 crianças diagnosticadas com síndrome congênita devido ao Zika Vírus e 7.606 crianças sem alteração congênita. As etapas de pré-processamento foram as seguintes:

- (1) *Binarização de atributos*. Todos os atributos nominais não ordinais foram binarização.
- (2) *Seleção de atributos*: A correlação é uma medida padronizada da relação entre duas variáveis e indica a força e a direção do relacionamento linear entre duas variáveis. As variáveis diâmetro encefálico⁷ e o perímetro encefálico apresentam correlação muito forte. Mantivemos o perímetro encefálico, pois é um critério de confirmação de microcefalia adotado pela OMS. Também foram excluídos os atributos que estavam altamente correlacionados com o atributo de classificação, tais como os: resultados de exames de imagem, tipo de alteração congênita, quando foi detectada a alteração congênita e idade gestacional na detecção da microcefalia. Ao final do processo, a base de dados apresentava 38 atributos⁸.
- (3) *Dados inconsistentes*: Havia 95 instâncias em que a idade da gestante estava com os valores 2 e 3. Como se trata de uma idade incompatível fisiologicamente para uma concepção, excluímos estes valores e deixamos em branco.

Duas instâncias estavam com o valor do perímetro encefálico medindo 323,3 cm, o que não corresponde a um valor real. Vale ressaltar que os relatos da literatura referem média de perímetro cefálico em RN normais masculinos de 34,61 cm, com variação entre 32,14 e 37,08 cm, e em RN normais femininos média de 34,05 cm, com variação entre 31,58 e 36,52 cm [Brasil et al. 2017]. Como estes valores não correspondem a valores encontrados na literatura, excluímos estes valores, deixando-os ausentes.

- (4) *Dados ausentes*: Os dados ausentes são comuns em banco de dados na área de saúde. Por isso, o uso de métodos adequados torna-se fundamental para diminuir o impacto da ausência de informação. A base de dados original possuía cerca de 30% de dados ausentes. Realizamos tratamento de dados ausentes através de imputação de dados através da média/mediana.

⁶Estas crianças já possuíam microcefalia antes do surto da Zika e foram inseridas no RESP por orientação da Secretária de Saúde para fazer o acompanhamento destes casos [Brasil et al. 2017].

⁷A circunferência craniana é detectada de forma intrauterina através de exames de imagem.

⁸Uma descrição detalhada destes atributos está disponível em: <https://bitlybr.com/ZeU5ZKO>

- (5) *Balanceamento de Classes*: A base de dados possui um desbalanceamento de classes; ou seja, há a presença de uma classe majoritária (sem síndrome), com frequência muito maior que a classe minoritária (com síndrome).
Utilizamos o método *undersampling* ‘SpreadSubsample’ disponível na ferramenta Weka⁹ que realiza a subamostragem da classe majoritária. Após o balanceamento de dados, a base de dados ficou com 2904 para cada uma das classes consideradas (com e sem alteração congênita).
- (6) *Escolha dos algoritmos de aprendizado*: Neste trabalho foram utilizados os algoritmos de árvore de decisão C4.5 e Random Forest no ambiente WEKA. Todos os algoritmos foram executados com seus parâmetros default. A escolha dos métodos foi baseada na interpretabilidade da árvore e pelo Random Forest indicar os atributos mais relevantes no contexto da classificação.
- (7) *Criação dos modelos de aprendizado*: A partir da base de dados balanceada, foram reservadas 20% das instâncias para teste (Teste 1 da Tabela I). O conjunto de dados restante foi utilizado para criação dos modelos de aprendizado. Durante o processo de treinamento, utilizou-se a validação cruzada de 10 partições. Como foi utilizado o balanceamento *undersampling*, as instâncias descartadas no processo de balanceamento também foram reservadas para teste (corresponde ao teste 2 da Tabela I).
- (8) *Métricas de Avaliação*: Neste trabalho foram utilizadas as seguintes medidas de avaliação de desempenho: precisão¹⁰, *recall*¹¹ e F1-Score (ou f-measure)¹² e a área sob a curva ROC. A curva ROC demonstra o desempenho do modelo de Machine Learning por meio da relação da Taxa de Verdadeiro Positivo (Sensibilidade) e da Taxa de Falso Positivo (1 - Especificidade), variando o ponto de corte na probabilidade estimada.

4. RESULTADOS E DISCUSSÕES

4.1 Análise do perfil de mulheres com e sem a síndrome congênita

O perfil das mães são mulheres com média de idade de 26 anos, com idades variando entre 12 e 49 anos, com a maior frequência na faixa etária entre 18-26 anos e gravidez única (88.80%).

A grande maioria, 60.80%, são residentes da região Nordeste; contudo se analisarmos apenas as mães que tiveram filhos com alterações congênicas esta incidência aumenta para 67.60%.

E 40.30% das mães são da raça parda, contudo há um aumento na incidência, 45.60%, quando selecionamos apenas o universo de mães que tiveram infecção congênita devido ZIKV.

Em relação à realização dos exames de STORCH, grande parte das mães o realizaram, 88.51% do total. Contudo, em relação a detecção do Zika, apenas 13,94% das mães realizaram o exame, apresentando uma taxa de 65% dos registros sem informação e 21.06% não realizado.

Em relação ao histórico de arbovírus e mal formação congênita, ambos apresentaram altos índices de registros em branco com 74% e 72% respectivamente. A maioria das mães não apresentam histórico de arbovírus 23.54% e 26.17% de alteração congênita.

Os sintomas da Zika são leves e apenas 26.04% das mães tiveram manifestações, este resultado é condizente com dados mais recentes da literatura que mostram um pouco mais de 20% dos pacientes infectados são sintomáticos [Mitchell et al. 2019].

4.2 Análise do perfil das crianças/recém nascidos com e sem a síndrome congênita

A maioria dos casos notificados, 90.7%, são de recém nascidos. Em relação ao sexo, 59.9% são mulheres. 74.9% nasceram entre 37 e 42 semanas de gestação e todas as crianças que foram foram a óbito (2%) foram diagnosticadas com alteração congênita.

⁹The University of Waikato. Waikato environment for knowledge. Available at <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>, 2019.

¹⁰ $Precisao = \frac{VP}{VP+FP}$

¹¹ $Recall = \frac{VP}{VP+FN}$

¹² $F1Score = \frac{2*precisao*recall}{precisao+recall}$

Os padrões internacionais para peso, comprimento e circunferência craniana de recém nascidos, por idade gestacional e sexo são definidos pelo Intergrowth-21st¹³. Quando colocamos estas medidas por sexo no aplicativo para o cálculo dos desvios-padrão e percentis do Intergrowth¹⁴, os valores ficam abaixo dos valores padrões para ambos os sexos.

Na tabela InterGrowth para uma criança que nasceu com 37 semanas de gestação, a medida do perímetro encefálico de referência será 30,24 cm para meninas e 30,54 cm para meninos [Brasil et al. 2017].

4.3 Resultados dos modelos de aprendizado

A fim de avaliar as características mais relevantes para o diagnóstico de crianças com síndrome congênita devido à infecção pelo ZIKV, os seguintes algoritmos de classificação foram utilizados: C4.5 e Random Forest (RF). Os resultados obtidos a partir da criação do modelo e dos testes são apresentados na Tabela I. Vale ressaltar que o Teste 1 corresponde a 20% da base de dados após o balanceamento ('Sim'=579 e 'Não'=582), e o Teste 2 foi realizado com as instâncias excluídas durante o processo de balanceamento. Ou seja, como foi realizado o balanceamento *undersampling*, utilizou-se as sequências desconsideradas para realizar o teste 2. Por esta razão não tem instâncias da classe minoritária, classe 'Sim'.

Tabela I: Avaliação da qualidade dos modelos de aprendizado

		Precisao	Recall	F-Measure	ROC
Treino	C4.5				
	Classe Sim	83.7	81.2	82.4	85.7
	Classe Não	81.7	84.2	83.0	85.7
	RF				
	Classe Sim	92.7	93.2	92.9	97.8
	Classe Não	93.1	92.7	92.9	97.8
Teste 1		Precisão	Recall	F-Measure	ROC
	C4.5				
	Classe Sim	83.4	78.5	80.9	84.4
	Classe Não	79.7	84.3	81.9	84.4
	RF				
	Classe Sim	92.6	94.3	93.4	97.9
Classe Não	94.2	92.4	93.3	97.9	
Teste 2		Precisão	Recall	F-Measure	ROC
	C4.5				
	Classe Não	-	85.3	-	-
	RF				
Classe Não	-	88.3	-	-	

Observa-se que os resultados ficaram acima de 78.0% para todas as métricas. De modo geral, o modelo que apresentou melhor desempenho com taxas acima de 92.0% foi o classificador Random Forest.

Analisando os resultados de Teste 1, percebemos que o maior *recall* para a classe 'Sim' (com síndrome) foi de 94.3% utilizando o classificador RF, significando que o modelo conseguiu caracterizar muito bem a classe 'Sim'. Ou seja, de todos os usuários que tinham a síndrome, o modelo errou apenas 5.7% deles. Resultado bastante semelhante também foi encontrado na precisão, indicando uma taxa de falsos positivos bastante baixa. Vemos o mesmo comportamento para a classe 'não'; ou seja, o algoritmo Random Forest obteve o melhor desempenho para esta classe também, com todas as métricas acima de 92%. Observe, por exemplo, os valores das métricas F-measure e curva ROC com valores de 93.3% e 97.9%, respectivamente, para ambas as classes.

O Teste 2 é formado pelas instâncias que foram descartadas no processo de balanceamento, portanto só apresenta instâncias da classe 'Não'; assim só faz sentido falar na métrica *recall*. Seguindo o comportamento do Teste 1, o algoritmo Random Forest obteve o melhor desempenho com *recall*

¹³O projeto INTERGROWTH-21st desenvolveu uma curva de crescimento fetal padrão para uso internacional. O objetivo principal foi estudar o crescimento, saúde, nutrição e desenvolvimento neuromotor desde 14 semanas de gestação até os dois anos de idade. A curva padrão do INTERGROWTH-21st complementa a curva de crescimento de crianças da OMS, para ambos os sexos.

¹⁴<http://intergrowth21.ndog.ox.ac.uk/pt/ManualEntry>

de 88.3%. Ou seja, de todas as instâncias da classe majoritária desconsideradas no processo de balanceamento, o RF conseguiu um desempenho bastante satisfatório.

Analisando os atributos selecionados pelo algoritmo RF para a classificação das classes ‘Sim’ e ‘Não’ os atributos mais relevantes foram, respectivamente: peso do recém nascido, idade de gestante, comprimento do recém nascido, realização de exame de imagem, perímetro cefálico do recém nascido e a região em que a mãe mora.

Estes resultados estão de acordo com os trabalhos de [Brasil et al. 2017], que afirmam que o peso, o comprimento e o PC são um dos principais instrumentos utilizados para diagnóstico de alterações congênicas, sobretudo a microcefalia. Além disso, são medidas utilizadas na curva Intergrowth. Segundo a OMS, informações sobre a prevalência de microcefalia congênita são limitadas. Em todo o mundo, os registros de defeitos congênicos apresentam taxas de microcefalia congênita variando de (0,005%) a (0,1-0,2%) por 10.000 nascimentos, considerando a definição de perímetro cefálico menor ou igual a -3 desvios-padrão para idade e sexo (microcefalia severa) [Brasil et al. 2017].

A realização do exame de imagem é o principal critério de confirmação das anomalias do sistema nervoso central. Na maioria dos casos a definição da anomalia apenas pelo PC é incompleta, sendo necessário exames de imagem (ultrassonografia, tomografia computadorizada, ressonância nuclear magnética) para evidenciar a anomalia [Brasil et al. 2017].

A região de residência da mãe também é fator significante pois a região Nordeste concentra mais de 67.6% das mães que tiveram filhos com alteração congênita. Além disso, vários estudos já confirmaram que houve um surto de infecção por ZIKV na região Nordeste, no período de 2015 e 2016 [Brasil et al. 2017] [Nunes et al. 2016].

Além disso, estudos encontraram a associação estatística significativa entre malformações congênicas e algumas variáveis maternas, como a idade da mãe.

As chances de mães adolescentes com gestações múltiplas gerarem uma criança com malformação é 6 vezes maior, comparadas às adolescentes com gestação única, para mães tardias as chances são 11 vezes maiores quando comparadas as mães com idade entre 20 a 34 anos [Brasil et al. 2017] [Nunes et al. 2016].

5. ANÁLISE DAS REGRAS DE CLASSIFICAÇÃO OBTIDAS

Além de analisar os atributos mais relevantes a partir do Random Forest, buscamos também avaliar como estes atributos estão relacionados. Para isso, foram extraídas as regras de classificação a partir do algoritmo C4.5. As regras mais significativas foram:

- Se a gestante fez o exame de imagem e o perímetro cefálico do recém-nascido é menor que 30.4 cm ele foi classificado “com síndrome congênita”, com cobertura de 54%.
- Caso a mãe tenha realizado exame de imagem durante a gestação, e o recém nascido tem perímetro encefálico maior que 30.4 cm, e a mãe não apresentou sintoma de erupção cutânea, e o peso do recém-nascido foi menor que 3035 gramas, o paciente não tem alteração congênita, classificando 13% das instâncias.
- Se gestante fez o exame de imagem, e o perímetro cefálico do recém-nascido é maior que 30.4 cm, e durante a gravidez a mãe apresentou sintoma de erupção cutânea, a criança tem alteração congênita classificando 11% das instâncias.

Em relação aos exames de imagens apenas 11% das gestantes realizaram, e deste percentual menos da metade das mães tiveram filhos com alteração congênita, portanto não há alta correlação entre a realização do exame de imagem com a classificação. A importância do exame de imagem deve-se ao fato de ele ser um dos principais critérios de confirmação de alterações congênicas.

As regras são bastantes significativas, pois o atributo perímetro cefálico é um dos critérios de diagnóstico de alterações congênicas devido ao Zika Vírus, sobretudo a microcefalia. Em relação à medida do perímetro encefálico de 30.4, este fator está em conformidade com a medida adotada pelo OMS como padrão para recém-nascidos (a termo) que são 31,5 cm para meninas e 31,9 cm para meninos [Brasil et al. 2017].

Em relação aos sintomas, a maioria das mães não tiveram manifestações durante a gravidez, sendo uma das principais características desta enfermidade [Almeida et al. 2019]. As mães que apresentaram sintoma da Zika e tiveram filhos com alteração houve uma prevalência de erupção cutânea.

Quanto ao peso, 3035 gramas é um peso normal para bebês não prematuros¹⁵. Alguns autores afirmam que o baixo peso ao nascer e a prematuridade são apontados como desfechos adversos mais frequentes da infecção pelo ZKV [Ribeiro et al. 2018]. Outros entendem que apesar de crianças com microcefalia apresentarem maior probabilidade de ter baixo peso ao nascer, a frequência de prematuridade não é alta [Herber et al. 2019].

A regra ao combinar perímetro cefálico normal e peso inferior a 3035 está de acordo com a literatura, já que é sabido que crianças com microcefalia devido ao ZKV, mesmo quando nascem com peso e comprimento normais, apresentam agravamento de seu estado nutricional, apresentando baixo peso duas vezes maior em relação às crianças não infectadas com a mesma idade [Brasil et al. 2017].

6. CONSIDERAÇÕES FINAIS

A abordagem apresentada neste artigo identificou as características mais relevantes para o diagnóstico precoce de síndrome congênita associada a infecção do vírus Zika.

Foi realizada uma comparação entre os algoritmos C45 e Random Forest, e o algoritmo que apresentou melhor desempenho foi o classificador Random Forest com valores das métricas acima de 92%.

Os atributos comuns indicados pela Random Forest e pela árvore de decisão C4.5 foram: exame de imagem, perímetro cefálico e peso do recém nascido. Todos os atributos identificados são utilizados como critério de confirmação da alteração congênita e já foram preconizados em vários estudos científicos, validando o método para identificação de características mais relevantes para o diagnóstico precoce.

Em trabalhos futuros serão investigados outros métodos de balanceamento e tratamento de dados ausentes.

REFERÊNCIAS

- ALMEIDA, I., RAMOS, C., RODRIGUES, D., SOUSA, A. D., NASCIMENTO, M. C. D., SILVA, M. D., BATISTA, F., SANTOS, J. P. D., OLIVEIRA, R., SOARES, F., XAVIER, S., AND CARVALHO-COSTA, F. Clinical and epidemiological aspects of microcephaly in the state of Piauí, northeastern Brazil, 2015-2016. *Jornal de Pediatria* vol. 95, pp. 466 – 474, 08, 2019.
- BRASIL, DA SAÚDE, M., DE VIGILÂNCIA EM SAÚDE, S., AND DE ATENÇÃO À SAÚDE, S. Orientações integradas de vigilância e atenção à saúde no âmbito da emergência de saúde pública de importância nacional. 2017.
- HERBER, S., SILVA, A. A., SANSEVERINO, M. T. V., FRIEDRICH, L., RANIERI, T. M., FAVRETO, C., FRAGA, L. R., TERRA, A. P., SCHWARTZ, I. V., AND SCHULER-FACCINI, L. Prevalence and causes of congenital microcephaly in the absence of a Zika virus outbreak in southern Brazil. *Jornal de Pediatria* 95 (5): 600 – 606, 2019.
- MITCHELL, P. K., Y TERAN-ROMERO, L. M., BIGGERSTAFF, B. J., DELOREY, M. J., AUBRY, M., CAO-LORMEAU, V.-M., LOZIER, M. J., CAUCHEMEZ, S., AND JOHANSSON, M. A. Reassessing serosurvey-based estimates of the symptomatic proportion of Zika virus infections. *American Journal of Epidemiology* vol. 188, pp. 206–213, 2019.
- NUNES, M., CARLINI, C., MARINOWIC, D., KALIL N, F., FIORI, H., SCOTTA, M., ZANELLA, P. . A., SODER, R., AND COSTA, J. Microcephaly and Zika virus: a clinical and epidemiological analysis of the current outbreak in Brazil. *Jornal de Pediatria* vol. 92, pp. 230 – 240, 06, 2016.
- PRATA-BARBOSA, A. M., MELO, M., GUASTAVINO, A. B., AND CUNHA, A. J. L. A. D. Effects of Zika infection on growth. *Jornal de Pediatria* vol. 95, pp. S30 – S41, 00, 2019.
- RIBEIRO, I., MR, A., SILVA, J., ZM, S., COSTA, M., MACS, V., FMA, B., H, G., MY, W., AND E, S. Microcefalia no Piauí, Brasil: estudo descritivo durante a epidemia do vírus Zika, 2015-2016. *Epidemiologia e Serviço de Saúde* vol. 27, pp. 11, 00, 2018.

¹⁵O bebê nascido antes de completar 37 semanas (pré-termo) e com peso inferior a 2.500 gramas é considerado prematuro [Brasil et al. 2017]