

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Programa de Pós-Graduação em Modelagem Computacional

**Rafael Valente Veiga**

**Aplicação de programação genética gramatical multiobjetiva no estudo do  
efeito de múltiplas infecções e ambiente no desenvolvimento de atopia e  
fenótipos de asma**

Juiz de Fora

2017

**Rafael Valente Veiga**

**Aplicação de programação genética gramatical multiobjetiva no estudo do efeito de múltiplas infecções e ambiente no desenvolvimento de atopia e fenótipos de asma**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração em Modelagem Computacional, como requisito parcial para obtenção do título de Doutor em Modelagem Computacional.

Orientador: Prof. Dr. Helio José Corrêa Barbosa

Coorientadores: Prof. Dr. Heder Soares Bernardino; Prof<sup>a</sup>. Dr<sup>a</sup>. Neuza Maria Alcântara-Neves

Juiz de Fora

2017

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Valente Veiga, Rafael.

Aplicação de programação genética gramatical multiobjetiva no estudo do efeito de múltiplas infecções e ambiente no desenvolvimento de atopia e fenótipos de asma / Rafael Valente Veiga. -- 2017.

149 f.

Orientador: Helio José Corrêa Barbosa

Coorientadores: Heder Soares Bernardino, Neuza Maria Alcântara-Neves

Tese (doutorado) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2017.

1. Asma. 2. Atopia. 3. Programação Genética Gramatical. 4. Classificador. 5. Multiobjetivo. I. Corrêa Barbosa, Helio José, orient. II. Soares Bernardino, Heder, coorient. III. Alcântara-Neves, Neuza Maria, coorient. IV. Título.

Rafael Valente Veiga

Aplicação de Programação genética gramatical multiobjetiva no estudo do efeito de múltiplas infecções e ambiente no desenvolvimento de atopia e fenótipos de asma

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração em Modelagem Computacional, como requisito parcial para obtenção do título de Doutor em Modelagem Computacional.

Aprovada em: 08 Março de 2017

BANCA EXAMINADORA



---

Prof. D.Sc. Helio José Corrêa Barbosa - Orientador  
Universidade Federal de Juiz de Fora  
Laboratório Nacional de Computação Científica



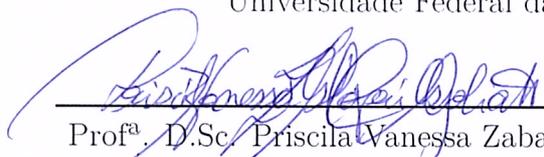
---

Prof. D.Sc. Heder Soares Bernardino - Coorientador  
Universidade Federal de Juiz de Fora



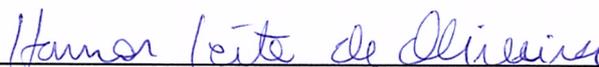
---

Prof<sup>a</sup>. D.Sc. Neuza Maria Alcântara-Neves -  
Coorientadora  
Universidade Federal da Bahia



---

Prof<sup>a</sup>. D.Sc. Priscila Vanessa Zabala Capriles Goliatt  
Universidade Federal de Juiz de Fora



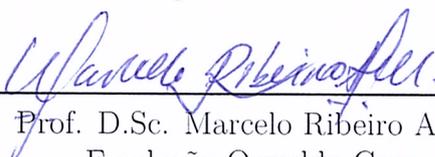
---

Prof. D.Sc. Itamar Leite de Oliveira  
Universidade Federal de Juiz de Fora



---

Prof. D.Sc. Douglas Adriano Augusto  
Fundação Oswaldo Cruz



---

Prof. D.Sc. Marcelo Ribeiro Alves  
Fundação Oswaldo Cruz

*Dedico este trabalho a família e amigos que sempre me apoiaram, em especial à minha  
noiva Larissa.*

## AGRADECIMENTOS

Agradeço ao meu orientador e coorientadores Prof. Dr. Helio José Corrêa Barbosa, Prof. Dr. Heder Soares Bernardino e Prof<sup>a</sup>. Dr<sup>a</sup>. Neuza Maria Alcântara-Neves pela orientação, atenção e disponibilidade para a elaboração deste trabalho.

Aos membros da banca Prof<sup>a</sup>. Dr<sup>a</sup>. Priscila Vanessa Zabala Capriles Goliatt, Prof. Dr. Itamar Leite de Oliveira, Prof. Dr. Douglas Adriano Augusto e Prof. Dr. Marcelo Ribeiro Alves pela disponibilidade de participação e contribuições acerca da tese;

À banca de qualificação, Prof. Dr. Douglas Adriano Augusto, Prof. Dr. Luis Paulo da Silva Barra e Prof<sup>a</sup>. Dr<sup>a</sup>. Priscila Vanessa Zabala Capriles Goliatt pelas sugestões e contribuições no exame de qualificação para melhoria do trabalho;

À minha noiva Larissa pela amor, atenção, carinho e a paciência para superar a distância o tempo exigido para realização deste trabalho.

Aos colegas do programa de pós-graduação em modelagem computacional pela amizade e bom humor que amenizaram momentos difíceis;

Agradeço ao grupo SCAALA pela disponibilidade do banco de dados e todo treinamento e conhecimento adquirido sobre estudos epidemiológicos em alergias e asma.

Aos funcionários do Programa de Pós-graduação em Modelagem Computacional, Reginaldo, Samantha e Renata pela disponibilidade e atenção nos serviços prestados;

Ao Nicolas Carels e à Fiocruz pela disponibilização de computadores para realização de experimentos computacionais.

Aos meu pai e minha mãe por todo apoio e suporte emocional fundamental para conclusão deste trabalho.

Ao meu avô Prof. Luiz Valente por incontáveis aulas sobre diversos assuntos que me fez gostar de sempre aprender.

Aos meus irmãos Gustavo e Marcelo e minhas tias e meus tios que estiveram sempre presentes mesmo distantes.

Ao Programa de Pós-graduação em Modelagem Computacional;

À Universidade Federal de Juiz de Fora;

À Capes pelo apoio financeiro concedido;

Ao Wellcome Trust pelo financiamento para obtenção dos dados.

“Tenho a impressão de ter sido uma criança brincando à beira-mar, divertindo-me em descobrir uma pedrinha mais lisa ou uma concha mais bonita que as outras, enquanto o imenso oceano da verdade continua misterioso diante de meus olhos”. (Isaac Newton)

## RESUMO

Nas últimas décadas os casos de asma e demais doenças alérgicas tiveram grande aumento em todo o mundo, sendo atualmente um grande problema de saúde pública. As causas do aumento da prevalência destas patologias são desconhecidas, porém a hipótese mais aceita é que seja oriunda da redução de infecções na infância como consequência da recente melhora nas condições de higiene. Esta redução das infecções pode levar ao desenvolvimento anômalo do sistema imune, aumentando assim a chance de desenvolver alergias e a asma. Tanto a asma como as demais alergias são patologias complexas, causadas por fatores genéticos e ambientais, de modo que o uso de ferramentas computacionais, tais como a programação genética podem contribuir para a compreensão destas doenças. Aplicou-se a técnica de Programação Genética Gramatical Multiobjetivo (MGGP) em dados obtidos de um coorte de 1445 crianças entre 4 e 11 anos para gerar modelos os quais possam representar como as relações entre infecções e ambiente podem explicar o desenvolvimento de atopia e asma. Para avaliar a presença de asma foi usado um questionário do ISAAC fase II e para avaliar atopia foram realizadas medições de anticorpos IgE contra alérgenos comuns e teste de reatividade cutânea. Os resultados obtidos mostram que os modelos gerados pela MGGP apresentam desempenho em acurácia competitivos aos obtidos pelo algoritmo C4.5 e regressão logística múltipla para os diferentes desfechos. Os resultados obtidos pela MGGP são de fácil interpretação, e capazes de encontrar relações complexas relevantes para o entendimento destas patologias, sendo assim, a MGGP é uma poderosa ferramenta para ajudar a compreender essas condições.

Palavras-chave: Asma. Atopia. Programação Genética Gramatical. Classificador. Multi-objetivo. Hipótese da Higiene.

## ABSTRACT

In the last decades cases of asthma and other allergic diseases have increased greatly throughout the world, being nowadays a major public health problem. The causes of this increased prevalence are unknown, however the most accepted hypothesis is that it comes from the reduction of childhood infections as a consequence of the recent improvement in hygiene conditions. This reduction of infections can lead to anomalous development of the immune system, thus increasing the chance of developing allergies, including asthma. Asthma and other allergies are complex pathologies caused by genetical and environmental factors, so the use of computational tools such as genetic programming can contribute to the understanding of these diseases. We applied the Multiobjective Grammatical Genetic Programming (MGGP) technique to data obtained from a cohort of 1445 children to generate models which may represent how the relationships between infections and environment may explain the development of allergies and asthma. To assess the presence of asthma, a questionnaire was used and allergy measurements were performed on IgE antibodies against common allergens and skin reactivity test. The results obtained show that the models generated by MGGP show a performance in accuracy that is competitive with those obtained by the algorithm C4.5 and multiple logistic regression for the different outcomes. The results obtained by MGGP are easy to interpret, and capable of finding complex relationship relevant to the understanding of these complex pathologies, therefore MGGP is a powerful tool to help understand these conditions.

Key-words: Asthma. Allergy. Grammatical Genetic Programming. Classifier. Multiobjective. Hygiene hypothesis.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Tipos de imunidade adaptativa . . . . .	18
Figura 2 – A prevalência de asma no mundo . . . . .	23
Figura 3 – Reação de hipersensibilidade imediata . . . . .	24
Figura 4 – O aumento de patologias alérgicas e asma . . . . .	25
Figura 5 – A hipótese da higiene . . . . .	26
Figura 6 – Esquema geral dos algoritmos evolutivos . . . . .	35
Figura 7 – Uso de árvore para representação de expressões . . . . .	36
Figura 8 – Exemplo do operador recombinação em GP . . . . .	37
Figura 9 – Exemplo de ordenação por dominância apresentando dois objetivos os quais se deseja minimizar. . . . .	39
Figura 10 – Exemplo do operador mutação em GGP . . . . .	41
Figura 11 – Exemplo do operador recombinação em GGP . . . . .	41
Figura 12 – Balanceamento dos dados . . . . .	52
Figura 13 – Relação entre acurácia e escolha do parâmetro para asma em C4.5 . . .	58
Figura 14 – Relação entre acurácia e escolha do parâmetro para asma atópica em C4.5	58
Figura 15 – Relação entre acurácia e escolha do parâmetro para asma não atópica em C4.5 . . . . .	59
Figura 16 – Relação entre acurácia e escolha do parâmetro para IgE em C4.5 . . .	59
Figura 17 – Relação entre acurácia e escolha do parâmetro para SPT em C4.5 . . .	60
Figura 18 – Relação entre acurácia e escolha do parâmetro para SPT com IgE em C4.5 . . . . .	60
Figura 19 – Relação entre tamanho da árvore e erro em MGGP para asma em diferentes gerações . . . . .	65
Figura 20 – Relação entre número de terminais e erro em MGGP para asma em diferentes gerações . . . . .	65
Figura 21 – Acurácia no grupo teste das melhores soluções MGGP em todas exe- cuções com desfecho asma utilizado os critérios tamanho máximo de árvore e número de terminais . . . . .	66
Figura 22 – Relação entre complexidade e erro em MGGP para IgE em diferentes gerações . . . . .	67
Figura 23 – Relação entre complexidade e erro em MGGP para SPT em diferentes gerações . . . . .	67
Figura 24 – Relação entre complexidade e erro em MGGP para SPT com IgE como possível exposição em diferentes gerações . . . . .	68
Figura 25 – Relação entre complexidade e erro em MGGP para asma atópica em diferentes gerações . . . . .	68
Figura 26 – Relação entre complexidade e erro em MGGP para asma não atópica em diferentes gerações . . . . .	69

Figura 27 – Relação entre complexidade e erro em MGGP para asma e sua comparação com os erros obtidos por RL e C4.5 . . . . .	70
Figura 28 – Relação entre complexidade e erro em MGGP para asma atópica e sua comparação com os erros obtidos por RL e C4.5 . . . . .	71
Figura 29 – Relação entre complexidade e erro em MGGP para asma não atópica e sua comparação com os erros obtidos por RL e C4.5 . . . . .	71
Figura 30 – Relação entre complexidade e erro em MGGP para IgE e sua comparação com os erros obtidos por RL e C4.5 . . . . .	72
Figura 31 – Relação entre complexidade e erro em MGGP para SPT e sua comparação com os erros obtidos por RL e C4.5 . . . . .	72
Figura 32 – Relação entre complexidade e erro em MGGP para SPT contendo IgE como possível exposição e sua comparação com os erros obtidos por RL e C4.5 . . . . .	73
Figura 33 – Mediana das acurácias no grupo de teste das melhores soluções obtidas por MGGP para asma usando duas abordagens distintas . . . . .	75
Figura 34 – Mediana das acurácias no grupo de teste das melhores soluções obtidas por MGGP para SPT usando duas abordagens distintas . . . . .	76
Figura 35 – Mediana das acurácias no grupo de teste das melhores soluções obtidas por MGGP para IgE usando duas abordagens distintas . . . . .	76
Figura 36 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma . . . . .	77
Figura 37 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para asma . . . . .	78
Figura 38 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma atópica . . . . .	82
Figura 39 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para asma atópica . . . . .	83
Figura 40 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma não atópica . . . . .	86
Figura 41 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para asma não atópica . . . . .	86
Figura 42 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para IgE	89
Figura 43 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para IgE . . . . .	90
Figura 44 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para SPT	93
Figura 45 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para SPT . . . . .	94
Figura 46 – Erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para SPT com IgE como possível exposição . . . . .	97

Figura 47 – Acurácia no grupo teste das soluções obtidas por RL, C4.5 e MGGP separadas por faixas de complexidades para SPT com IgE como possível exposição . . . . . 97

## LISTA DE TABELAS

Tabela 1 – Analogias entre Seleção Natural e Computação Evolucionista, retirada de Eiben e Smith (2003). . . . .	34
Tabela 2 – Variáveis usadas no estudo MGGP . . . . .	53
Tabela 3 – Teste de normalidade <i>Shapiro-Wilk</i> das acurácias na RL . . . . .	54
Tabela 4 – Resultados obtidos da execução da RL - Mostra a média, mediana, desvio padrão, máximo e mínimo da acurácia obtida de 30 execuções com a média da especificidade e sensibilidade. . . . .	56
Tabela 5 – Resultados obtidos da execução da C4.5 - Mostra a média, mediana, desvio padrão, máximo e mínimo da acurácia obtida de 30 execuções com a média da especificidade e sensibilidade. . . . .	62
Tabela 6 – Comparação das acurácias obtidas por RL e C4.5 . . . . .	63
Tabela 7 – Resumo dos resultados referentes a asma . . . . .	78
Tabela 8 – Exemplos de relações encontradas pelo MGGP para Asma . . . . .	80
Tabela 9 – Resumo dos resultados referentes a asma atópica . . . . .	82
Tabela 10 – Exemplos de relações encontradas pelo MGGP para asma atópica . . . . .	84
Tabela 11 – Resumo dos resultados referentes a asma não atópica . . . . .	87
Tabela 12 – Exemplos de relações encontradas pelo MGGP para asma não atópica . . . . .	88
Tabela 13 – Resumo dos resultados referentes a IgE . . . . .	90
Tabela 14 – Exemplos de relações encontradas pelo MGGP para IgE . . . . .	91
Tabela 15 – Resumo dos resultados referentes a SPT . . . . .	93
Tabela 16 – Exemplos de relações encontradas pelo MGGP para SPT . . . . .	95
Tabela 17 – Resumo dos resultados referentes a SPT com IgE . . . . .	98
Tabela 18 – Exemplos de relações encontradas pelo MGGP para SPT com IgE . . . . .	99
Tabela 19 – Matriz de covariância . . . . .	111
Tabela 20 – Comparação dos grupo estudado e excluídos do estudo . . . . .	112

## LISTA DE ABREVIATURAS E SIGLAS

HSV	Virus da hérpes
HAV	virus da hepatite
EBV	virus eptein barr.
<i>FcεRI</i>	Receptores de IgE de alta afinidade
GGP	Programação Genética gramatical
GP	Programação Genética
GINI	Índice de Renda
IFN- $\gamma$	Interferon gama
IgA	Anticorpo imunoglobulina A
IgE	Anticorpo imunoglobulina E
IgD	Anticorpo imunoglobulina D
IgG	Anticorpo imunoglobulina G
IgM	Anticorpo imunoglobulina M
IL-10	Interleucina 10
IL-21	Interleucina 21
IMC	Índice de massa corpórea
MHC	Complexo de histocompatibilidade principal
MGGP	Programação Genética Gramatical Multiobjetivo
NK	Células assassinas naturais
NSGA-II	do inglês <i>Non-Dominated Sorting Genetic Algoritm II</i>
OR	Razão de chance
RL	Regressão logística múltipla
SCAALA	do inglês <i>Social Change, Allergy and Asthma in Latin America</i>
SPT	Teste de reatividade cutânea
T <sub>H</sub> 1	célula T <i>helper</i> 1

T <sub>H</sub> 2	célula T <i>helper</i> 2
T <sub>H</sub> 17	célula T <i>helper</i> 17
T-reg	célula T regulatória

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO À ASMA E ALERGIAS</b> . . . . .	<b>17</b>
1.1	INTRODUÇÃO À RESPOSTA IMUNE . . . . .	17
1.2	RESPOSTA IMUNE ADAPTATIVA . . . . .	18
1.3	INTRODUÇÃO À ASMA E OUTRAS ALERGIAS . . . . .	20
1.4	IMPORTÂNCIA DO ESTUDO DE ASMA E ALERGIAS . . . . .	23
1.5	HIPÓTESE DA HIGIENE . . . . .	24
1.6	OBJETIVOS . . . . .	29
1.6.1	<b>Objetivo geral</b> . . . . .	29
1.6.2	<b>Objetivos específicos</b> . . . . .	29
<b>2</b>	<b>FUNDAMENTOS TEÓRICOS DA INTELIGÊNCIA COMPUTACIONAL</b> . . . . .	<b>30</b>
2.1	INTRODUÇÃO AO ESTUDO EPIDEMIOLÓGICO . . . . .	30
2.2	REGRESSÃO LOGÍSTICA . . . . .	32
2.3	ALGORITMO C4.5 . . . . .	33
<b>3</b>	<b>PROGRAMAÇÃO GENÉTICA</b> . . . . .	<b>34</b>
3.1	COMPUTAÇÃO EVOLUTIVA . . . . .	34
3.2	PROGRAMAÇÃO GENÉTICA - GP . . . . .	35
3.2.1	<b>Representação</b> . . . . .	36
3.2.2	<b>Definição da função de avaliação</b> . . . . .	37
3.2.3	<b>Operadores de variação</b> . . . . .	37
3.3	Non-Dominated Sorting Genetic Algorithm II - NSGA-II . . . . .	38
3.4	PROGRAMAÇÃO GENÉTICA GRAMATICAL - GGP . . . . .	39
3.4.1	<b>Gramática formal</b> . . . . .	39
3.5	APLICAÇÕES DE GP . . . . .	40
<b>4</b>	<b>METODOLOGIA PARA OBTENÇÃO DOS DADOS EPIDEMIOLÓGICOS</b> . . . . .	<b>42</b>
4.1	POPULAÇÃO DO ESTUDO . . . . .	42
4.2	QUESTIONÁRIO . . . . .	42
4.3	DEFINIÇÃO DE ASMA . . . . .	42
4.4	TESTE DE PUNTURA CUTÂNEA (SPT) . . . . .	43
4.5	TESTES SOROLÓGICOS PARA DETECÇÃO DE ANTICORPOS IGE ANTI-ALÉRGENOS AMBIENTAS . . . . .	43

4.6	TESTES SOROLÓGICOS PARA DETECÇÃO DE ANTICORPOS IGG CONTRA INFECÇÕES . . . . .	44
4.7	ESTUDO PARASITOLÓGICO . . . . .	44
4.8	ESTUDO NUTRICIONAL . . . . .	44
4.9	ESTUDO DE DESORDEM PSICOLÓGICA NA MÃE . . . . .	45
4.10	DESENVOLVIMENTO DOS MODELOS E ESCOLHA DAS VARIÁVEIS	45
4.11	CONSTRUÇÃO DOS BANCOS DE DADOS . . . . .	46
<b>5</b>	<b>EXPERIMENTOS COMPUTACIONAIS . . . . .</b>	<b>49</b>
5.1	EXPERIMENTOS COMPUTACIONAIS PARA REGRESSÃO LOGÍSTICA . . . . .	49
5.2	EXPERIMENTOS COMPUTACIONAIS PARA O C4.5 . . . . .	50
5.3	EXPERIMENTOS COMPUTACIONAIS PARA A MGGP . . . . .	50
5.4	RESULTADOS OBTIDOS . . . . .	52
5.4.1	<b>Resultados obtidos pela regressão logística . . . . .</b>	<b>52</b>
5.4.2	<b>Resultados obtidos pelo algoritmo C4.5 . . . . .</b>	<b>57</b>
5.4.3	<b>Resultados MGGP . . . . .</b>	<b>64</b>
5.4.4	<b>Resultados Asma . . . . .</b>	<b>77</b>
5.4.5	<b>Resultados IgE . . . . .</b>	<b>89</b>
5.4.6	<b>Resultados SPT . . . . .</b>	<b>92</b>
5.4.7	<b>Resultados SPT com IgE como possível exposição . . . . .</b>	<b>96</b>
<b>6</b>	<b>CONCLUSÕES . . . . .</b>	<b>101</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>102</b>
<b>A</b>	<b>Resultados adicionais . . . . .</b>	<b>111</b>
<b>B</b>	<b>Artigo publicado . . . . .</b>	<b>113</b>
<b>C</b>	<b>Questionário aplicado . . . . .</b>	<b>122</b>

# 1 INTRODUÇÃO À ASMA E ALERGIAS

Este capítulo descreve o sistema imune, asma e alergias. Na Seção 1.1 são explicados conceitos básicos sobre o funcionamento da resposta imune. A Seção 1.2 apresenta uma descrição sobre a resposta imune adaptativa com seus componentes e processos. A Seção 1.3 apresenta os mecanismos envolvidos na resposta imune da alergia e da asma. A Seção 1.4 fala sobre a importância do estudo de asma e alergias e seu crescimento nas últimas décadas. A Seção 1.5 explica a hipótese da higiene e suas evidências epidemiológicas.

## 1.1 INTRODUÇÃO À RESPOSTA IMUNE

O sistema imune é um sistema de estruturas e processos biológicos que protege o organismo contra doenças e substâncias nocivas (ABBAS, 2012). De modo a funcionar corretamente, o sistema imune deve detectar uma imensa variedade de agentes, como bactérias, vírus parasitos e toxinas, e distingui-los do tecido saudável do próprio corpo. O sistema imune pode ser separado em imunidade inata e imunidade adaptativa.

A imunidade inata (também denominada imunidade natural) proporciona a linha de defesa inicial sendo formada por: (1) barreiras físicas e químicas como os epitélios e substâncias químicas anti-microbianas produzidas pelas células epiteliais; (2) células fagocitárias como neutrófilos, macrófagos, células dendríticas e células NK (assassinas naturais); (3) proteínas do sangue como membros do sistema complemento e mediadores inflamatórios; e (4) proteínas denominadas citocinas que regulam e coordenam muitas das atividades de células da resposta imune. Os mecanismos da resposta imune inata são específicos para estruturas comumente encontradas em uma grande variedade de patógenos, tendo assim baixa especificidade (ABBAS, 2012).

A imunidade adaptativa é estimulada especificamente contra agentes infecciosos, tumores, alérgenos, toxinas, entre outros, e sua resposta cresce em magnitude e capacidade de defesa a cada encontro com o agente. O sistema imune adaptativo é capaz de reconhecer e reagir especificamente a um grande número de substâncias microbianas e não microbianas. Além disso, possui uma extraordinária capacidade de distinguir entre substâncias estreitamente relacionadas. Por causa deste fato, o sistema imune adaptativo também é conhecido por imunidade específica. Os principais componentes da imunidade adaptativa são células denominadas linfócitos e anticorpos (ABBAS, 2012). A resposta imune adaptativa tem grande importância nas doenças alérgicas por isso é explicado com maior profundidade.

## 1.2 RESPOSTA IMUNE ADAPTATIVA

Existem dois tipos de resposta imune adaptativa, a imunidade celular e a imunidade humoral que são representadas na Figura 1. A imunidade humoral é mediada por anticorpos que são produzidos exclusivamente pelos linfócitos B. Os anticorpos reconhecem e se ligam de forma específica a substâncias estranhas como toxinas e patógenos, neutralizando-os e ajudando na sua eliminação. Os anticorpos são especializados (IgM, IgD, IgA, IgG, IgE) podendo ativar diferentes mecanismos efetores. A imunidade humoral é eficiente principalmente contra toxinas e patógenos extracelulares (MURPHY, 2014).

A imunidade celular é mediada pelos linfócitos T. Os microrganismos intracelulares, como vírus e algumas bactérias, sobrevivem e proliferam no interior das células sendo, portanto, inacessíveis aos anticorpos que não podem atuar no meio intracelular. A defesa contra essas infecções é feita através dos linfócitos T que ativam os macrófagos para induzir a morte das células infectadas, eliminando as infecções e o reservatório da infecção (MURPHY, 2014).

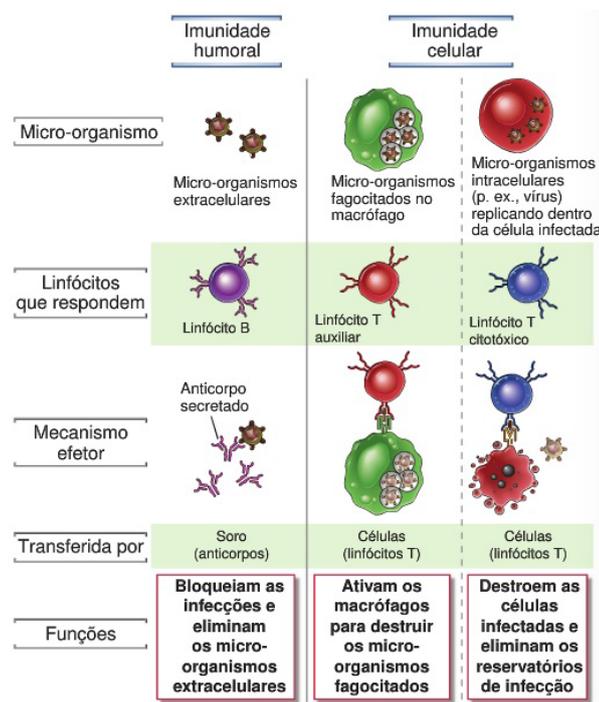


Figura 1 – Na Imunidade humoral, os linfócitos B secretam anticorpos que dificultam as infecções por microrganismos extracelulares e os eliminam. Na imunidade celular, os linfócitos T auxiliares ativam os macrófagos para destruir os microrganismos fagocitados, ou os linfócitos T citotóxicos destroem diretamente as células infectadas. Figura obtida de (ABBAS, 2012).

Os linfócitos são as células que reconhecem e respondem especificamente aos antígenos os quais são substância capazes de desencadear uma resposta imune. Cada linfócito reconhece um antígeno específico desta forma assume um papel central na

resposta imune adaptativa tanto na regulação da resposta imune quanto no mecanismo efetor. Existem subpopulações distintas de linfócitos que se diferem na forma de reconhecer antígenos e na sua função (MURPHY, 2014; ABBAS, 2012).

Os Linfócitos B são as únicas células capazes de produzir anticorpos, por isso têm uma grande importância para o sistema imune humoral. Os linfócitos T apenas reconhecem antígenos ligados ao complexo de histocompatibilidade principal (MHC) que são expressos na membrana das células nucleadas e constantemente apresenta aos linfócitos T amostras de peptídeos presentes no citoplasma (MURPHY, 2014; ABBAS, 2012).

Os linfócitos T apresentam populações funcionalmente distintas como linfócitos T citotóxicos que têm como função eliminar as células que apresentam através do MHC o antígeno ao qual o linfócito T citotóxico reconhece, e linfócitos T auxiliares que em resposta ao antígeno e estímulos produzidos pelas células apresentadoras de antígenos (macrófagos, células dendríticas, etc.), secretam citocinas que são proteínas que servem como sinalizadoras entre as células para coordenar a resposta imune (MURPHY, 2014).

Com o estudo da resposta imune às infecções, logo tornou-se evidente que existe uma grande heterogeneidade em como o sistema imune atua para combater diferentes infecções. A percepção de que essas reações imunológicas fenotipicamente diversificadas são dependentes de linfócitos T auxiliares levou ao aprofundamento nos estudos destas células e foi descoberto que os linfócitos T auxiliares podem se diferenciar em distintas subpopulações celulares de acordo com seu modo de ação, podendo ser:  $T_H1$  (T *helper* 1),  $T_H2$  (T *helper* 2),  $T_H17$  (T *helper* 17) e T-reg (T regulatório) (MURPHY, 2014).

A diferenciação em cada subtipo de linfócito é induzida pelos tipos de patógenos que cada subtipo pode combater melhor e a coordenação das subpopulações de células T auxiliares é feita através de processos de retroalimentação positivos e negativos. Linfócitos  $T_H1$  secretam interferon gama (IFN- $\gamma$ ) que estimula outros linfócitos a se diferenciarem em  $T_H1$  e inibe a geração de células  $T_H2$  e  $T_H17$ . Da mesma forma, a interleucina 4 (IL-4) produzida pelas células  $T_H2$  promove a diferenciação em  $T_H2$  e inibe a células  $T_H1$ . A interleucina 21 (IL-21) produzida pelas células  $T_H17$  promove diferenciação em  $T_H17$ , e a interleucina 10 (IL-10) produzida principalmente pelas células T-reg inibe  $T_H1$  e  $T_H2$  assim como alguns de seus mecanismos efetores (MURPHY, 2014; ABBAS, 2012).

A diferenciação dos linfócitos T auxiliares no fenótipo  $T_H1$  é estimulada por infecções intracelulares como infecções virais, bactérias intracelulares e alguns parasitas intracelulares. Essa subpopulação de linfócitos estimula o sistema imune através de citocinas pro-inflamatórias para privilegiar a imunidade celular que é mais eficiente contra estas infecções. O fenótipo  $T_H17$  é estimulado principalmente por bactérias e fungos.

Os linfócitos T-reg são uma população especializada que agem para suprimir a ativação do sistema imune, mantendo-o em homeostase e tolerante a antígenos próprios.

Estudos recentes indicam que a baixa produção de IL-10 excretado por essas células pode estar relacionada ao desenvolvimento de atopia <sup>1</sup> (FJELBYE et al., 2015; YANG et al., 2015).

O fenótipo  $T_H2$  é estimulado em resposta aos helmintos e alérgenos (que são antígenos que podem desencadear uma reação alérgica). Este fenótipo estimula a imunidade humoral rica em anticorpos do tipo IgE e eosinofilia, sendo uma resposta particularmente eficiente contra helmintos. O estímulo para a fagocitose desencadeado pela resposta efetora dos linfócitos  $T_H1$  não é eficiente pois, os helmintos são muito grandes em relação aos neutrófilos e macrófagos para serem fagocitados por eles. A resposta  $T_H2$  desencadeia mecanismos efetores capazes de eliminar infecções helmínticas, entre estes mecanismos podemos destacar a liberação de óxido nítrico que é um poderoso veneno e mediador inflamatório que ajuda na eliminação do parasito. A liberação do oxido nítrico pelos macrófagos é estimulado pelas citocinas Th2 (MURPHY, 2014; ABBAS, 2012).

### 1.3 INTRODUÇÃO À ASMA E OUTRAS ALERGIAS

Na maioria dos indivíduos saudáveis, o sistema imune não reage contra substâncias ambientais comuns e inofensivas, porém para parte da população, o sistema imune reage a uma ou mais substâncias ambientais comuns de forma anormal, levando a uma reação de hipersensibilidade a este alérgeno. Existem 4 tipos de reações de hipersensibilidade (ABBAS, 2012):

- Hipersensibilidade do tipo I, é um tipo de reação mediada por anticorpos do tipo IgE, que apresenta seus sintomas entre 2-30 minutos após a exposição ao alérgeno e sua resposta ocorre nas doenças alérgicas e nas helmintíases.
- Hipersensibilidade do tipo II, é mediada por IgM e IgG, causando a incompatibilidade de transfusões, eritroblastose fetal, reações auto-imunes, reações induzidas por drogas, entre outras. Ocorre entre 5-8 horas após a exposição.
- Hipersensibilidade do tipo III, é mediada por complexos imunes, levando a doenças como artrites e glomerulonefrites, ocorre entre 2-8 horas após a exposição.
- Hipersensibilidade do tipo IV, é mediada por células, é característica de alergias à metais e infecções por patógenos intracelulares. Ocorre entre 24 e 72 horas após a exposição.

Este trabalho estuda hipersensibilidade do tipo I que é mediada por IgE.

<sup>1</sup> que é a predisposição que algumas pessoa têm de produzir anticorpos do tipo IgE contra substâncias que são inócuas para maioria da população

As reações de hipersensibilidade do tipo I são caracterizadas por:(i) durante a fase de sensibilização, as células apresentadoras de antígeno apresentam o antígeno e fazem a coestimulação ao linfócito T; (ii) O linfócito T se diferencia em linfócito T auxiliar que estimula a ativação e mudança de isoforma de anticorpos dos linfócitos B; (iii) Os linfócitos B reconhecem o antígeno e produzem IgE específica contra o alérgeno;(iv) O IgE livre se liga aos receptores de IgE de alta afinidade ( $Fc\epsilon RI$ ) na superfície dos mastócitos.

Quando o indivíduo já sensibilizado ao alérgeno é re-exposto, ele já possui mastócitos revestidos de IgE em sua superfície, e essa nova exposição ao antígeno desencadeia ligações cruzadas das IgE da sua superfície com os antígenos, estimulando a liberação de mediadores inflamatórios como histamina, prostaglandinas, leucotrienos, citocinas e quimiocinas os quais são os responsáveis pela sintomatologia destas patologias (ABBAS, 2012). Os mecanismos da hipersensibilidade imediata são apresentados na Figura 3.

Os indivíduos geralmente sintetizam pequenas quantidades de IgE, produzindo preferencialmente outros isotipos de anticorpos, como IgM e IgG. A atopia é caracterizada pela capacidade genética do indivíduo de se sensibilizar e produzir anticorpos do isotipo IgE em resposta a alérgenos comuns, que de um modo geral são inócuos para os indivíduos não atópicos (JOHANSSON et al., 2004).

Uma vez que os indivíduos são atópicos, eles podem, em contato com alérgenos ambientais comuns, desenvolver reações alérgicas. As reações alérgicas se manifestam de diferentes maneiras, dependendo dos tecidos afetados, incluindo erupções cutâneas, sinusite, constrição brônquica, dor abdominal, diarreia e choque sistêmico (ABBAS, 2012).

A técnica de expor a pele dos indivíduos a diferentes alérgenos comuns e observar se ocorrem reações alérgicas é um importante teste clínico para alergia, sendo denominado teste de punctura cutânea (SPT, do inglês *skin prick test*). Recentes estudos epidemiológicos têm sugerido que existe uma dissociação entre ser atópico (ter altas doses de IgE) e ser positivo para SPT principalmente em países em desenvolvimento (COOPER et al., 2008; PONTE et al., 2011). Apesar de alergias serem patologias bem estudadas e de se conhecer boa parte dos seus mecanismos, pouco se sabe sobre quais os fatores levam ao desenvolvimento de alergias. Evidências epidemiológicas sugerem que a alergia seja uma patologia multifatorial com importantes componentes genéticos e ambientais (CAMPBELL et al., 2015).

A asma brônquica é uma doença heterogênea, com sinais e sintomas que variam de um indivíduo a outro e em um mesmo indivíduo no decorrer do tempo. Dentro de uma classificação simplista possui os seguintes fenótipos: asma atópica e asma não atópica (JOHANSSON et al., 2004).

A atopia é caracterizada pela capacidade genética do indivíduo de se sensibilizar e produzir anticorpos do isotipo IgE em resposta a alérgenos comuns, que de um modo geral

são inócuos para os indivíduos não atópicos. A asma não atópica apresenta os níveis séricos da IgE total normais ou baixos e testes cutâneos negativos para alérgenos. Entretanto, os pacientes podem apresentar características como infiltrado eosinofílico e citocinas de perfil Th2 (IL-4, IL-5 e IL-13), assim como na asma atópica (HUMBERT et al., 1999), recentemente foi mostrado que a asma não atópica apresenta alta produção de IFN- $\gamma$ , e aumento de neutrófilos e monócitos sanguíneos em crianças (FIGUEIREDO et al., 2012).

Outros fenótipos têm sido descritos como: asma de início precoce (infância) e asma de início tardio (adulto); asma com componente irreversível de limitação ao fluxo aéreo e asma com limitação totalmente reversível do fluxo aéreo; asma com eosinofilia no escarro induzido e asma com neutrofilia no escarro induzido; asma controlada com tratamento e asma de difícil controle (WENZEL, 2006).

O termo “endofenótipo” foi descrito como um interno ou intermediário fenótipo (ou seja, não é evidente a olho nu), que preenche a lacuna na cadeia causal entre genes e doenças distais (GOTTESMAN; SHIELDS, 1973). A definição de endofenótipos da asma abre a possibilidade de uma classificação da doença mais precisa, da identificação de fatores de risco específicos e de biomarcadores para o diagnóstico, prognóstico e monitorização da resposta ao tratamento. Foram descritos quatro subtipos da asma quanto á celularidade do esputo induzido: asma neutrofílica, asma eosinofílica, asma granulocítica mista e asma paucigranulocítica (SIMPSON et al., 2006; BAINES et al., 2014).

Um dos problemas de se estudar asma era a dificuldade de comparação dos estudos pelo uso de diferentes metodologias. Em 1998, a realização do Estudo Internacional para Asma e Alergia em Crianças (*“International Study of Asthma and Allergy in Childhood”* – ISAAC), usando como instrumento de coleta de dados um questionário para avaliar os sintomas de doenças alérgicas e asma, serviu de modelo podendo-se comparar estudos a nível mundial. Ficou evidente que não somente países desenvolvidos apresentavam altas prevalências de asma, mas também países em desenvolvimento, como o Brasil, possuíam também altas prevalências (BEASLEY; ASTHMA et al., 1998). Trabalhos mais recentes têm confirmado a alta prevalência desta patologia (GINA, 2015) como pode ser visto na Figura 2.

Os estudos em países desenvolvidos têm mostrado que a maioria dos casos de asma é atribuída à ocorrência de atopia. Porém, Cunha et al. (2010) mostrou que em países em desenvolvimento a maioria dos casos de asma não pode ser atribuído à atopia, indicando que o fenótipo majoritário de asma em países desenvolvidos é diferente daqueles observados em países em desenvolvimento.

Existe uma grande heterogeneidade na prevalência de asma e alergias no mundo (GINA, 2015) como pode ser visto na Figura 2, sendo maior no ambiente urbano do que no rural (RODRIGUEZ et al., 2015; ENDARA et al., 2015).

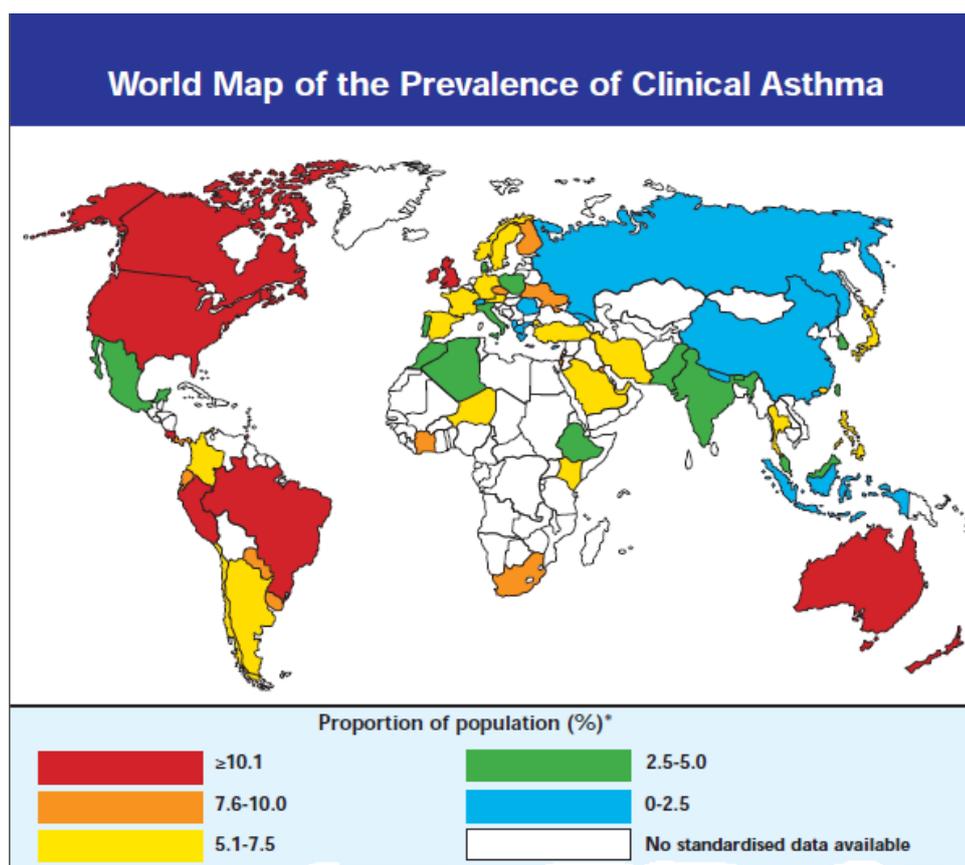


Figura 2 – A prevalência de asma clínica no mundo. Figura obtida de (GINA, 2015).

#### 1.4 IMPORTÂNCIA DO ESTUDO DE ASMA E ALERGIAS

A asma e alergias são considerados grandes problemas de saúde pública atualmente, causando grandes impactos econômicos tanto nas famílias afligidas quanto nos sistemas de saúde pública no mundo, produzindo também perdas de produtividade e deterioração da qualidade de vida (FRANCO et al., 2007). A asma destaca-se como a mais comum das doenças crônicas da infância, afeta cerca de 300 milhões de pessoas em todo mundo e estima-se que é responsável por cerca de 1 em cada 250 mortes em todo o mundo (GINA, 2015).

A prevalência de asma e alergias tem crescido acentuadamente nas últimas décadas, tendo este aumento sido observado em muitos países, incluindo o Brasil (BUTLAND et al., 2006; KABIR et al., 2011; WEHRMEISTER et al., 2012; DEVEREUX, 2006). O aumento na prevalência da asma tem sido associado a um aumento da sensibilização atópica e é acompanhada por aumentos semelhantes em outras desordens alérgicas tais como eczema e rinite (GINA, 2015) como pode ser visto na Figura 4.

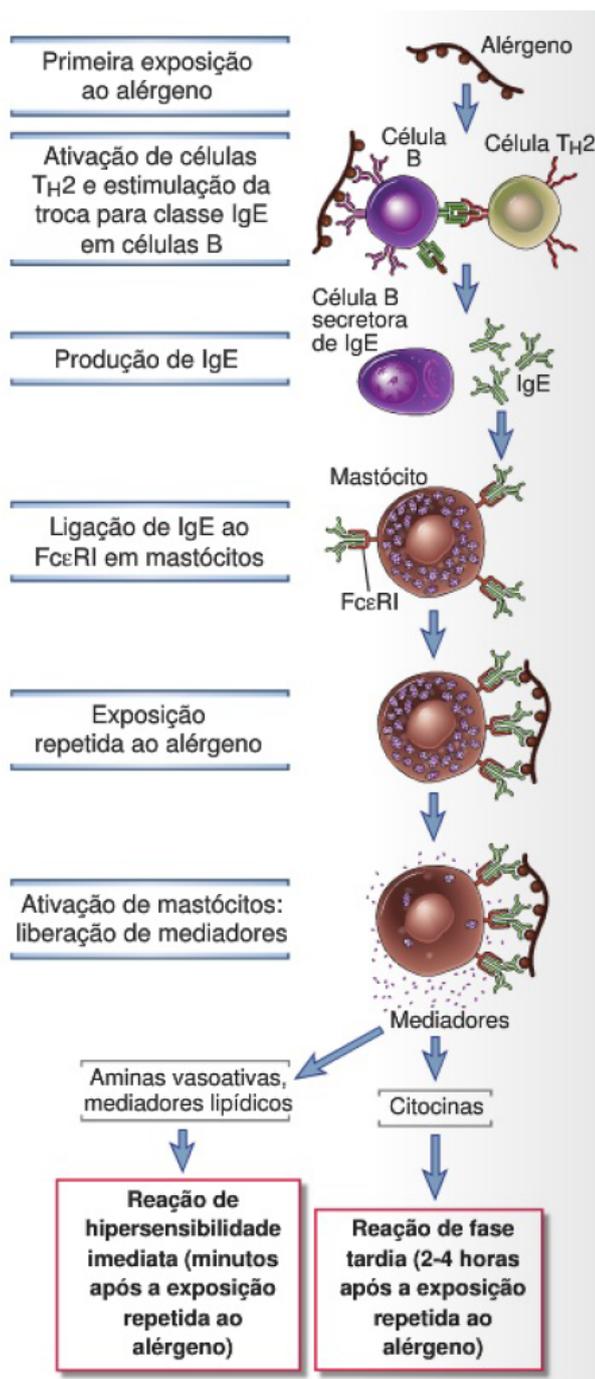


Figura 3 – As patologias causadas por reação de hipersensibilidade imediata são iniciadas com a exposição a um alérgeno, o que estimula linfócitos  $T_H2$  a induzir a produção de IgE. A IgE se liga aos receptores de alta afinidade ( $Fc\epsilon RI$ ) presente nos mastócitos, e a consequente exposição ao alérgeno ativa o mastócito para secretar os mediadores inflamatórios que são responsáveis pelas reações patológicas da alergia. Figura adaptada de Abbas (2012).

## 1.5 HIPÓTESE DA HIGIENE

Grandes esforços foram realizados para descobrir as causas deste rápido aumento das doenças alérgicas. Embora as razões para estas tendências permaneçam especulativas,

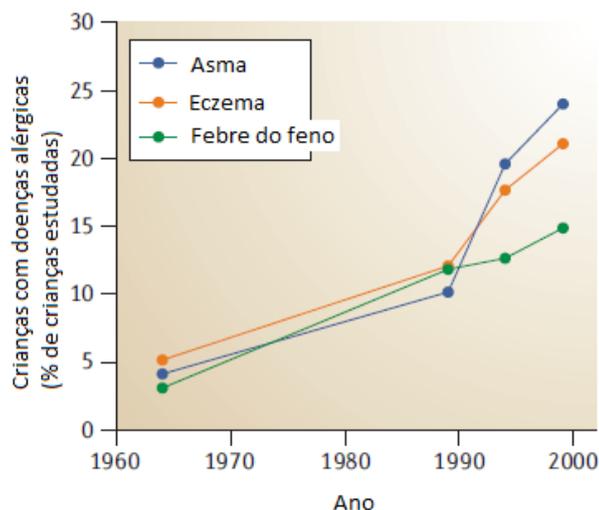


Figura 4 – O aumento da prevalência de asma, eczema e febre do feno desde 1964. Resultados de quatro inquéritos transversais de escolares de 9-12 anos em Aberdeen, Reino Unido, entre 1964 e 1999. Figura obtida de Devereux (2006).

a rapidez com que as taxas de prevalência das doenças alérgicas têm aumentado nos países afetados sugere fortemente que fatores ambientais tiveram um papel dominante sobre os fatores genéticos (TSE; HORNER, 2008). Assim, ocorreu grande procura de fatores ambientais que pudessem explicar este aumento, como por exemplo poluentes; porém os resultados foram inconclusivos (GOWERS et al., 2012).

Strachan (1989), utilizando evidências epidemiológicas de que o número de irmãos e a presença de irmãos mais velhos estão associados com proteção contra febre do feno, propôs a hipótese da higiene que afirma que o aumento de alergias pode ser explicado pela redução de infecções provocada pela melhora das condições de higiene.

Os humanos e seus sistema imune evoluíram em um ambiente repleto de infecções, produtos bacterianos, helmintos e baixa higiene. Com a recente revolução nas práticas de higiene, desenvolvimento dos antibióticos e vacinação, vive-se em um ambiente dito “limpo”. O sistema imune humano foi desenvolvido para um ambiente com muitas infecções e não amadurece adequadamente em um ambiente com poucas infecções, o que leva a respostas imunes desbalanceadas e inadequadas como as alergias (GERN; BUSSE, 2002).

Na década de 1990, com os estudos e caracterizações das linhagens de linfócitos  $T_H1$  e  $T_H2$ , surgiu uma possível explicação imune para a hipótese da higiene que seria: após o nascimento, os recém-nascidos apresentam uma resposta imune desviada em  $T_H2$ , para evitar rejeição durante a gestação. Porém, se durante seu desenvolvimento não ocorrerem exposições por infecções, o sistema imune não amadurecerá adequadamente uma resposta imune  $T_H1$ , mantendo-se a resposta imune desviada para  $T_H2$  aumentando a chance de desenvolvimento de patologias alérgicas, incluindo a asma (GERN; BUSSE, 2002). Esta primeira ideia do mecanismo imune envolvido da hipótese da higiene pode ser sintetizada

na Figura 5.

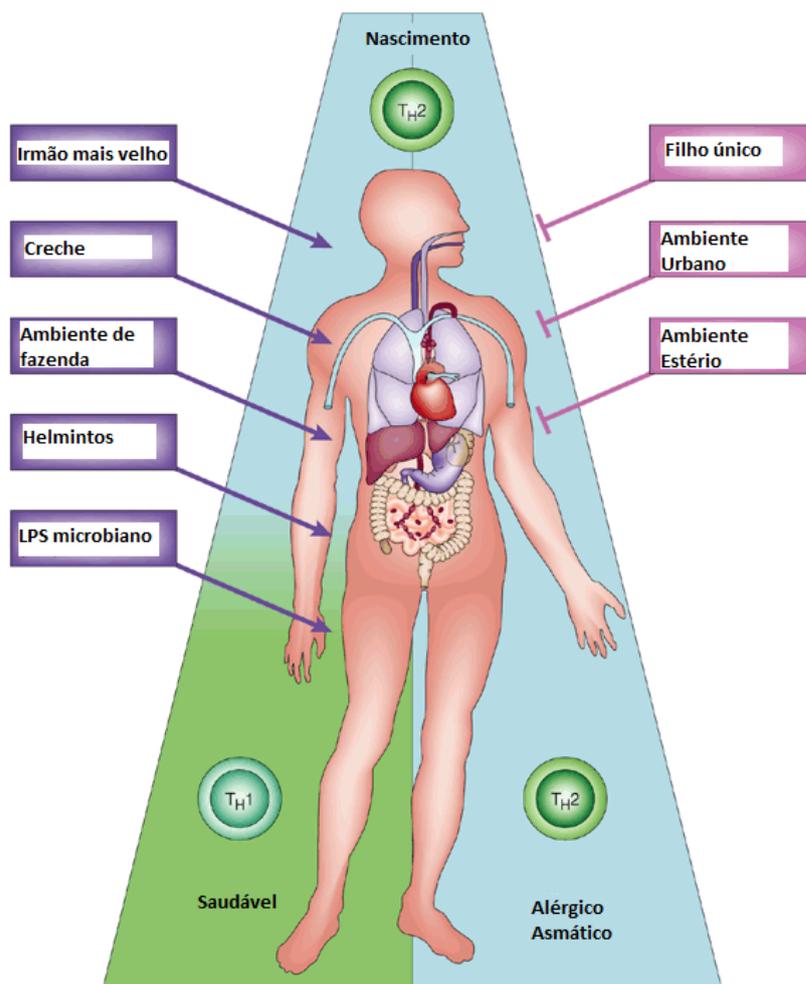


Figura 5 – O sistema imune é imaturo ao nascimento e desviado para ter maior resposta  $T_H2$ . Certos estímulos como infecções (obtidas dos irmãos mais velhos, creche), pode ajudar o sistema imune a desenvolver um balanço entre  $T_H1$  e  $T_H2$ , sendo assim saudável. Na ausência de estímulos, o sistema persiste desviado a ter uma resposta  $T_H2$ , assim aumentando o risco de desenvolver asma e alergias. Figura adaptada de Gern e Busse (2002).

Porém, as doenças auto imunes também apresentaram crescimento no mesmo período que as alergias. Doenças auto imunes apresentam mecanismos imunes típicos de respostas das subpopulações  $T_H1$ , diferentemente da resposta  $T_H2$  das alergias. Posteriormente, com o melhor entendimento dos mecanismos imunes e a descoberta de subpopulação de linfócitos T-reg, ficou evidente que a hipótese da higiene como foi concebida inicialmente necessitava ser revista. Assim, a hipótese da higiene foi aprimorada levando a ideias de que a ausência de estímulos para o amadurecimento do sistema imune podem estar também levando à resposta deficitária do tipo T-reg, aumentando tanto respostas exacerbadas  $T_H2$  quanto  $T_H1$  (KAMRADT; GÖGGEL; ERB, 2005; TSE; HORNER, 2008).

Existem muitas evidências que corroboram a hipótese da higiene, tanto originárias de estudos epidemiológicos como em estudos usando testes em animais (ALCANTARA-NEVES et al., 2012; MENDONÇA et al., 2012; VEIGA et al., 2011; YANG et al., 2015; ALCANTARA-NEVES et al., 2010; RODRIGUES et al., 2008; RIFFO-VASQUEZ et al., 2004). Algumas infecções bacterianas (por bactérias intra-celulares) podem induzir uma forte resposta  $T_H1$ , produzindo IFN- $\gamma$  que faz realimentação negativa na resposta  $T_H2$ ; assim espera-se que algumas infecções crônicas bacterianas possam reduzir as chances de ter alergias. Alguns trabalhos mostraram que a bactéria *Mycobacterium tuberculosis*, causadora da tuberculose, pode proteger contra atopia e asma (JANSON et al., 2007; THOMAS; SMITH, 1990), porém nem todos trabalhos acharam esta relação (THOMAS; SMITH, 1990). O estudo da bactéria *Helicobacter pylori* encontrou associação negativa com atopia (JANSON et al., 2007), porém outros trabalhos não encontraram essas associações (CHANG; HU, 2013; MATRICARDI et al., 2000).

A maioria das infecções virais induzem resposta imune  $T_H1$  e, assim infecções virais crônicas são possíveis candidatos a proteger seus hospedeiros contra doenças alérgicas. Matricardi et al. (2000) estudando a associação de infecções com atopia em uma população de cadetes italianos, mostrou que hepatite A protege contra o desenvolvimento de atopia, resultados que não foram reproduzidos no Brasil (VEIGA et al., 2011; ALCANTARA-NEVES et al., 2012). O vírus da família *Herpesviridae* como *Herpes simplex* que causa herpes e *Varicela zoster* que causa varicela e herpes zoster, e *Epstein Barr* causador da mononucleose infecciosa, causam infecções crônicas e comuns com grande capacidade de imunomodular a resposta do hospedeiro. Assim, são infecções que poderem estar associadas com patologias alergias.

Janson et al. (2007) estudando 1249 indivíduos adultos oriundos da Islândia, Suécia e Estônia, demonstraram que infecção por *Herpes simplex* protege contra atopia. No Brasil, estudos sugerem que o *Herpes simplex* também protege contra atopia (VEIGA et al., 2011; ALCANTARA-NEVES et al., 2012). Outra infecção viral desta mesma família que protege contra atopia é *Epstein Barr* (VEIGA et al., 2011; NILSSON et al., 2005). Esta infecção viral apresenta como mecanismo de escape do sistema imune, a produção de uma proteína similar à IL-10 que induz a resposta T-reg e, assim, pode suprimir as respostas  $T_H1$ ,  $T_H2$  e a desgranulação dos mastócitos, bloqueando a resposta exacerbada das desordens alérgicas.

Apesar de infecções helmínticas induzirem uma resposta  $T_H2$ , o que a principio levaria a aumentar as chances de desenvolver alergias, foram propostos mecanismos que explicariam como helmintos poderiam proteger contra patologias alérgicas (COOPER et al., 2003; COOPER et al., 2008; FIGUEIREDO et al., 2011):

- Apesar de tanto infecções helmínticas quanto alergias induzirem uma grande produção

de IgE, as respostas em alguns aspectos parecem ser diferentes. Os helmintos coevoluindo com os seres humanos, desenvolveram uma série de adaptações que os habilitaram a evadir do nosso sistema imune e sobreviver parasitando nosso corpo causando pouco impacto na sobrevivência do hospedeiro. Desta forma, estas infecções são geralmente de longa duração, assintomáticas e induzem a produção de citocinas T-reg como IL-10 que suprime resposta  $T_H2$  e a desgranulação dos mastócitos, reduzindo os sintomas típicos do mecanismo efetor da resposta  $T_H2$ .

- A presença de helmintos faz com que aumente a diversidade de anticorpos do tipo IgE, cada um reconhecendo um diferente epítipo (região da molécula do alérgeno que é reconhecido pelo anticorpo). Consequentemente, a diversidade de anticorpos que recobre o mastócito fica maior, fazendo com que a probabilidade da realização de ligação cruzada seja reduzida, portanto, a probabilidade de desgranulação dos mastócitos fica menor, evitando a resposta efetora típica da alergia.

Muitos trabalhos têm estudado a associação de helmintos com atopia e asma (LEONARDI-BEE; PRITCHARD; BRITTON, 2006; FLOHR; QUINNELL; BRITTON, 2009). Cooper et al. (2003) mostraram que infecções helmínticas protegem contra atopia. Rodrigues et al. (2008) mostraram que crianças novas (menores do que 4,6 anos) com alta carga de infecções por *Trichuris trichiura* apresentam menos hipersensibilidade cutânea quando ficam mais velhas (entre 4,6 e 11,3 anos).

Apesar das evidências de que o desenvolvimento de asma e atopia são influenciadas por infecções e condições de higiene, muitos estudos tem evidenciado que outros fatores também são importantes no desenvolvimento destas patologias complexas e multifatoriais como: (1) fatores ambientais e socioeconômicos (IDRIS et al., 2016; DICK et al., 2014; GALO BARDES et al., 2015; CHEN, 2014); (2) nutricionais (SILVA et al., 2013; D'INNOCENZO et al., 2014; WOOD; GIBSON, 2009); (3) psico-sociais (FEITOSA et al., 2011; MCQUAID; KOPEL; NASSAU, 2001); (4) genéticos (CAMPBELL et al., 2015; COSTA et al., 2015). Evidências recentes têm indicado que a interação entre estes diferentes fatores relacionados podem ser importantes para explicar essas patologias (CANN et al., 2011; RABINOVITCH et al., 2011; CAMPBELL et al., 2015).

Asma e demais doenças alérgicas têm crescido nas últimas décadas sendo atualmente importantes problemas de saúde pública em todo mundo. Estas patologias são complexas e multifatoriais com forte influência ambiental. Pouco se sabe como as infecções e o ambiente interagem para influenciar o desenvolvimento destas doenças. Este trabalho aplica a técnica de programação genética gramatical multiobjetivo em dados epidemiológicos para criar modelos que tentam explicar como o ambiente e as infecções influenciam no desenvolvimento de doenças alérgicas.

## 1.6 OBJETIVOS

### 1.6.1 **Objetivo geral**

Encontrar relações entre infecções, fatores ambientais, sociais, psicológicos e nutricionais que expliquem o desenvolvimento da asma e atopia.

### 1.6.2 **Objetivos específicos**

- Comparar o desempenho de uma MGGP com a regressão logística (RL) e o algoritmo de geração de árvore de decisão C4.5.
- Descrever as possíveis relações entre os fatores de risco com a asma e atopia.
- Desenvolver um método para o MGGP que seja mais adequado para o estudo do problema proposto.

## 2 FUNDAMENTOS TEÓRICOS DA INTELIGÊNCIA COMPUTACIONAL

Este capítulo descreve os fundamentos teóricos das técnicas empregadas neste trabalho. A Seção 2.1 descreve conceitos importantes de epidemiologia e as dificuldades de estudar problemas epidemiológicos. A Seção 2.2 explica os fundamentos da técnica de regressão logística. A Seção 2.3 descreve brevemente os fundamentos da técnica de inteligência computacional C4.5. A Seção 3.3 explica o que é otimização multiobjetivo e explica o algoritmo NSGA-II.

### 2.1 INTRODUÇÃO AO ESTUDO EPIDEMIOLÓGICO

A epidemiologia foi definida por Last, Abramson e Freidman (2001) como “o estudo da distribuição e dos determinantes de estados ou eventos relacionados à saúde em populações específicas, e sua aplicação na prevenção e controle dos problemas de saúde”. Essa definição deixa claro que os epidemiologistas estão preocupados com a melhora da saúde da população e seus indicadores. A Organização Mundial de Saúde (OMS) em 1948 definiu que “Saúde é um estado de completo bem-estar físico, mental e social e não apenas a mera ausência de doença” (BONITA; BEAGLEHOLE; KJELLSTRÖM, 2008). Assim, um dos grandes objetivos da epidemiologia é buscar quais características da população estão relacionados com a melhora ou piora no estado de saúde.

Existem muitas formas de medir associação entre características populacionais e o seu estado de saúde. Uma das principais medidas de associação aplicadas na epidemiologia em variáveis dicotômicas é a razão de chance ( $OR$ ) que é definida por

$$OR = \frac{p/(1-p)}{q/(1-q)} = \frac{p(1-q)}{q(1-p)} \quad (2.1)$$

onde  $p$  é a probabilidade de ficar doente entre os indivíduos que apresentam a característica avaliada e  $q$  a probabilidade de ficar doente entre os indivíduos que não apresentam a característica avaliada. A  $OR$  mede quantas vezes a chance de ficar doente no grupo exposto  $p/(1-p)$  é maior do que a chance no grupo não exposto  $q/(1-q)$ . Assim quando a  $OR > 1$  existe mais chance de estar doente no grupo exposto do que no grupo não exposto, sendo então a exposição um fator de risco para a doença. Quando a  $OR < 1$ , o grupo exposto apresenta menos chance de estar doente do que o não exposto, sendo essa exposição um fator de proteção.

Em estudos laboratoriais convencionais, quando se quer observar o efeito que uma exposição tem em uma doença, são comumente usados animais experimentais em ambientes extremamente controlados. Estes animais são usualmente isogênicos (com a mesma constituição genética), tem a mesma alimentação, temperatura, fatores de estresse e são minimamente manipulados por profissionais qualificados. Todo este cuidado é para

assegurar que o único fator que diferencie o grupo caso (animais expostos) do grupo controle (animais não expostos) é a exposição estudada. Assim pode-se garantir que as diferenças da saúde nos grupos caso e controle após os experimentos são originários da diferença na exposição. Porém não é ético expor humanos a ambientes extremamente controlados, e nem expor propositalmente a doenças e seus fatores de risco. Apesar de exames laboratoriais controlados com animais serem ideais para o entendimento de doenças, existem situações no qual não se podem fazer estudos em animais:

- Os animais podem apresentar mecanismos e resposta imune diferentes quando comparados aos humanos.
- Existem doenças específicas para humanos, para as quais não existem modelo experimental animal.
- Algumas exposições não se aplicam em animais, como por exemplo avaliação do impacto da renda na saúde e avaliação do impacto do divórcio dos pais na saúde.

Cada pessoa tem características genéticas específicas, uma história de vida própria, e é exposta a incontáveis fatores de risco e exposições desconhecidos. Então, como garantir que as diferenças encontradas na saúde do grupo caso e controle é causada pela exposição estudada? O viés de confundimento é o erro devido a uma diferença não aleatória na distribuição dos fatores de risco entre os grupos expostos e não expostos. A variável de confundimento está distribuída desigualmente entre os grupos comparados. Por exemplo: um grupo é mais idoso ou fuma mais que o outro.

Existem dois tipos de estratégias para reduzir o problema de viés de confundimento: através do desenho do estudo ou através de análise. O desenho de estudo caso controle pareado tenta evitar esse problema pela seleção para cada caso de um controle com mesma idade, gênero, condições de vida similares e proximidade genética. Contudo, esse estudo é mais caro, difícil de ser aplicado em grandes populações, nem sempre sendo viável selecionar controles adequados.

Também existem estratégias utilizando análise de dados que visam evitar o viés de confundimento. A análise estratificada é a mais simples pois simplesmente separa a amostra de acordo com as características e analisa a exposição e saúde entre os indivíduos com características iguais, ou seja, para cada característica a população é separada. Assim, quando separado por diferentes características como faixa etária, gênero, renda, se é fumante ou não, e se faz atividade física, existem tantas combinações de características possíveis que o número de pessoas com as mesmas características é muito pequeno. Sendo necessário uma amostra populacional muito grande para cada estrato ter uma quantidade de indivíduos representativa. Deixando essa solução inviável para aplicações práticas usuais. A prática mais adotada para reduzir os problemas causados pelo viés de confundimento é

o uso de modelos matemáticos, dentre eles podendo-se destacar o uso de regressão logística múltipla (RL) e regressão linear múltipla.

## 2.2 REGRESSÃO LOGÍSTICA

A regressão logística é o modelo mais usado pela Epidemiologia em estudos de associação quando o desfecho apresenta distribuição binomial. Este trabalho escolheu utilizar a regressão logística para comparar seu desempenho com o MGGP pois, esta técnica é amplamente aplicada em estudos epidemiológicos. A regressão logística mostrada na equação 2.2 é usada para encontrar um modelo que melhor explique uma variável qualitativa dicotômica  $Y$  codificada em  $[0,1]$  que é chamada de variável dependente ou desfecho. Esta variável comumente descreve o estado de saúde em modelos epidemiológicos. Para descrever a variável dependente, a regressão logística usa variáveis independentes ou preditoras  $X_1, X_2, \dots, X_k$ , onde  $k$  é o número de variáveis independentes. As variáveis independentes comumente representam as exposições e variáveis confundidoras, podendo apresentar valores discretos e contínuos (DOMÍNGUEZ-ALMENDROS; BENÍTEZ-PAREJO; GONZALEZ-RAMIREZ, 2011).

$$Y = \text{logit}(p) = \ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

ou ainda:

$$p = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k))} + e \quad (2.2)$$

onde  $p$  é probabilidade de um estado do desfecho,  $\beta$  é o vetor dos coeficientes da equação que define a magnitude da associação de cada variável independente  $X$  e  $e$  significa o erro. Apesar da regressão logística ser uma ferramenta adequada para encontrar associações e medir sua magnitude, ela apresenta algumas limitações por assumir independência das variáveis de exposição e pressupor que o desfecho pode ser modelado usando uma combinação linear das variáveis de exposição. Assim o desempenho da regressão logística quando usada em variáveis interdependentes ou condicionalmente dependentes é ineficiente e sua interpretação inadequada.

Muitos fenômenos biológicos são fundamentalmente não lineares com relações de condicionalidade entre as exposições, limitando o uso destas ferramentas estatísticas tradicionais, como a regressão logística. As patologias alérgicas e a asma são problemas complexos com múltiplos fatores de risco os quais apresentam inter-relações complexas ainda desconhecidas. Sendo um ótimo candidato ao uso de outras ferramentas capazes de modelar relações não-lineares e condicionais entre as variáveis preditoras.

### 2.3 ALGORITMO C4.5

O C4.5 (QUINLAN, 1993) é um algoritmo de inteligência computacional capaz de encontrar relações não lineares entre preditores e condicionais. Este algoritmo foi escolhido para ter seus resultados comparados com o MGGP por ter um desempenho bom e ser muito utilizado na literatura (LUSTGARTEN et al., 2017; SANZ et al., 2017). Este algoritmo se baseia em árvore de decisão ou árvore de classificação, que usa uma representação em árvore para tomar decisões. Na representação em árvore tem-se um nó inicial que é chamado de raiz e este nó pode ter filhos, os quais também podem ter seus filhos e assim sucessivamente, sendo os nós que não apresentam filhos chamados de folhas. Em uma árvore de decisão cada nó apresenta uma regra que define a qual nó filho deve ser submetida a avaliação dos dados para a regra subsequente. A folha define qual classe deve ser predita para estes indivíduos. O C4.5 é um algoritmo para criar árvore de decisão baseado no ganho da informação (WEBB, 1999), que é calculado através da variação da entropia da árvore após a tomada de uma decisão.

O C4.5 determina qual deverá ser o critério de decisão da árvore através da avaliação de todos os critérios possíveis, escolhendo aquele que apresenta o maior ganho de informação.

### 3 PROGRAMAÇÃO GENÉTICA

Este capítulo descreve a técnica de programação genética. A Seção 3.1 explica os fundamentos de computação evolutiva. A Seção 3.2 descreve a técnica de programação genética. A Seção 3.4 descreve a programação genética gramatical. A Seção 3.5 mostra as diversas áreas de estudo onde a programação genética tem sido empregada.

#### 3.1 COMPUTAÇÃO EVOLUTIVA

Poucas teorias tiveram o impacto no conhecimento e no modo de pensar humano como a teoria de seleção natural proposta por Darwin em 1858. A ideia principal é que o ambiente é repleto de indivíduos competindo por recursos limitados (alimentos, parceiros reprodutivos, moradia, etc.), assim os mais aptos (com maior valor adaptativo) tem maior chance de deixar mais descendentes, os quais compartilham mais características em comum com os progenitores do que com o restante da população. Portanto, a frequência de características mais adaptativas ao ambiente tende a crescer na população.

A observação da incrível capacidade que a seleção natural tem em solucionar problemas serviu de inspiração para desenvolver métodos de otimização baseados na ideia de seleção natural para resolver problemas complexos. A computação evolutiva apareceu como um ramo da ciência da computação que usa inspiração na seleção natural para resolver problemas complexos. O uso do processo de seleção natural para resolver problemas segue a analogia mostrada na Tabela 1 (EIBEN; SMITH, 2003).

Tabela 1 – Analogias entre Seleção Natural e Computação Evolucionista, retirada de Eiben e Smith (2003).

Seleção Natural	Computação Evolucionista
Ambiente	Problema
Indivíduo	Solução candidata
Valor adaptativo	Qualidade da solução

Apesar de existir uma grande variedade de métodos de computação evolucionista, as ideias subjacentes a estas técnicas são parecidas. Basicamente, tem-se uma população de indivíduos (soluções candidatas), o ambiente faz pressão de seleção sobre os indivíduos (sobrevivência do mais apto) e esta população sobrevivente sofre variações mutações/recombinação, reproduz e o ciclo recomeça. A cada geração os novos indivíduos tendem a ser mais bem adaptados que os das gerações anteriores. O esquema básico é mostrado na Figura 6. Existem dois componentes principais que atuam na computação evolucionista:

- operadores de variação, como mutação e recombinação, que servem para explorar o espaço de busca.

- seleção, que propicia maiores chances para as melhores soluções transmitirem suas características para próximas gerações.

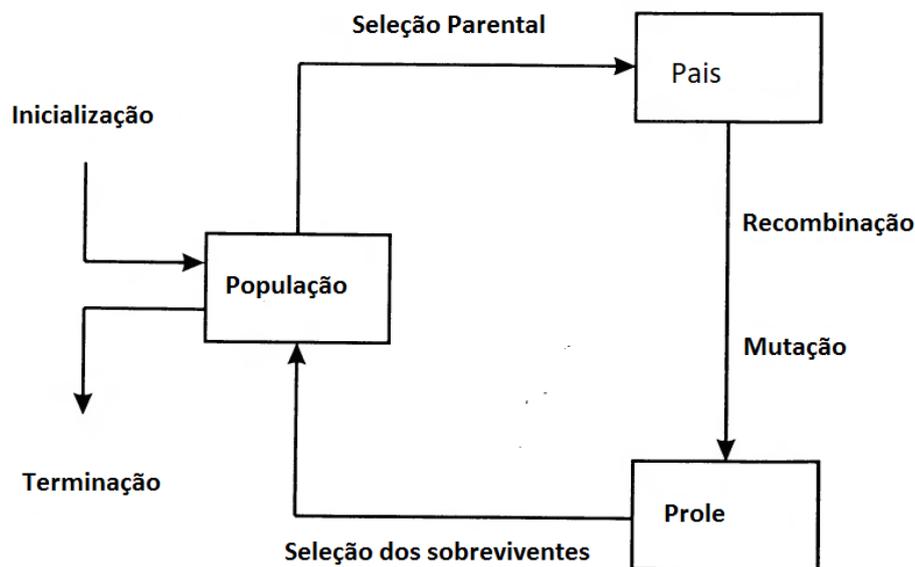


Figura 6 – Um esquema geral dos algoritmos evolutivos, adaptado de Eiben e Smith (2003)

O uso de técnicas de algoritmos evolutivos requer a especificação de alguns componentes, processos e operações para ser aplicado a um problema específico, sendo os principais:

- Representação da solução.
- Definição da função de avaliação (que mensura o valor adaptativo da solução)
- Tamanho da população.
- Mecanismo de seleção dos reprodutores.
- Operadores de variação (mutação e recombinação).
- Mecanismo de seleção dos sobreviventes.

### 3.2 PROGRAMAÇÃO GENÉTICA - GP

A programação genética (GP, do inglês "Genetic Programming") é um algoritmo evolutivo, sendo assim, uma meta-heurística estocástica de otimização. Muitos algoritmos parecidos com GP foram desenvolvidos na década de 80, sendo que pode-se destacar o de Cramer (1985), porém o desenvolvimento do GP moderno foi atribuído a Koza (1992) que foi capaz de desenvolver e formalizar a GP como um importante campo da computação evolucionista.

A ideia central da GP é realizar um processo evolutivo para criar artefatos computacionais, podendo ser programas de computador escrito em determinada linguagem de programação, expressões aritméticas, condicionais, etc. Estes artefatos, que serão chamados aqui de modelos para facilitar o entendimento, normalmente mapeiam entradas em saídas. Na GP, cada solução candidata (indivíduo) é um modelo. Assim a GP em vez de achar uma solução específica para um problema como acontece usualmente nos algoritmos de otimização, é útil para achar modelos (expressões matemáticas e lógicas, programas, etc.) capazes de resolver um problema.

### 3.2.1 Representação

A representação destes modelos na GP é tipicamente feita através de árvores, como pode ser observado na figura 7. Porém outras estruturas podem ser usadas para este fim, destacando-se as estruturas lineares e em grafos (EIBEN; SMITH, 2003).

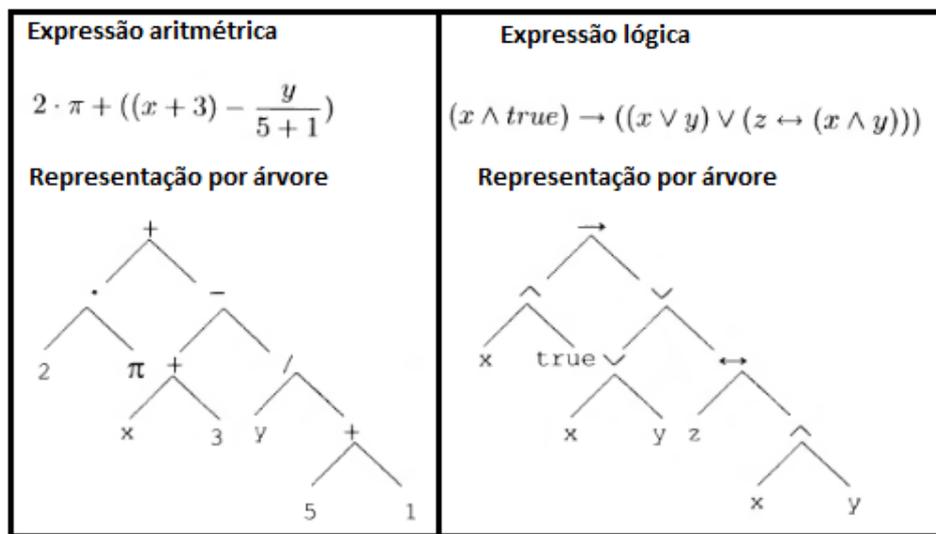


Figura 7 – Exemplos de como representar expressões aritméticas e lógicas usando-se árvore, adaptado de Eiben e Smith (2003).

A construção dos modelos requer uma sintaxe que vai delimitar um domínio sob o qual o processo evolutivo pode atuar. Esta sintaxe precisa ser fixada previamente, sendo definida por um conjunto de funções  $\mathbb{F}$  e terminais  $\mathbb{T}$ , dependentes do domínio de aplicação.

O conjunto das funções  $\mathbb{F}$  é formado por operadores matemáticos (+, -, produto, divisão, exponencial, etc.), operadores lógicos (ou, e, XOR, etc.), operadores condicionais (se  $x$  então  $y$  senão  $z$ ), laços iterativos (enquanto  $x$  faça  $y$ ). O conjunto dos terminais  $\mathbb{T}$  é formado por operandos ou seja, variáveis ou atributos ( $x_1, x_2$ , gênero, idade, temperatura, etc.), constantes ( $\pi, e$ , etc) e funções que não requerem argumentos (aleatório, etc.).

### 3.2.2 Definição da função de avaliação

A função de avaliação depende muito do domínio da aplicação, algumas vezes o problema é identificar artefatos computacionais que minimizam ou maximizam uma medida (como por exemplo quando se quer minimizar a energia livre de um sistema, ou quando se quer maximizar os lucros de um processo, etc.), nestes casos, podemos usar diretamente esta medida ou aplicar uma função sobre esta medida (como por exemplo uma função transformação para normalização, uma inversa, etc.). Quando o problema é encontrar artefatos computacionais que tenham um comportamento esperado, como por exemplo encontrar modelos preditores, usualmente são usadas medidas de erros entre os valores esperados e aqueles encontrados pelo modelo.

### 3.2.3 Operadores de variação

A GP usa como operadores de variação a mutação e recombinação. A mutação é uma alteração aleatória, usualmente “pequena” que atua sobre um único indivíduo. Por exemplo, a mutação pode operar mudando um elemento dentro do modelo do conjunto  $\mathbb{F}$  por outro ou um elemento do conjunto  $\mathbb{T}$  por outro. O operador recombinação toma partes de indivíduos diferentes (modelos diferentes) e substitui por alguma parte de outro indivíduo de forma aleatória. Um exemplo de recombinação é mostrado na Figura 8.

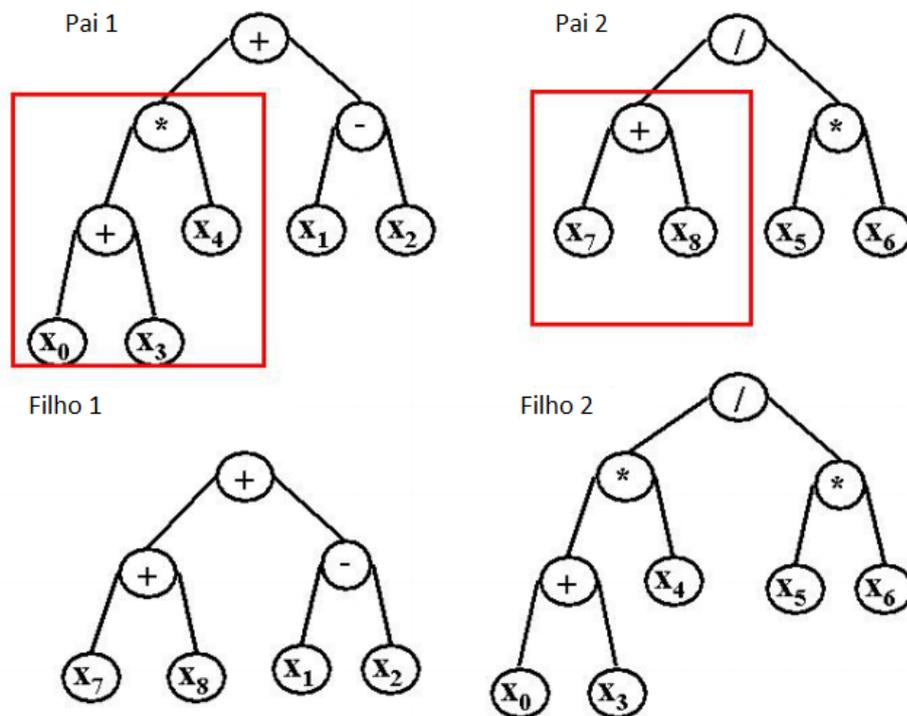


Figura 8 – Mostra exemplo de como funciona o operador recombinação na programação genética, adaptado de Ritchie et al. (2003).

### 3.3 Non-Dominated Sorting Genetic Algorithm II - NSGA-II

Um problema de otimização busca encontrar uma solução que maximize ou minimize um objetivo, como por exemplo encontrar um menor trajeto percorrido, um menor custo de produção, uma maior acurácia de um modelo de predição, etc. Contudo, muitos problemas requerem encontrar as melhores soluções de acordo com múltiplos objetivos, os quais, muitas vezes, são conflitantes entre si, sendo assim problemas de otimização multiobjetivo.

Um exemplo de problema de otimização multiobjetivo é quando se quer comprar um produto, pode-se considerar dois objetivos: (1) minimizar o preço do produto e (2) maximizar a qualidade do produto. Assim não é possível definir se um produto mais barato e de pior qualidade é melhor do que um produto de boa qualidade porém mais caro. Contudo, um bom algoritmo multiobjetivo para este problema deve ser capaz de encontrar um conjunto das melhores soluções com diferentes relações entre preços e qualidades.

O *Non-Dominated Sorting Genetic Algorithm II* (NSGA-II) é um algoritmo evolutivo multiobjetivo amplamente utilizado e com excelente desempenho em problemas com poucos objetivos. O NSGA-II foi proposto por Deb et al. (2002), que é uma versão aprimorada daquela proposta por Srinivas e Deb (1994).

O NSGA-II usa dois critérios para escolha das melhores soluções baseados nos objetivos:

- **Dominância:** todas soluções para as quais não existem soluções superiores em todos os objetivos simultaneamente são chamadas de soluções não dominadas. A população de soluções é ordenada de acordo com a dominância da seguinte forma: as soluções não dominadas são classificadas como posto 1, todas as soluções que são apenas dominadas pelas soluções do posto 1 são classificadas como posto 2, e este processo continua até que todas as soluções estejam classificadas. Quanto menor o posto melhor a solução. A Figura 9 ilustra um exemplo da ordenação por dominância no caso de dois objetivos.
- **Critério de aglomeração:** o cálculo da distância de aglomeração requer ordenar a população de acordo com o valor de cada função objetivo. Posteriormente, para cada função objetivo, as soluções de contorno (soluções com valores de função menores e maiores) são atribuídas à distância de aglomeração um valor infinito. Para todas as outras soluções intermediárias é atribuído um valor de distância igual à diferença padronizada dos valores de função de duas soluções adjacentes. Este cálculo é continuado com outras funções objetivas. O valor geral da distância de aglomeração é calculado como a soma dos valores de distância individuais correspondentes a cada objetivo.

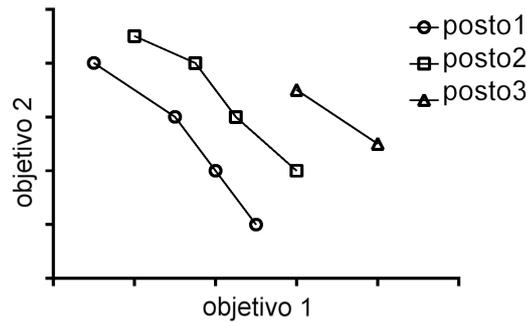


Figura 9 – Exemplo de ordenação por dominância apresentando dois objetivos os quais se deseja minimizar. O posto 1 não é dominado por nenhuma solução, o posto 2 é dominado por apenas as soluções do posto 1 e o posto 3 é dominado por soluções dos postos 1 e 2.

### 3.4 PROGRAMAÇÃO GENÉTICA GRAMATICAL - GGP

A programação genética tradicional é uma poderosa ferramenta para gerar programas. Porém, a GP pode apresentar dificuldades em aplicar restrições e limitar representações dos modelos, o que dificulta seu uso em problemas reais. A programação genética gramatical (WHIGHAM et al., 1995) é uma alternativa com maior capacidade de aplicação em problemas reais, pois aceita a aplicação de tipagem em elementos e apresenta uma sintaxe que contribui na interpretabilidade dos problemas e uma busca mais adequada da solução. A limitação do espaço de busca é feita através da restrição sobre a linguagem na qual os modelos podem evoluir. Para definir a sintaxe, é criada uma gramática formal.

#### 3.4.1 Gramática formal

Uma gramática formal  $G$  pode ser definida pela equação

$$G = \{N, \Sigma, R, S\}$$

onde  $N$  é um conjunto finito de não-terminais (ou seja de estruturas auxiliares da gramática),  $\Sigma$  é o conjunto finito de terminais ou alfabeto da linguagem (conjunto de elementos que aparecem na linguagem),  $R$  é o conjunto finito de regras (ou produções) e  $S \in N$  é o símbolo inicial. Cada regra pertencente a  $R$  segue a definição:

$$(\Sigma \cup N)^* N (\Sigma \cup N)^* ::= (\Sigma \cup N)^*, \quad (3.1)$$

onde  $*$  é o operador estrela de Kleene e  $\cup$  é a união dos conjuntos (HOPCROFT; ULLMAN, 1969).

Este trabalho usa uma gramática formal livre de contexto, ou seja, o lado esquerdo da equação 3.1 consiste de apenas um não-terminal. Assim a equação 3.1 pode ser

simplificada como

$$N ::= (\Sigma \cup N)^*$$

ou seja,

$$\langle \text{não terminal} \rangle ::= \text{expressão}$$

onde o  $\langle \text{não terminal} \rangle \in N$  e  $\text{expressão} \in (\Sigma \cup N)^*$ .

É possível existir várias expressões para cada não-terminal. Assim, por convenção, as expressões são delimitadas por "|" (lê-se; ou). Um exemplo de uma gramática pode ser visto a seguir:

$$G = \{N, \Sigma, R, S\}$$

$$N = \{\langle \text{expr} \rangle, \langle \text{op} \rangle, \langle \text{uop} \rangle, \langle \text{var} \rangle\}$$

$$\Sigma = \{\sin, \cos, \log, \exp, \sqrt{\quad}, +, -, \times, \div, x, y, 1, 2, 3, \text{pow}, (, )\}$$

$$S = \langle \text{expr} \rangle$$

$R :$

$$\langle \text{expr} \rangle ::= (\langle \text{expr} \rangle \langle \text{op} \rangle \langle \text{expr} \rangle) \mid \langle \text{uop} \rangle (\langle \text{expr} \rangle) \mid \langle \text{var} \rangle$$

$$\langle \text{op} \rangle ::= + \mid - \mid \times \mid \div \mid \text{pow} \mid \log \mid \exp$$

$$\langle \text{uop} \rangle ::= \sin \mid \cos \mid \sqrt{\quad}$$

$$\langle \text{var} \rangle ::= x \mid y \mid 1 \mid 2 \mid 3$$

Esta gramática pode gerar muitas expressões aritméticas como  $\sqrt{x + (2 \times (1 + y))}$ , porém existem expressões que não podem ser representadas por esta gramática específica, como por exemplo  $\tan(x)$ . O uso dos não-terminais como  $\langle \text{op} \rangle$  e  $\langle \text{uop} \rangle$  definem operações que recebem dois argumentos e um argumento, respectivamente. Assim, é possível garantir consistência, mantendo uma ampla variedade de operações possíveis. O que expande a possibilidade de usar GP em problemas mais complexos e realistas (BERNARDINO, 2012).

O uso da gramática delimita a criação da população inicial bem como a aplicação dos operadores variacionais como mutação e recombinação. Tanto para mutação e recombinação, somente é permitido trocar um não terminal  $N$  por outro do mesmo tipo, assim, mantendo a consistência dos modelos. Por exemplo, um não-terminal  $\langle \text{expr} \rangle$  só pode ser substituído por outro  $\langle \text{expr} \rangle$ . Exemplos de operadores de mutação e recombinação são mostrados nas Figuras 10 e 11 respectivamente.

### 3.5 APLICAÇÕES DE GP

Koza (2010) mostrou que a GP tem sido aplicada para encontrar soluções de problemas em diferentes áreas do conhecimento. Produzindo patentes e cerca de 76 resultados

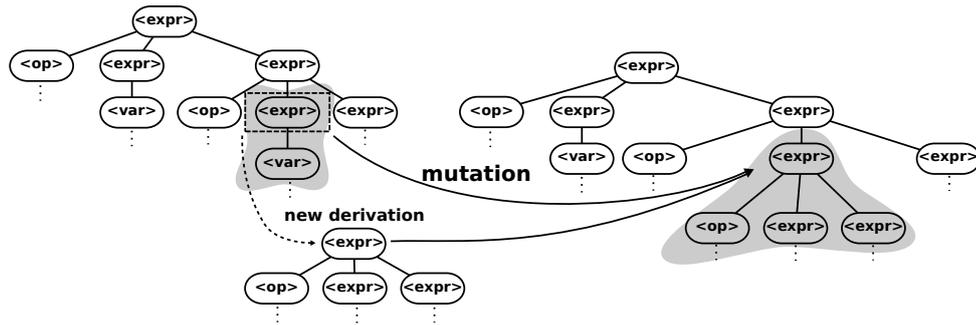


Figura 10 – Ilustração de um operador mutação na programação genética gramatical, Figura obtida de Barbosa e Bernardino (2010).

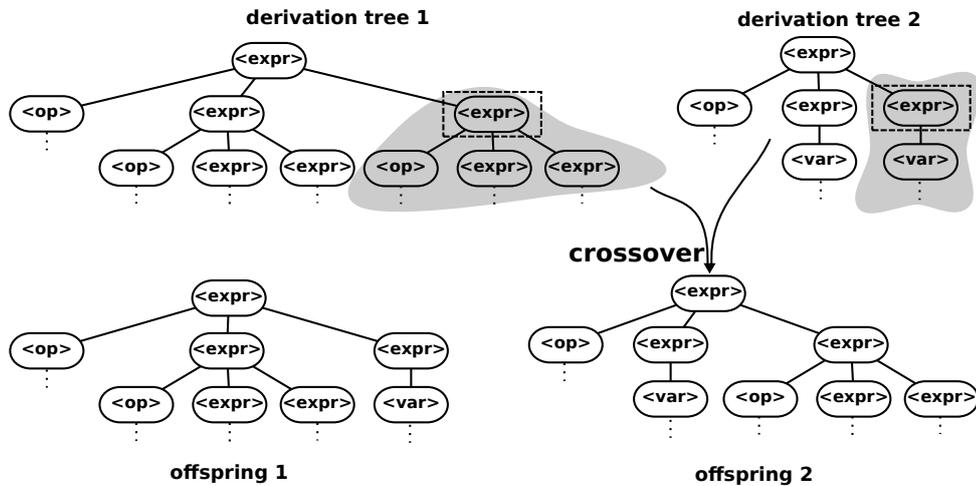


Figura 11 – Ilustração de um operador recombinação na programação genética gramatical. Figura obtida de Barbosa e Bernardino (2010).

que equivalem ou superam as soluções encontradas por humanos especialistas em seus campos de pesquisa. Entre os campos de pesquisas pode-se destacar o desenvolvimento de circuitos elétricos e quânticos (TOULOUSE, 2006; KOZA et al., 1996), o desenvolvimento de antenas de comunicação (LOHN; HORNBY; LINDEN, 2005), a álgebra finita (SPECTOR et al., 2008), o reconhecimento de imagem (TORRES et al., 2009), a regressão simbólica (UY et al., 2011) e a engenharia reversa (BONGARD; LIPSON, 2007).

No estudo de sistemas biológicos, a GP tem sido pouco aplicada, porém, recentemente diversos trabalhos tem aplicado GP no estudo de expressão gênica (YU et al., 2007; HONG; CHO, 2006; MITRA et al., 2006), modelagem de crescimento de algas (MUTTIL; CHAU, 2006), predição de câncer (BOJARCZUKA; LOPESB; FREITASC, 2001; GUO; NANDI, 2006; PAUL; IBA, 2009), predição de diagnóstico médico (TAN et al., 2003), na identificação e classificação de diferentes tipos de escoliose (WONG et al., 2000). Uma área que a GP tem atraído grande interesse é no estudos de associação ampla em genomas (SZE-TO et al., 2013; ESTRADA-GIL et al., 2007; NUNKESSER et al., 2007; MOORE; WHITE, 2007).

## 4 METODOLOGIA PARA OBTENÇÃO DOS DADOS EPIDEMIOLÓGICOS

### 4.1 POPULAÇÃO DO ESTUDO

O estudo foi realizado em uma coorte de 1.445 crianças do Programa SCAALA (Social Change, Allergy and Asthma in Latin America), em Salvador, Bahia, Brasil, uma cidade com população de 2,5 milhões, com alta prevalência de asma (SOLÉ et al., 2006). Esta coorte de crianças com 4 a 13 anos de idade foi originada de um estudo anterior sobre o impacto da instalação de saneamento básico na periferia da cidade sobre diarreia infantil (BARRETO et al., 2007). O desenho e metodologia do estudo atual foram descritos anteriormente por Barreto et al. (2006). Resumidamente, crianças nascidas entre 1994 e 2001, provenientes de 24 áreas geográficas (estas denominadas áreas sentinelas) foram selecionadas para representar a população sem saneamento. Estas foram novamente contatadas em 2005 e 2013 e foi aplicado um questionário do ISAAC fase II (respondido pelos pais ou responsáveis). Adicionalmente, dados demográficos foram re-coletados. Os dados aqui apresentados referem-se aos resultados obtidos nas coletas realizadas em 2005 e 2013. As crianças e adolescentes foram submetidas ao teste de puntura cutâneo para sete alérgenos e foram coletadas amostras de sangue, sendo as alíquotas dos soros obtidos estocadas a  $-70^{\circ}C$  até o uso, para dosagem de anticorpos IgG contra os patógenos estudados e anticorpos IgE anti-alérgenos.

As condutas médicas deste trabalho foram aprovadas pelo Comitê de Ética de Pesquisa do Instituto de Saúde Coletiva, Universidade Federal da Bahia e do CONEP (Conselho Nacional de Ética de Pesquisa). O termo de consentimento escrito para participação no trabalho, contendo informações sobre história de alergia e seus fatores de risco, foi obtido dos pais ou responsáveis pelas crianças.

### 4.2 QUESTIONÁRIO

Foi aplicado um questionário baseado no do ISAAC fase II, que contém questões validadas mundialmente sobre fatores de risco de asma e alergias, traduzido para o português e sendo adicionadas questões sobre ambiente, condições da residência, doenças alérgicas e potenciais fatores de risco. O questionário, mostrado no anexo C, foi aplicado por técnicos devidamente treinados e respondido pelos pais ou responsáveis.

### 4.3 DEFINIÇÃO DE ASMA

Asma foi definida de acordo com informações obtidas do questionário ISAAC fase II, sendo consideradas asmáticas as crianças que responderam positivamente para presença de sibilo nos últimos 12 meses e pelo menos uma resposta positiva para estas questões:

(1) sibilância <sup>1</sup> após exercício; (2) ter apresentado quatro ou mais episódios de sibilos nos últimos 12 meses; (3) acordar com falta de ar; (4) asma diagnosticada pelo médico.

A asma é uma patologia complexa apresentando diferentes fenótipos. Este trabalho estuda dois fenótipos: (1) asma não atópica, definida como a positividade para asma em uma população composta apenas de indivíduos IgE negativos (ser asmático e IgE negativo usando como grupo de referência IgE negativos e não asmático); (2) asma atópica, definida como a positividade para asma em uma população composta somente por indivíduos IgE positivos (ser asmático e IgE positivo usando como grupo de referência IgE positivo e não asmático). O uso do grupo de referência para asma atópica como sendo não asmático e positivo para IgE, em vez de não asmático, é porque quando é comparado asmático e IgE positivos com asmático negativo para IgE são mudadas concomitantemente duas características (ser asmático e ter IgE). Desta forma, não é possível saber se o efeito medido vem da asma ou da atopia. Quando é usado como referência o grupo não asmático e IgE positivo, é medida apenas a característica ser asmático na presença de IgE, definindo melhor a asma atópica sem problemas em definir a origem do efeito.

#### 4.4 TESTE DE PUNTURA CUTÂNEA (SPT)

O teste de puntura cutânea (SPT) para os seguintes alérgenos identificados como relevantes regionalmente, por trabalhos anteriores não publicados: ácaros (*Dermatophagoides pteronyssinus*, *Blomia tropicalis*), baratas (*Blattella germanica*, *Periplaneta americana*), epitélio de cão e gato, e mistura de fungos (*Aspergillus amstelodami*, *Aspergillus fumigatus*, *Aspergillus niger*, *Aspergillus terreus*, *Penicillium brevicompactum*, *Penicillium expansum*, *Penicillium notatum*, *Penicillium roqueforti*, *Cladosporium fulvum*, and *Cladosporium herbarum*) usando kits ALK- Abello (São Paulo, Brasil). O controle positivo e negativo usados nos testes foram respectivamente histamina 10mg/mL e solução salina. Alérgenos e controles foram aplicados no antebraço direito. Após 15 minutos foi feita a medida da reação, sendo considerado positivo quando a média de dois diâmetros perpendiculares dos testes fossem 3mm maior do que a média de dois diâmetros perpendiculares do controle negativo. Crianças positivas para pelo menos um dos alérgenos estudados foram classificadas como positivas. As crianças que usavam medicamentos anti-alérgicos foram excluídas do estudo.

#### 4.5 TESTES SOROLÓGICOS PARA DETECÇÃO DE ANTICORPOS IGE ANTI-ALÉRGENOS AMBIENTAS

A presença de anticorpos IgE no soro das crianças contra *Dermatophagoides pteronyssinus*, *Blomia tropicalis*, *Periplaneta americana* e *Blattella germanica* foi determinada

<sup>1</sup> Sonoridade aguda e/ou chiada produzida pelas vias respiratórias

usando kits comercialmente disponíveis (ImmunoCap system, Phadia AB, Uppsala, Sweden). Os resultados foram expressos em  $kU/L$  (que corresponde à  $2,4ng/mL$ ) obtidos a partir de uma curva padrão produzida através de diluições seriadas de IgE humano contra um soro padrão para IgE da Organização Mundial de Saúde (standard WHO 75/502). Foi adotado o ponto de corte de  $0,70kU/L$ , sendo considerado positivo para crianças cujo valor obtido é maior do que este ponto de corte para ao menos um dos alérgenos estudados.

#### 4.6 TESTES SOROLÓGICOS PARA DETECÇÃO DE ANTICORPOS IGG CONTRA INFECÇÕES

Foi escolhido infecções crônicas com potencial de imunomodular o sistema imune. Para avaliar esta infecções foi avaliado a presença de anticorpos IgG contra estas infecções no soro das crianças contra os seguintes infecções: Herpes simplex (HSV), Herpes zoster (VZV), Epstein-Barr (EBV), *Toxoplasma gondii* e *Helicobacter pylori* foram determinadas por ELISA indireta usando kits da Diamedix (Miami, Florida, USA) e Hepatite A (HAV) usando kits da ADALTIS (Toronto, Canada) seguindo as instruções do fabricante. Os pontos de corte foram determinados pelo índice obtido pela relação:  $\frac{\text{absorbância da amostra}}{\text{absorbância do calibrador}}$ . Índices maiores que 1,1 foram considerados positivos, menores que 0,9 considerados negativos e entre 0,9 e 1,1 foram considerados indeterminados e estas crianças foram removidas do estudo.

#### 4.7 ESTUDO PARASITOLÓGICO

A detecção de infecções por *Ascaris lumbricoides* e *Trichuris trichiura* foi feita através da realização de duas coletas de exames de fezes por indivíduo, em tempos distintos para detecção de ovos destes parasitas. As fezes foram analisadas usando a técnica de sedimentação de Hoffman et al. (1934). Duas lâminas foram analisadas por amostra e a quantificação foi realizada através da técnica de Katz et al. (1972). Todas as crianças com resultados positivos foram tratadas com drogas anti-helmínticas apropriadas.

#### 4.8 ESTUDO NUTRICIONAL

Informações sobre os padrões alimentares foram obtidas com base em questionário de frequência alimentar, validado por Matos et al. (2012). Este questionário consiste em 98 alimentos, relacionados ao consumo alimentar nos últimos 12 meses. Utilizou-se uma análise de componentes principais para se obter quatro padrões alimentares que foram discretizados utilizando-se de valores relativos aos seus tercís. A análise de detalhes para definir o padrão alimentar pode ser encontrada no trabalho de D’Innocenzo et al. (2011). O padrão alimentar 1 foi caracterizado pela predominância de frutas, hortaliças, leguminosas, cereais e peixes. O padrão alimentar 2 foi caracterizado pela predominância

de leite e produtos lácteos, ketchup / maionese / mostarda e frango. O padrão alimentar 3 foi caracterizado pela predominância de alimentos fritos, doces, lanches, refrigerante / suco artificial. O padrão alimentar 4 foi caracterizado pela predominância de salsichas, ovos e carne vermelha.

#### 4.9 ESTUDO DE DESORDEM PSICOLÓGICA NA MÃE

O questionário SRQ-20 foi utilizado para avaliar distúrbios psiquiátricos menores na mãe. Este instrumento foi desenvolvido pela Organização Mundial da Saúde (World Health Organization, 1994) e validado no Brasil por Mari e Williams (1986). Este questionário é composto por 20 perguntas com respostas dicotômicas (sim / não) referentes à presença ou ausência de sintomas de depressão, ansiedade e distúrbios somáticos no mês anterior. A presença de 8 ou mais respostas positivas às questões indica suspeita de transtornos psiquiátricos na mãe. Embora não caracterizando um diagnóstico psiquiátrico, indica sofrimento psíquico significativo. Esse ponto de corte foi definido de acordo com estudos previamente realizados no Brasil (MARI; WILLIAMS, 1986).

#### 4.10 DESENVOLVIMENTO DOS MODELOS E ESCOLHA DAS VARIÁVEIS

As variáveis usadas como desfecho para a elaboração dos modelos foram: (1) Asma, definida de acordo com o item 4.3; (2) Asma atópica definida de acordo com o item 4.3; (3) Asma não-atópica definida de acordo com o item 4.3; (4) SPT, que é a presença de reatividade cutânea contra alérgenos ambientais comuns, definido de acordo com o item 4.4, e (5) IgE, que é a presença de anticorpos do tipo IgE contra alérgenos ambientais comuns, definido de acordo com o item 4.5.

As infecções estudadas são infecções comuns (com alta frequência na população) e escolhidas por apresentarem alta capacidade imunomoduladora por longo período de tempo, sendo estas as infecções virais HSV, VZV, EBV e HAV (descritos no item 4.6). Os modelos de regressão logística usando os desfechos descritos anteriormente, e as suas associações com infecções virais ajustados por fatores ambientais foi publicado anteriormente por Veiga et al. (2011). Uma comparação de modelos equivalentes aos publicados pelo trabalho descrito anteriormente Veiga et al. (2011) com modelos gerados por GGP foi publicado por Veiga et al. (2015). Outras infecções comuns crônicas escolhidas para o presente trabalho com grande capacidade imunomoduladora por longo tempo são: as infecções causadas pelo protozoário *T. gondii* e pela bactéria *H. pylori* descritas no item 4.6, e as infecções causadas pelos helmintos *A. lumbricoides* e *T. trichiura* descritos no item 4.7. Os modelos usando regressão logística com os desfechos e a exposição a todas as infecções escolhidas para o presente trabalho foram publicados anteriormente por Alcantara-Neves et al. (2012).

As outras variáveis escolhidas para participar da criação dos modelos foram aquelas

que apresentam características importantes que poderiam modificar as relações entre as infecções e os desfechos anteriormente descritos. Outra característica importante usada para definir essas variáveis do estudo é se sua frequência apresenta valores intermediários. Foram escolhidas como variáveis que representam características individuais (aquelas características individuais que são independentes dos outros preditores) o gênero, a idade e o histórico de asma nos pais. A variável foi usado o índice de renda (GINI) foi escolhida para representar fatores socioeconômicos. Os fatores ambientais escolhidos foram: número de irmãos, se a criança frequentou creche, presença de fumantes na residência, presença de rede de esgoto na residência, presença de água tratada na residência, se troca roupa de cama uma ou mais vezes por semana, a presença de gatos na residência, a presença de cachorro na residência, a presença de mofo ou umidade na residência. 4 padrões alimentares e suspeita de desordem psicológica na mãe.

Foram excluídos da análise os indivíduos que apresentam valores ausentes ou indeterminados para qualquer variável do estudo. Para avaliar se a amostra estudada mantém a representatividade da população total (ou seja, para avaliar a inferência estatística), foi realizado teste para analisar se existe diferença significativa entre a população estudada e a excluída usando o teste qui-quadrado para variáveis nominais e o teste T para idade (variável contínua). Em todos os modelos, a variável idade foi categorizada para torná-la discreta. Os resultados são mostrados na Tabela 20 no apêndice.

#### 4.11 CONSTRUÇÃO DOS BANCOS DE DADOS

Foram realizados estudos preliminares aplicando programação genética mono-objetivo (GGP). Estes estudos obtiveram dificuldades em encontrar modelos iguais em diferentes execuções, bem como ocorreu uma grande variação na acurácia nas diferentes execuções. As modificações realizadas nas variáveis no estudo da MGGP em relação à GGP foram realizadas com o intuito de:

- Ampliar o espaço de busca para a possível formação de modelos com melhor acurácia.
- Diversificar o conjunto de características potencialmente relacionados com asma e atopia, com a inserção de características nutricionais e saúde mental materna.
- Escolher variáveis que representem melhor uma característica potencialmente relacionada a atopia e asma.

Em todos os experimentos realizados para cada desfecho, foi realizada uma única separação dos grupos para evitar que as diferenças encontradas entre as técnicas estudadas (MGGP, RL e C4.5) sejam originárias de variações geradas por separações diferenciais entre os grupos. Uma vez definidos os grupos para cada desfecho, estes mesmos grupos foram usados nas diferentes técnicas.

Todos os indivíduos que apresentaram dados faltantes ou indeterminados para qualquer uma das variáveis do estudo foram excluídos da análise. Foram sorteados 10% dos indivíduos (instâncias) para formar o grupo de teste. Este sorteio foi feito mantendo a frequência do desfecho no grupo igual à frequência do mesmo na população original. Dos 90% restantes, foi realizada por 6 vezes validação cruzada onde a população foi dividida aleatoriamente em 5 partes, mantendo as proporções em relação ao desfecho equivalentes ao da população original. A primeira parte da população é definida como grupo de validação e seu respectivo grupo de treinamento é constituído pelas outras 4 partes. A parte dois é definida como outro grupo de validação e seu respectivo grupo de treinamento é constituído pelas outras 4 partes. O processo continua até ter 5 grupos de validação com seus respectivos 5 grupos de treinamento.

Assim, para cada desfecho ao final, obtêm-se um único grupo de teste, onde nenhuma de suas instâncias pertence a nenhum outro grupo, 30 grupos de validação e seus respectivos 30 grupos de treinamento. Todos os grupos apresentam frequências relativas do desfecho similares às da população original. Os mesmos grupos foram utilizados em todas as análises.

Os grupos de treinamento foram usados para gerar soluções pelas técnicas estudadas, os grupos de validação foram usados para escolher as melhores soluções geradas pelas técnicas, e o grupo de teste é um grupo único, o qual contém dados que não foram usados nem para gerar os modelos nem para selecionar os melhores modelos, e assim este grupo é usado para avaliar o desempenho dos modelos encontrados nas diferentes técnicas.

Um dos grandes problemas observados nos estudos preliminares realizados usando uma abordagem mono-objetiva (GGP) foi a grande variação da acurácia do melhores modelos em diferentes execuções. Esta grande variação fez com que não existisse homogeneidade das soluções entre diferentes execuções, assim, gerando a cada execução um modelo diferente. Essa grande variação de soluções encontradas em diferentes execuções pode ser gerada por:

- variações aleatórias nos grupos de validação gerados por seleção.
- mudanças na acurácia geradas por diferenças na frequência do desfecho nos diferentes grupos.
- baixa convergência das soluções devido ao baixo número de gerações da GGP.

No estudo da MGGP foram realizadas mudanças para evitar variações das soluções encontradas em diferentes execuções no estudo da GGP. Para reduzir o efeito aleatório das separações dos grupos, o grupo de validação foi aumentado de 10% para 18%, estando menos sujeito a variações aleatórias geradas pela seleção dos grupos, gerando uma melhor

escolha dos modelos. Também foi fixada a frequência do desfecho em todos os grupos criados, evitando que diferenças nas frequências do desfecho geradas por fatores aleatórios pudessem causar diferenças nas acurácias encontradas nos grupos, uma vez que a frequência do desfecho tem grande impacto na acurácia dos modelos.

As diferentes técnicas foram aplicadas para os mesmos grupos. Isso possibilitou que as comparações entre as diferentes técnicas não apresentassem muita influência no viés da seleção dos grupos. Além disso, a escolha de um único grupo de teste o qual suas instâncias não foram usadas na construção de soluções em nenhuma técnica estudada, serviu de grupo de comparação ideal para observar o desempenho de cada técnica avaliada em dados desconhecidos da mesma população.

Como os dados apresentaram problemas devido ao desbalanceamento dos dados foi usada a técnica de *Oversampling* (CHAWLA, 2005). Esta técnica consiste em escolher aleatoriamente instâncias pertencentes ao grupo minoritário e replicar esta instância na base de dados, assim aumentando o número de instâncias no grupo minoritário. Assim, em todos os grupos de estudo (teste, treinamento e validação) pra todos os desfechos foi aplicado o *Oversampling* da seguinte forma: dentro do grupo minoritário foi sorteado aleatoriamente um indivíduo para ter seus dados replicados, o processo continuou até o tamanho do grupo minoritário ser igual ao tamanho do grupo majoritário. Assim, ao final do *Oversampling* o número de instâncias positivas para o desfecho é igual ao número de instâncias negativas para o desfecho, equilibrando a base de dados.

## 5 EXPERIMENTOS COMPUTACIONAIS

Este capítulo descreve os experimentos computacionais e resultados obtidos do estudo da MGGP. A Seção 5.1 explica como foram realizados os experimentos computacionais da RL. A Seção 5.2 explica como foram realizados os experimentos computacionais do C4.5. A Seção 5.3 explica como foram realizados os experimentos computacionais da MGGP. A Seção 5.4 descreve e discute os resultados obtidos.

### 5.1 EXPERIMENTOS COMPUTACIONAIS PARA REGRESSÃO LOGÍSTICA

os mesmos 30 grupos de validação e o mesmo único grupo de teste usado pelas outras técnicas estudadas; sendo estes grupos com a mesma frequência em relação ao desfecho e, posteriormente, foi aplicado a técnica de balanceamento. Foi aplicada regressão logística múltipla em cada um dos 30 grupo de treinamento. Cada modelo gerado foi aplicado no grupo de treinamento que gerou o modelo, no seu respectivo grupo de validação e no grupo de teste. Para avaliar o desempenho do modelo aplicado nos grupos, foi utilizada a acurácia. Como se quer comparar o modelo de RL com os modelos de classificação de MGGP e C4.5, foi aplicada uma função degrau no valor  $p$  que é a probabilidade de um indivíduo (instância) ser positivo para o desfecho. Esta função degrau é configurada para quando  $p > 0,5$  esta retorne um valor de 1 ou seja, este indivíduo seja classificado como positivo para o desfecho, caso contrario tenha valor 0 e seja classificado como negativo para o desfecho. Ao final da avaliação de todos os indivíduos no grupo são obtidas as classes em relação ao desfecho predito pelo modelo e este é confrontado com os valores conhecidos para obter a acurácia deste modelo no grupo.

Em cada grupo de treinamento ocorre a seleção das variáveis que devem entrar no modelo. Esta seleção é necessária, pois a inserção de todas as variáveis no modelo pode gerar problemas de superajuste (o modelo se ajusta demais aos dados e perde a capacidade de descrever dados novos com precisão), e multicolinearidade (problema causados pela inserção de duas variáveis colineares).

Na seleção das variáveis, as variáveis gênero e idade foram consideradas variáveis confundidoras *a priori*, ou seja, sempre entram nos modelos RL. Logo, todos os modelos RL apresentam ao menos o gênero e a idade. A seleção das outras variáveis que podem entrar no modelo foi realizada por *Stepwise* bidirecional (BILLINGS; VOON, 1986), ou seja, foi começado com apenas as variáveis *a priori*. Depois é testado se as outras variáveis se inseridas no modelo são significantes para este modelo (no grupo de treinamento). Se a variável com menor p-valor for inferior à 0,05 (for significativa) então ela é adicionada ao modelo. Em seguida, se alguma outra variável das que foram adicionadas perder a significância ela sai do modelo e pode ser reinserida posteriormente; este processo de revezamento de inserção e retirada de variáveis continua até que não seja possível nem

adicionar nem retirar alguma variável do modelo. As análises foram realizadas usando WEKA v3.6.11.

## 5.2 EXPERIMENTOS COMPUTACIONAIS PARA O C4.5

O C4.5 (WEBB, 1999) foi executado em cada grupo de treinamento. Assim como na RL, os modelos gerados no grupo de treinamento foram avaliados nos seus respectivos grupos de teste e validação. Para evitar superajuste, o parâmetro de número mínimo de instâncias por folha foi escolhido tal que maximiza a média de acurácia nos grupos de validação. Dessa forma escolhe-se o parâmetro que gera, em média, modelos com melhor capacidade de explicar dados novos. A escolha do parâmetro foi feita nos grupos de treinamento variando o parâmetro de 1 até o número de instâncias no grupo de validação. Em seguida, os modelos foram avaliados no seus respectivos grupos de validação. O valor do parâmetro que apresentou maior média de acurácia nos grupos de validação foi escolhido. As análises foram realizadas usando WEKA v3.6.11.

## 5.3 EXPERIMENTOS COMPUTACIONAIS PARA A MGGP

A evolução dos modelos foi realizada através do algoritmo NSGA-II descrito na Seção 3.3. As funções objetivo adotados no NSGA-II foram: (i) minimização do erro na predição do modelo, buscando gerar modelos mais acurados (ii) Minimização da complexidade do modelo, definida pelo número de terminais da representação do modelo em árvore, com isso busca-se soluções mais parcimoniosas. O objetivo da MGGP é achar modelos que apresentem melhores compromissos entre baixo erro e menor complexidade.

A MGGP foi executada 30 vezes independentemente, para cada grupo de treinamento. Uma população inicial de 500 modelos foi gerada aleatoriamente. Foi determinado através de estudos preliminares que seria necessário o número de gerações de 20.000 na MGGP.

A seleção dos modelos parentais foi feita através de torneio. Foram sorteados dois modelos aleatoriamente e o melhor deles escolhido como um dos pais. O mesmo procedimento foi repetido para a escolha de um segundo pai. Os dois pais geraram dois filhos por recombinação, cujo exemplo foi ilustrado na Figura 11. Estes filhos sofreram mutação, cujo exemplo foi ilustrado na Figura 10.

Este processo de geração de modelos filhos continua até se obter 500 modelos filhos. Estes são avaliados baseando-se no grupo de treinamento e unidos aos 500 modelos pais. Os 500 melhores modelos (baseado no grupo de treinamento) entre os modelos pais e os filhos são selecionados para compor a próxima geração.

Ao fim das 20.000 gerações obtêm-se a população final, que compreende 500 modelos. Os melhores modelos em cada execução são os modelos não-dominados em relação ao

grupo de validação.

A gramática aplicada para construir os modelos foi uma gramática livre de contexto que usa operadores de comparação ( $>$ ,  $<$ ,  $>=$ ,  $<=$ ,  $==$  e  $!=$ ), operadores lógicos ( $AND$ ,  $OR$ ,  $XOR$ ) e o operador ternário *if – then – else*. Para facilitar a avaliação dos modelos foi adotada a notação pós-fixa. A notação tradicional para expressões aritméticas, que representa uma operação binária na forma  $x + y$ , ou seja, com o operador entre seus dois operandos, é conhecida como notação infixa. Uma notação alternativa para esse tipo de expressão é a notação pós-fixa, na qual o operador é expresso após seus operandos na forma  $xy+$ . A grande vantagem da representação pós-fixa é que esta dispensa o uso dos parênteses, sendo assim mais fácil de manipular computacionalmente (VASUDEVA, 2006). A gramática adotada é composta das seguintes regras ( $R$ ):

$$\begin{aligned}
\langle expr \rangle &::= \langle expr1 \rangle \langle expr1 \rangle \langle bool \rangle \textit{if - else} \mid \langle binaryClass \rangle \\
\langle expr1 \rangle &::= \langle expr2 \rangle \langle expr2 \rangle \langle bool \rangle \textit{if - else} \mid \langle binaryClass \rangle \\
\langle expr2 \rangle &::= \langle binaryClass \rangle \langle binaryClass \rangle \langle bool \rangle \textit{if - else} \mid \\
&\quad \langle binaryClass \rangle \\
\langle bool \rangle &::= \langle varBin \rangle \langle binaryClass \rangle \langle compbin \rangle \mid \\
&\quad \langle varCat \rangle \langle CatClass \rangle \langle compcat \rangle \mid \\
&\quad \langle varNum \rangle \langle numValue \rangle \langle compCont \rangle \mid \\
&\quad \langle bool \rangle \langle bool \rangle \langle log \rangle \\
\langle compcat \rangle &::= < \mid <= \mid == \mid >= \mid > \mid != \\
\langle compbin \rangle &::= == \mid != \\
\langle compCont \rangle &::= < \mid <= \mid > \mid >= \\
\langle log \rangle &::= \textit{and} \mid \textit{or} \mid \textit{xor} \\
\langle binaryClass \rangle &::= 0 \mid 1
\end{aligned}$$

onde o símbolo “|” é usado para delimitar as múltiplas proposições, “ $\langle varBin \rangle$ ” é o conjunto de variáveis binárias, “ $\langle varCat \rangle$ ” é o conjunto de variáveis categóricas, “ $\langle CatClass \rangle$ ” é o conjunto de possíveis valores adotados por cada variável categórica, “ $\langle varNum \rangle$ ” é o conjunto de variáveis numéricas, e “ $\langle numValue \rangle$ ” é o conjunto de valores possíveis para cada variável numérica.

Pode ser observado que a gramática adota um limite no número de operadores *if – else* aninhados em 3, esta escolha foi outra modificação realizada na MGGP em relação à GGP. A limitação destes operadores aninhados tem por objetivo evitar que ocorra uma subdivisão da população que leva à aplicação de uma nova regra a uma subpopulação muito pequena, gerando operações com pouco significado real e limitada a uma subpopulação muito restrita, assim aumentando a possibilidade de superajuste dos modelos ao grupo de treinamento.

Foi adotado um limite máximo (de 25) na altura da representação do modelo na

forma de árvore. Este limite evita que se formem modelos excessivamente grandes, os quais têm difícil interpretação e podem impactar no desempenho da MGGP nas últimas gerações já que têm alto custo computacional para serem avaliados. O MGGP foi implementado na linguagem de programação c++.

## 5.4 RESULTADOS OBTIDOS

A população de estudo foi formada por 1445 crianças/adolescentes das quais 1047 apresentaram resultados válidos (não foram excluídas da análise por algum dos critérios descritos na Seção 4.10). As variáveis estudadas, bem como suas frequências, são descritas na Tabela 2.

Esta população de estudo teve alta prevalência de asma 22.9%, SPT 30.3% e IgE 38.6%, maior prevalência de asma e atopia do que a maioria dos países em desenvolvimento e muitos países desenvolvidos (ASHER et al., 2006).

### 5.4.1 Resultados obtidos pela regressão logística

A alta prevalência dos desfechos estudados tem como consequência o número de casos positivos para essas patologias próximo do número de casos negativos. Assim, não é esperado que esta população tenha grandes problemas de desbalanceamento de dados. Entretanto, como mostrada na Figura 12, a aplicação da técnica de *Oversample* teve um efeito profundo na melhora em prever casos positivos para estas patologias, desta forma, equilibrando os erros do tipo 1 e tipo 2. Além disso, outros estudos também mostraram a importância do balanceamento de dados nos algoritmos de classificação em estudos epidemiológicos (RAMEZANKHANI et al., 2016; LEE, 2014).

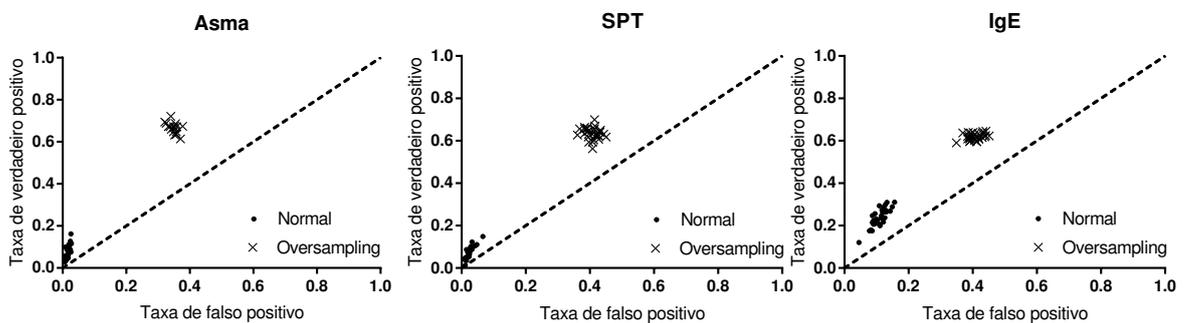


Figura 12 – Espaço ROC para os resultados obtidos por RL nos grupos de treinamento, mostrando as diferenças obtidas antes e após a aplicação da técnica de balanceamento *Oversampling*.

Não foram encontradas evidências de que a distribuição das acurácias dos modelos

Tabela 2 – Variáveis usadas no estudo MGGP

Variáveis <i>N</i> = 1046	Tipo Variáveis desfecho	Frequência %
IgE (positivos)	binárias	38.6%
SPT(positivos)	binárias	30.3%
Asma (positivos)	binárias	22.9%
Variáveis de exposição		
Gênero (masculinos)	binárias	52.7%
Idade	categórico	
4 e 5		35.9%
6 e 7		35.1%
8 até 11		29.0%
Asma nos pais (presença)	binárias	12.6%
HSV (positivos)	binárias	54.9%
VZV (positivos)	binárias	45.8%
EBV (positivos)	binárias	88.4%
HAV (positivos)	binárias	16.7%
<i>T. gondii</i> (positivos)	binárias	18.4%
<i>H. pylori</i> (positivos)	binárias	27.6%
<i>A. lumbricoides</i> (positivos)	binárias	16.2%
<i>T. trichiura</i> (positivos)	binárias	11.2%
número de irmãos	categórico	
nenhum		18.9%
1		35.2%
2		24.0%
3 ou mais		21.9%
Frequentou creche (sim)	binárias	15.4%
Fumante em casa (presença)	binárias	27.1%
Rede de esgoto (presença)	binárias	83.5%
Frequência da troca da roupa de cama ( $\geq 1$ por semana)	binárias	45.0%
Gato em casa (presença)	binárias	17.6%
Cachorro em casa (presença)	binárias	39.8%
Mofa ou umidade nas paredes da casa (presença)	binárias	68.6%
Água tratada (presença)	binárias	91.9%
Pavimentação da rua (ausência)	binárias	35.1%
Mosca em casa (presença)	binárias	51.5%
Desordem psicológica materna (suspeita)	binárias	37.2%
Padrões de dieta 1 à 4	categórica por tercil	
Calorias diárias ( <i>média(desvio padrão)</i> )Kcal	numérico	2210(929)
IMC (índice de massa corpórea)	categórico	
sobrepeso / obesidade		12.2%
eutrófico		75.1%
baixo peso		12.7%
GINI (índice econômico)	categórico por tercil	

Tabela 3 – Teste de normalidade *Shapiro-Wilk* das acurácias na RL \* Valores referente à 30 execuções da RL, resultados de p-valor menores do que 0,05 significa distribuição diferente da normal.

Modelos	Grupo	p-valor*
Asma	treinamento	0,153
	validação	0,861
Asma atópica	treinamento	0,215
	validação	0,226
Asma não atópica	treinamento	0,927
	validação	0,085
SPT	treinamento	0,504
	validação	0,957
IgE	treinamento	0,657
	validação	0,275

obtidos na RL apresentam uma distribuição diferente da normal, como mostrado na Tabela 3. Assim a média das acurácias dos modelos em cada grupo pode ser considerada uma boa medida de avaliação para estes modelos.

Os resultados dos 30 grupos de treinamentos e suas avaliações nos grupos de validação e teste são mostrados na Tabela 4. Entre os desfechos estudados, aqueles que geraram modelos com maior média de acurácia na RL do grupo de teste foram o SPT com IgE como possível exposição, seguido por asma atópica. A alta acurácia dos modelos “SPT com IgE” era esperada pois a presença de IgE contra alérgenos é necessária para ser positivo contra SPT para o mesmo alérgeno. Desta forma, tanto SPT quanto IgE fazem parte da mesma via causal, sendo assim, a IgE não deveria ser usada como exposição junto com outras exposições. Contudo trabalhos recentes têm mostrado que existe dissociação da IgE com SPT pois existem pessoas que são positivas para IgE e negativas para SPT (COOPER et al., 2008; PONTE et al., 2011). Assim, para um melhor entendimento da atopia, é necessário entender quais fatores são capazes de, na presença de IgE contra alérgenos, inibir a reatividade cutânea (SPT) e assim inibir os sintomas da alergia. Uma das formas de encontrar estes fatores inibitórios é por intermédio de inserção da IgE como uma exposição.

A asma atópica obteve uma boa acurácia média no grupo de validação. Contudo, as acurácias obtidas pelas execuções apresentaram um alto desvio padrão. Este alto desvio padrão pode ser originário do pequeno número de instâncias estudadas  $N = 404$ , pois para estudar o fenótipo de asma atópica foi usado como grupo de referência atópicos não asmáticos, ou seja, todas as instâncias não estudadas eram atópicas. Foi usado este grupo de referência, pois é difícil discernir se o grupo asmático atópico em relação ao não asmático atópico apresentaria este efeito por causa da atopia ou da asma. Usando como referência o grupo não asmático e atópico, é observado apenas o efeito da asma na presença da atopia. Assim, considerando apenas os atópicos, tem-se um  $N = 404$  e um

baixo tamanho amostral, o que aumenta o erro, implicando no maior desvio padrão da acurácia dos modelos. A maioria dos modelos para todos os desfechos tiveram 5 ou mais exposições.

Tabela 4 – Resultados obtidos da execução da RL - Mostra a média, mediana, desvio padrão, máximo e mínimo da acurácia obtida de 30 execuções com a média da especificidade e sensibilidade.

Desfecho	média	mediana	SD	max	min	sensibilidade	especificidade
IgE	treinamento	60,6058	1,1567	63,3909	58,5683	61,8401	39,7234
	validação	55,0468	3,3412	60,5263	47,4137	54,5322	55,5627
	teste	55,4286	55,2381	2,8929	63,8095	50,4762	60,8205
SPT	treinamento	61,3473	1,8155	64,4084	57,6996	63,5099	59,1848
	validação	54,3444	4,4173	65,1515	45,4545	53,2136	55,4751
	teste	54,1905	55,2381	3,5409	60,0000	44,7619	53,6530
SPT com IgE	treinamento	82,7489	0,8103	84,7908	81,3931	83,2888	82,2090
	validação	81,2634	81,2500	2,5180	85,9848	76,1538	81,4292
	teste	76,9206	77,1429	0,9096	79,0476	74,2857	74,4292
Asma	treinamento	65,4037	1,6856	69,0721	62,1551	65,7540	65,0534
	validação	57,8221	57,8767	3,5677	65,4109	50,3424	62,5852
	teste	56,6667	56,1905	4,7412	63,8095	45,7143	57,3251
Asma atópica	treinamento	67,2355	2,7765	72,5961	59,8557	68,4935	65,9775
	validação	53,1730	53,8461	6,7406	64,4230	39,4230	59,6153
	teste	72,0833	72,5000	6,8643	82,5000	57,5000	74,8276
Asma não atópica	treinamento	70,5651	70,8333	2,0780	75,4032	65,4569	69,4833
	validação	58,0122	58,0654	6,3042	67,0212	39,3617	65,3893
	teste	57,3958	57,8125	4,7209	67,1875	50,0000	58,3974

#### 5.4.2 Resultados obtidos pelo algoritmo C4.5

O algoritmo C4.5 é baseado na ideia de árvore de decisão. Na árvore de decisão cada escolha tomada divide a amostra em sub-amostras menores as quais são submetidas a novas decisões repartindo ainda mais as sub-amostras. Ao final de algumas decisões tomadas, os grupos gerados apresentam poucas instâncias, que pouco significado têm no entendimento da população. Assim, a tomada de decisão nestes grupos pequenos representa mais as características individuais da amostra estudada do que características gerais da população que se procura entender, fazendo com que apesar da capacidade de predição dos dados na base de dados ao qual foi aplicado a árvore de decisão ser muito alta, a capacidade de descrever novos dados usualmente é muito baixa. Este problema é chamado de superajuste e existem várias estratégias para evitar e combater este problema.

A estratégia adotada para evitar o problema de superajuste do C4.5 neste trabalho foi estabelecer um limite no número mínimo de instâncias as quais podem ser aplicada uma nova regra. Assim, cada regra só pode ser aplicada em uma sub-amostra com um número mínimo de instâncias. Como um grupo maior de instâncias é mais representativo em relação à população, então uma decisão adequada neste grupo tem maior probabilidade de ser uma decisão adequada para toda a população, evitando assim problemas de superajuste.

A fim de escolher do número mínimo de instâncias por nó, foi aplicado o C4.5 em todos os grupos de treinamento variando este parâmetro de 1 até o metade do número de instâncias do grupo de treinamento. Foram escolhidos os parâmetros que obtiveram maior média de acurácia nos seus respectivos 30 grupos de validação. As acurácias médias em relação ao número mínimo de instâncias por nó são mostradas nas Figuras 13 até 18.

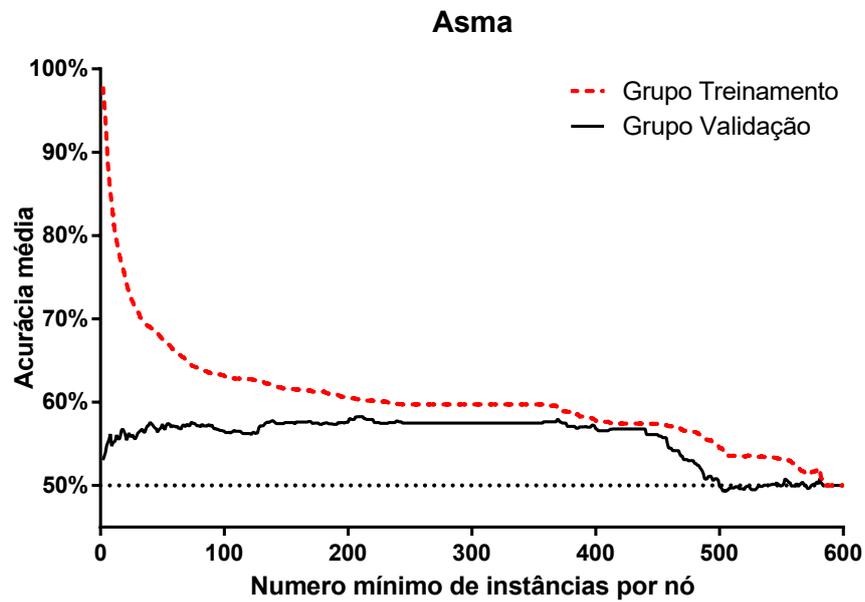


Figura 13 – Relação entre acurácia média nos grupos e a escolha do parâmetro para asma em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 210.

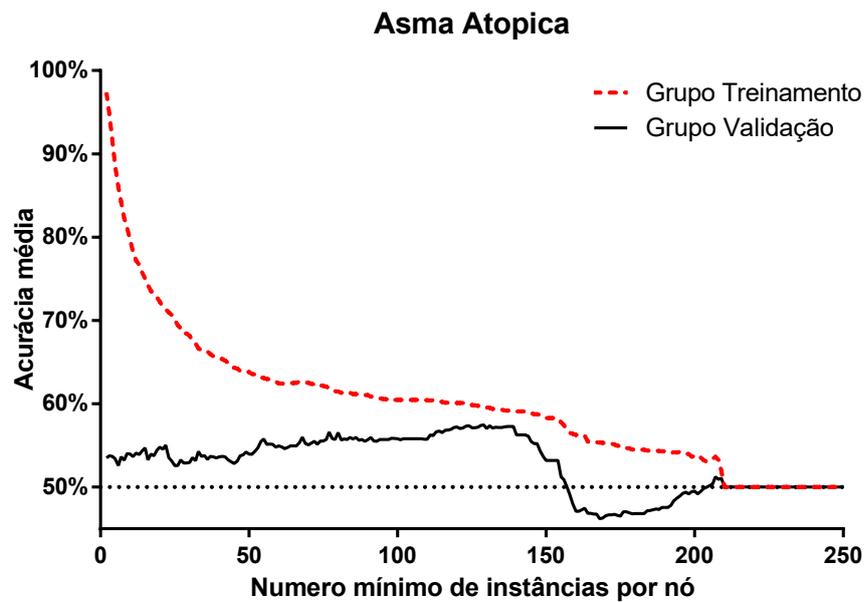


Figura 14 – Relação entre acurácia média nos grupos e a escolha do parâmetro para asma atópica em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 128.

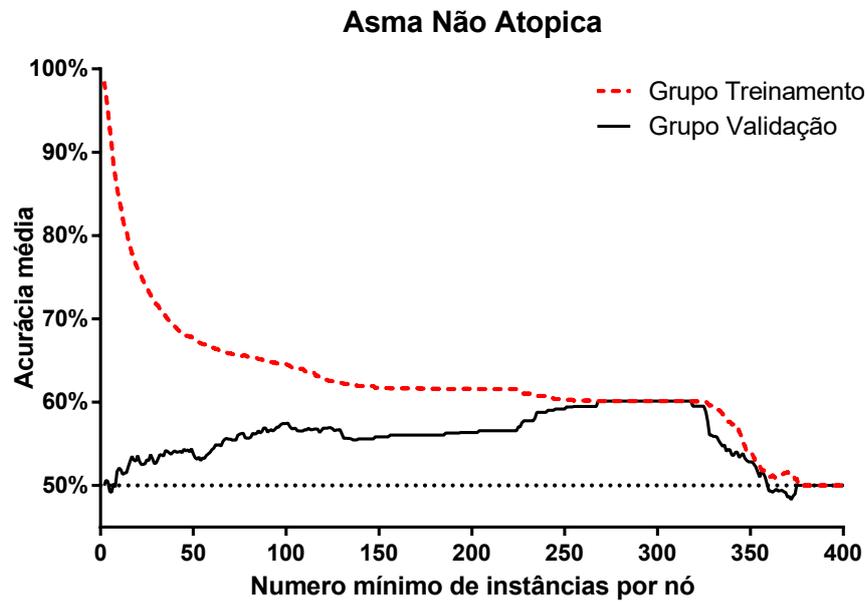


Figura 15 – Relação entre acurácia média nos grupos e a escolha do parâmetro para asma não atópica em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 300.

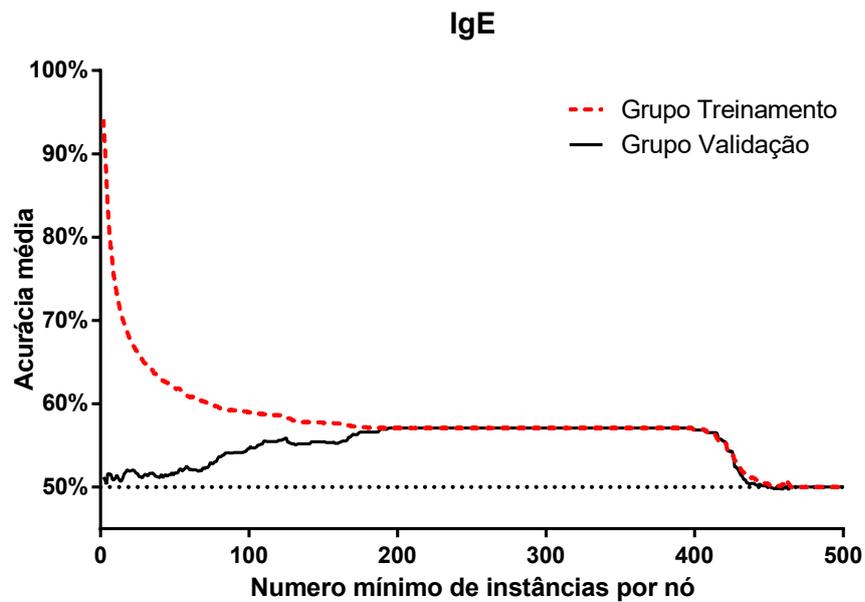


Figura 16 – Relação entre acurácia média nos grupos e a escolha do parâmetro para IgE em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 200.

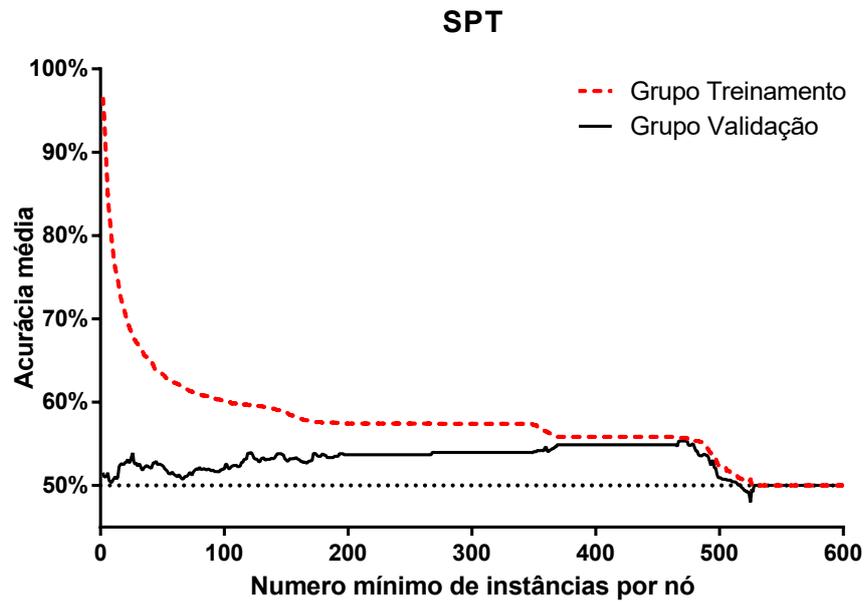


Figura 17 – Relação entre acurácia média nos grupos e a escolha do parâmetro para SPT em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 473.

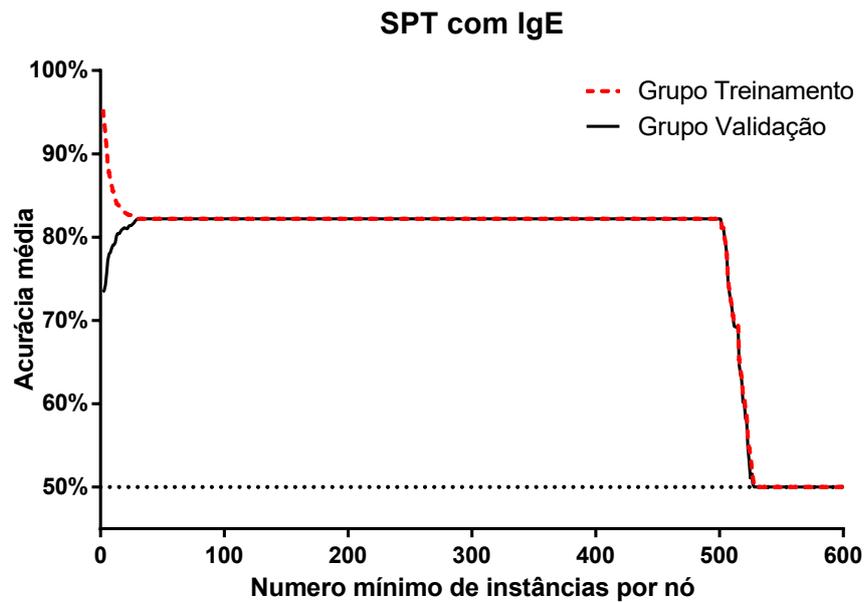


Figura 18 – Relação entre acurácia média nos grupos e a escolha do parâmetro para SPT com IgE em C4.5. O valor do parâmetro escolhido que foi usado nas análises foi 32.

As Figuras 13 até 18 indicam a necessidade de controlar o efeito de superajuste, pois quando o parâmetro número mínimo de instâncias é próximo de um, a acurácia média no grupo de treinamento fica acima de 95%, contudo, a acurácia média no grupo de validação é um pouco acima de 50%. Assim apesar de ter uma excelente capacidade em prever as instâncias do grupo de treinamento, esses modelos tem um pobre desempenho em prever dados novos como no caso do grupo de validação.

Os parâmetros escolhidos para cada desfecho foram: 210 para asma, 128 asma atópica, 300 asma não atópica, 200 IgE, 473 para SPT e 32 para SPT com IgE como possível exposição. Contrariamente ao que aconteceu na RL, os modelos gerados pelo C4.5 utilizaram poucas variáveis de exposição na construção de seus modelos, gerando modelos pequenos e simples, com uma ou duas variáveis de exposição.

Os resultados das execuções do C4.5 são mostrados na Tabela 5. As 30 execuções referentes a cada grupo de treinamento convergiram para um único modelo. Conseqüentemente, o desvio padrão no grupo de teste foi zero. Estas convergências ocorreram para os desfechos IgE, SPT com IgE como possível exposição e asma não atópica.

As comparações das médias das acurácias obtidas das técnicas C4.5 e RL são mostradas na Tabela 6. As médias das acurácias do C4.5 foram maiores do que RL para os desfechos IgE e SPT com IgE como possível exposição. Contudo, para SPT e asma não atópica, as médias das acurácias da RL foram superiores.

Tabela 5 – Resultados obtidos da execução da C4.5 - Mostra a média, mediana, desvio padrão, máximo e mínimo da acurácia obtida de 30 execuções com a média da especificidade e sensibilidade.

Desfecho	média	mediana	SD	max	min	sensibilidade	especificidade
IgE	treinamento	57,091	0,949	58,568	55,639	61,323	52,859
	validação	57,099	2,958	63,157	52,586	61,338	52,860
	teste	53,333	0,000	53,333	53,333	57,500	50,769
SPT	treinamento	55,646	1,368	58,460	51,908	53,062	58,230
	validação	55,386	4,312	65,909	46,590	52,364	58,409
	teste	50,381	0,710	51,428	46,666	47,708	51,552
SPT com IgE	treinamento	82,215	0,961	84,410	79,657	83,028	81,402
	validação	82,216	2,719	87,307	75,378	83,030	81,402
	teste	77,142	0,000	77,142	77,142	84,375	73,972
Asma	treinamento	60,330	1,331	62,457	57,474	56,259	64,401
	validação	58,252	4,679	70,833	49,305	52,425	64,079
	teste	61,968	3,158	66,666	55,238	54,861	64,074
Asma atópica	treinamento	59,639	1,257	62,740	57,692	58,766	60,512
	validação	57,467	4,507	67,307	50,000	55,705	59,230
	teste	60,583	10,521	72,500	35,000	61,515	60,229
Asma não atópica	treinamento	60,138	2,025	63,903	56,317	55,471	64,806
	validação	60,131	61,297	72,340	44,021	55,460	64,801
	teste	45,312	0,000	45,312	45,312	25,000	50,000

Tabela 6 – Comparação das acurácias obtidas por RL e C4.5 - Comparação dos resultados RL e C4.5, valores em negritos são significantes à 95% usando Teste T para dados independentes. \* Todas as execuções de C4.5 convergiram para um único modelo.

Desfecho	Grupos	RL	C4.5	p-valor
		média	média	
IgE	treinamento	60,605	57,091	0,360*
	validação	55,046	57,099	0,586*
	teste	55,428	53,333	<b>&lt;0,001*</b>
SPT	treinamento	61,347	55,646	<b>0,038</b>
	validação	54,344	55,386	0,940
	teste	54,190	50,381	<b>&lt;0,001</b>
SPT com IgE	treinamento	82,748	82,215	0,609*
	validação	81,263	82,216	0,562*
	teste	76,920	77,142	<b>0,040*</b>
Asma	treinamento	65,403	60,330	0,191
	validação	57,822	58,252	0,258
	teste	56,666	61,968	<b>&lt;0,001</b>
Asma atópica	treinamento	67,235	59,639	<b>0,012</b>
	validação	53,173	57,467	0,078
	teste	72,083	60,583	<b>&lt;0,001</b>
Asma não atópica	treinamento	70,565	60,138	0,825*
	validação	58,012	60,131	0,225*
	teste	57,395	45,312	<b>&lt;0,001*</b>

### 5.4.3 Resultados MGGP

A altura máxima da árvore foi utilizada para limitar a complexidade da GGP e MGGP, contudo, árvores com mesma altura podem gerar modelos com um número de operações muito variado, pois árvores com poucos “*ramos*” usualmente gera modelos com poucas operações e árvores com muitos “*ramos*” geram modelos com muitas operações mesmo se tiver mesma altura máxima.

Ao usar a altura máxima da árvore como um dos critérios de avaliação multiobjetivo, foi observado que alguns dos modelos gerados apresentaram operações as quais tinha pouca ou nenhuma influência na acurácia. Essas operações não foram selecionadas negativamente pela evolução porque apenas representavam o acréscimos de novos ramos sem alterar a altura máxima e, assim, não influenciaram na seleção.

O número de terminais da representação dos modelos por meio de árvore foi outra medida de complexidade testada como critério de avaliação multiobjetivo. O número de terminais avalia o número de operações e variáveis usadas no modelo. O MGGP usando este critério de complexidade gerou modelos com algumas relações as quais tiveram pouca ou nenhuma influência na acurácia.

Outra diferença importante é que os modelos gerados usando o critério de número de terminais obtiveram uma variação de valores possíveis muito maior do que o critério de altura máxima da árvore. Enquanto os valores obtidos usando altura máxima da árvore usualmente variaram entre 4 e 20, os valores obtidos pelo número de terminais usualmente variaram entre 1 e 300. Assim o número de valores possíveis para complexidade usando altura máxima da árvore foi muito menor quando se usa número de terminais. AS Figuras 19 e 20 mostram as soluções não dominadas em uma execução utilizando o desfecho asma aplicando os critérios de complexidade tamanho árvore e número de terminais respectivamente. Pode-se observar que o número de soluções não dominadas quando se utiliza o critério número de terminais, é muito maior do que quando se utiliza tamanho da árvore. Isso ocorre porque diferenças entre complexidade como número de terminais as quais não representam diferenças na altura máxima podem propiciar a formação de um modelo mais acurado o qual não domina nem é dominado pela solução menos complexa.

Foi observado uma maior diversidade de soluções não dominadas quando se utiliza o número de terminais. Também foi observado que o uso de número de terminais proporcionou um menor erro de classificação principalmente transcorrido muitas gerações em relação às soluções encontradas quando se usa o critério de tamanho máximo da árvore.

Ao se observar outras execuções utilizando o desfecho asma, foi constatado que o desempenho do MGGP quando utiliza o critério número de terminais foi superior aos MGGP usando tamanho máximo da árvore. A Figura 21 mostra que os modelos obtidos

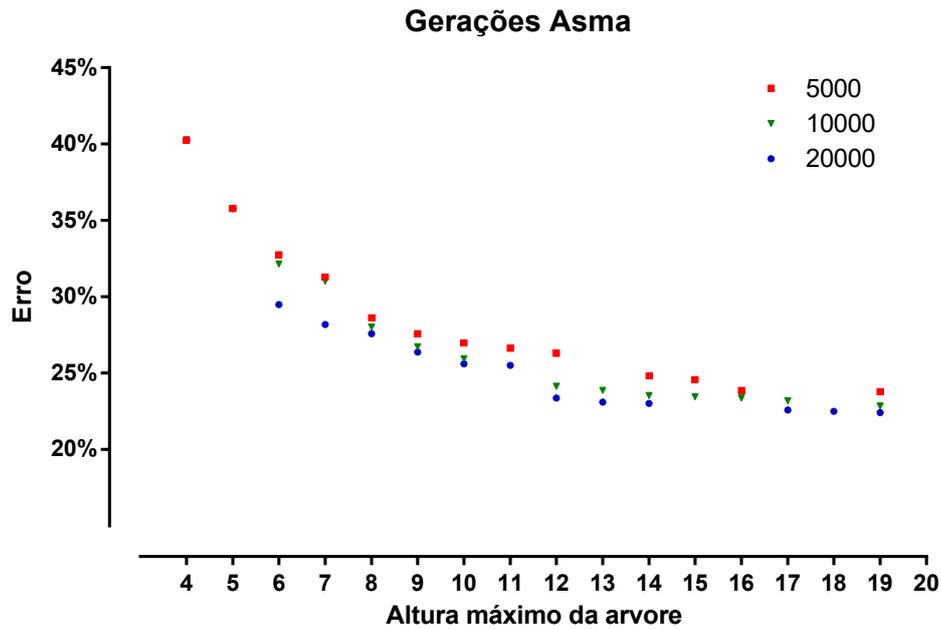


Figura 19 – Relação entre a complexidade medida através do tamanho da árvore e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma em uma execução por três diferentes gerações representadas por diferentes cores.

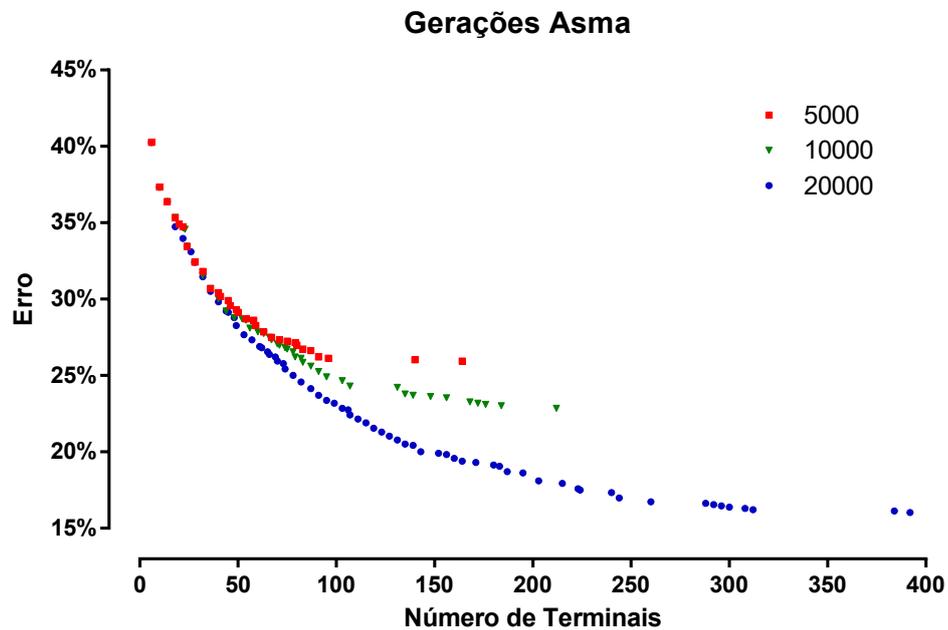


Figura 20 – Relação entre a complexidade medida através do número de terminais e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma em uma execução por três diferentes gerações representadas por diferentes cores.

por MGGP usando número de terminais obteve acurácias significativamente maiores do que usando tamanho máximo da árvore.

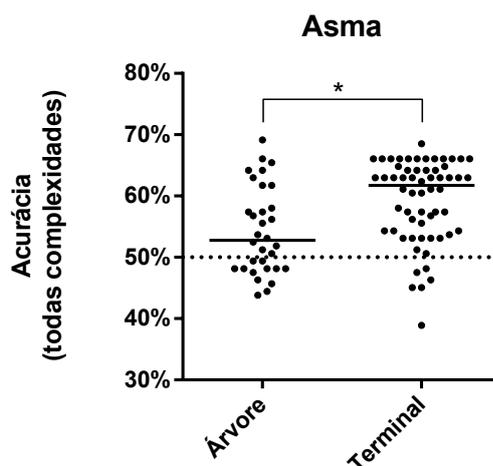


Figura 21 – Acurácia no grupo de teste das melhores soluções MGGP em todas execuções com desfecho asma utilizando os critérios tamanho máximo de árvore e número de terminais. A barra horizontal representa a mediana das soluções. \* diferença significativa entre os grupos, teste de Mann-Whitney com p-valor < 0,05.

Uma possível explicação para o melhor desempenho do MGGP quando utiliza número de terminais é que com a utilização do número de terminais existe um maior número de complexidades possíveis, e isso implicou em uma maior diversidade de soluções ótimas as quais puderam contribuir com suas características favoráveis possibilitando a evolução de soluções melhores. Assim, o uso de número de terminais como critério de complexidade obteve as vantagens de:

- gerar soluções com melhor acurácia;
- gerar maior diversidade das soluções ótimas;
- gerar soluções com menor número de relações de baixa ou nenhuma influência na acurácia;

Os erros de treinamento e complexidade das soluções não dominadas nas diferentes gerações em uma execução também é mostrado para IgE, SPT, SPT com IgE como potencial exposição, asma atópica e asma não atópica são mostrados nas Figuras 22, 23, 24, 25 e 26, respectivamente.

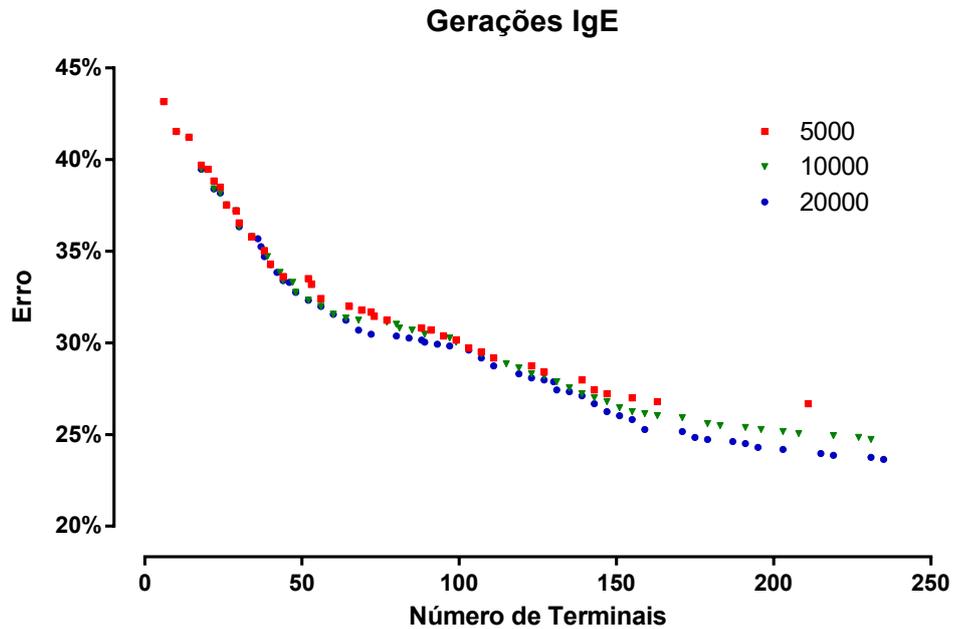


Figura 22 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para IgE em uma execução por três diferentes gerações representadas por diferentes cores.

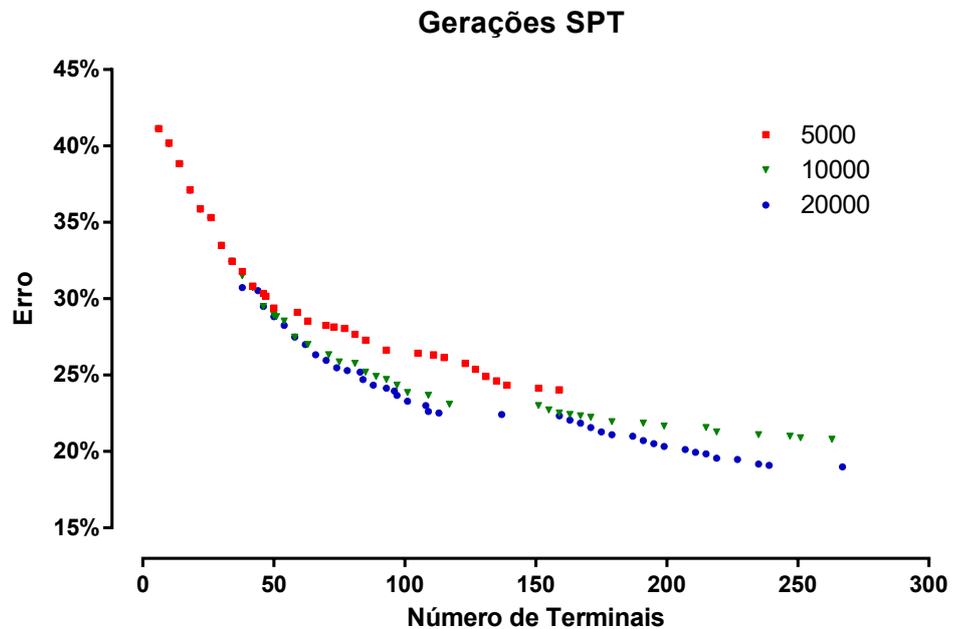


Figura 23 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para SPT em uma execução por três diferentes gerações representadas por diferentes cores.

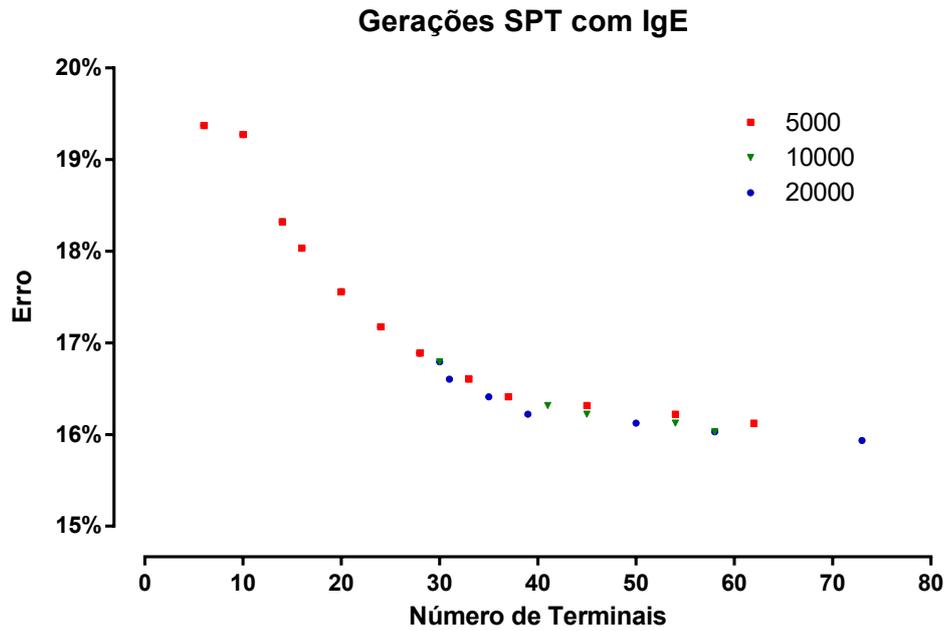


Figura 24 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para SPT com IgE como possível exposição em uma execução por três diferentes gerações representadas por diferentes cores.

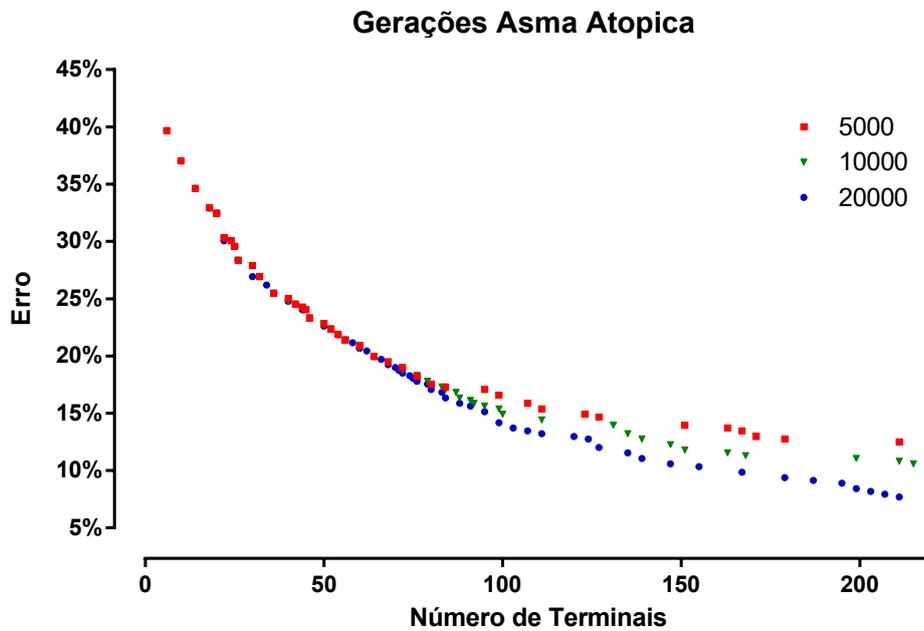


Figura 25 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma atópica em uma execução por três diferentes gerações representadas por diferentes cores.

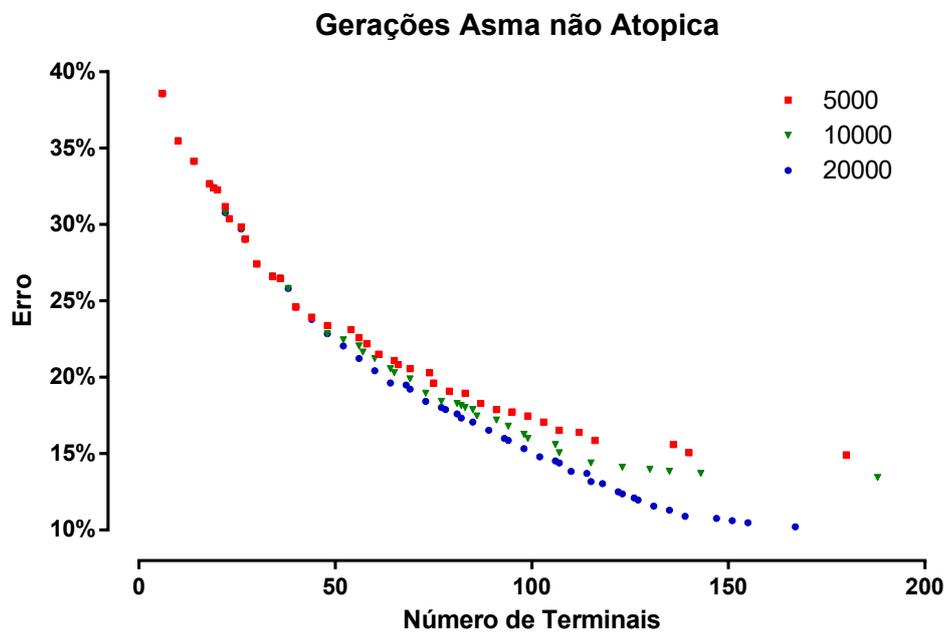


Figura 26 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma não atópica em uma execução por três diferentes gerações representadas por diferentes cores.

Pode ser visto nas Figuras 20,22, 23, 24, 25 e 26 que todos os desfechos apresentaram ganhos em acurácia em relação à mesma complexidade quando é comparadas às gerações 5000 com 10000 e 20.000. A Figuras 20 mostra que na asma o ganho da complexidade no passar das gerações é muito evidente, também pode-se perceber que o aumento das gerações propiciou a formação de modelos mais complexos com baixo erro.

Entretanto, quando se estudou os fenótipos de asma separando em asma atópica na Figura 25 e asma não atópica na Figura 25, o crescimento na acurácia nas gerações foi relativamente pequeno em relação a asma em geral. Contudo, as magnitudes de erros tanto na asma atópica quanto na asma não atópica foi menores que a asma em geral ao fim das 20.000 gerações (porém observando uma só execução por desfecho não pode-se afirmar isso).

As Figuras 27, 28, 29, 30, 31 e 32 mostram a relação entre o erro no grupo de treinamento e a complexidade das soluções não dominadas encontradas pela MGGP em uma execução e sua comparação com os erros encontrados pelas técnicas RL e C4.5 aplicadas ao mesmo grupo de treinamento. Estas figuras são referente aos desfechos asma, asma atópica, asma não atópica, IgE, SPT, SPT com IgE como possível exposição respectivamente.

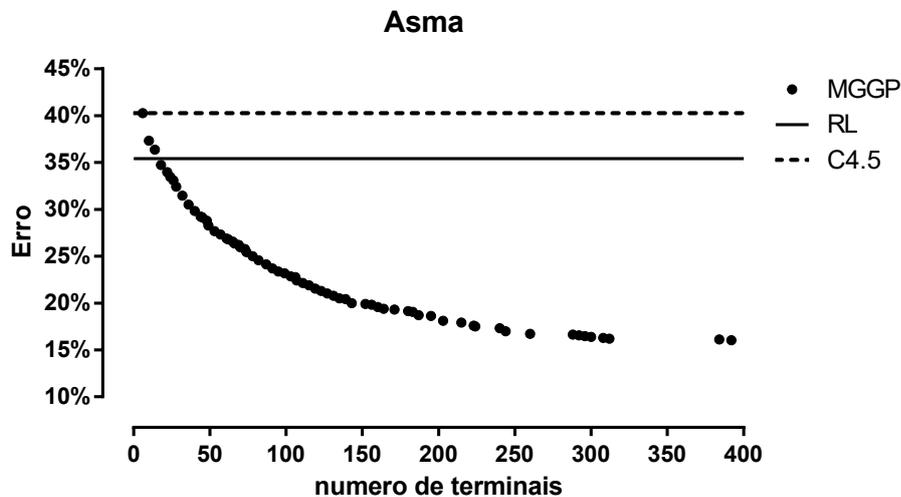


Figura 27 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

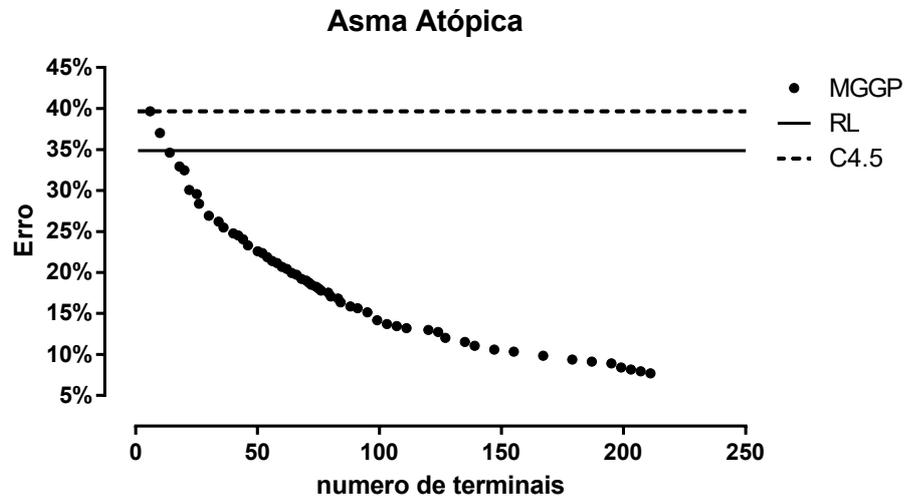


Figura 28 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma atópica em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

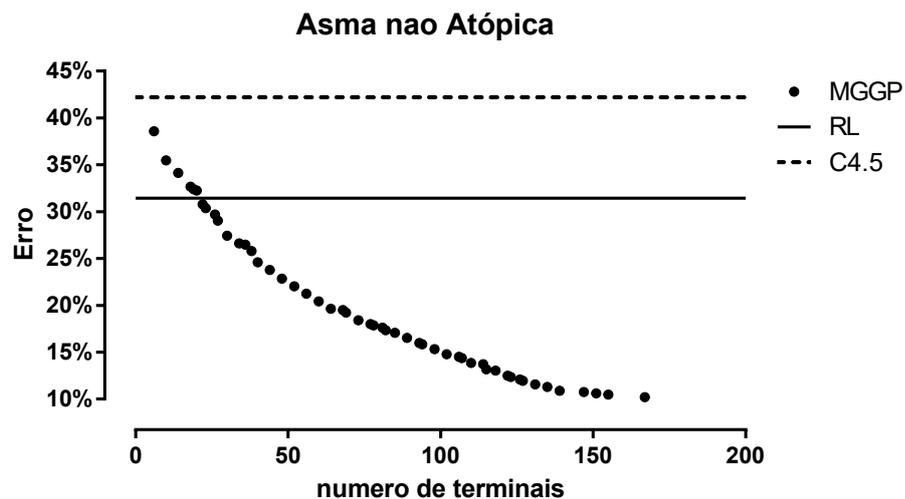


Figura 29 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para asma não atópica em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

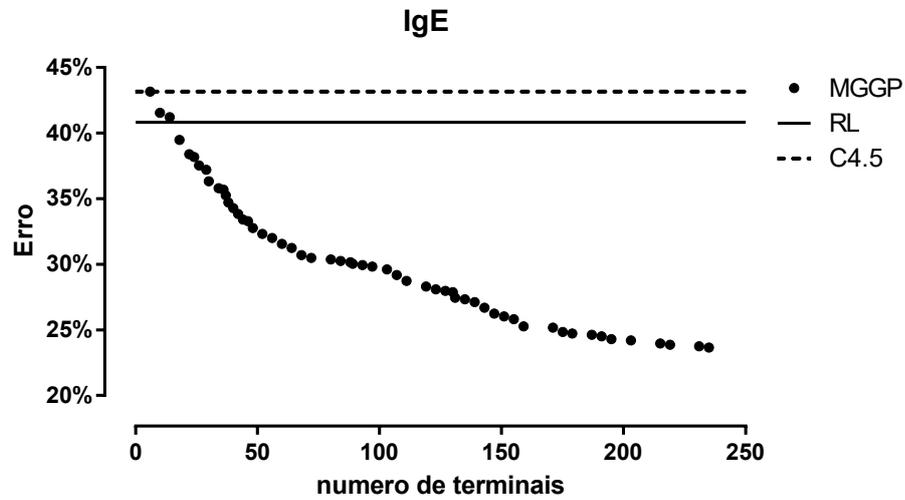


Figura 30 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para IgE em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

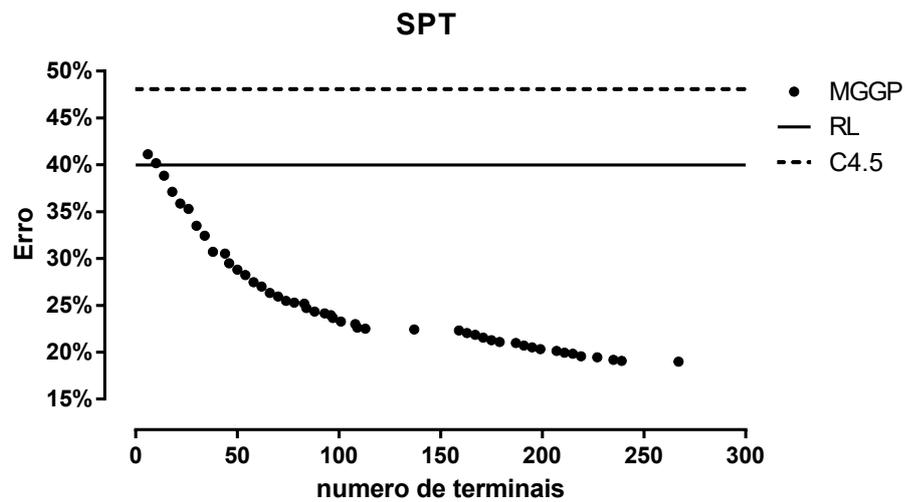


Figura 31 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para SPT em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

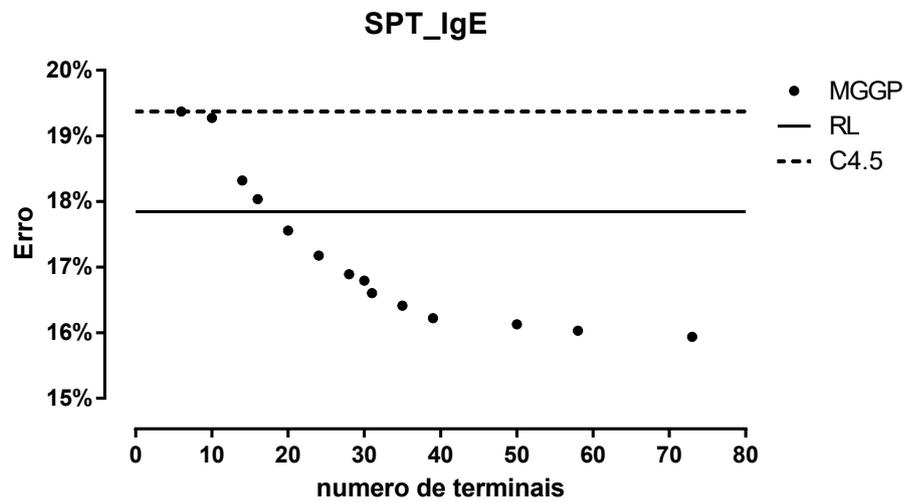


Figura 32 – Relação entre a complexidade e seu erro no grupo de treinamento em soluções não dominadas de MGGP para SPT contendo IgE como possível exposição em uma execução na população final e sua comparação com os erros obtidos por RL e C4.5.

As execuções mostradas nas Figuras 27, 28, 29, 30, 31 e 32 apresentaram um grande número de soluções não dominadas, isso aconteceu porque não foi possível obter uma solução muito acurada e ao mesmo tempo pouco complexa. Caso fosse possível obter uma solução boa para os dois critérios, esta solução dominaria boa parte das outras. Isso demonstra a importância de usar uma abordagem multiobjetivo em problemas epidemiológicos, pois quando se estuda doenças complexas e multifatoriais, é esperado encontrar diversas relações complexas que podem ser melhor entendidas quando se observa diversas relações entre complexidade das relações e acurácia dos modelos.

Mesmo com baixa complexidade o MGGP foi capaz de encontrar soluções com melhor acurácia que as encontradas por RL e C4.5, referente ao mesmo grupo de treinamento para todos os desfechos. Contudo, não dá para afirmar que as soluções encontradas por MGGP são melhores baseando-se apenas no erro de treinamento, pois pode-se estar cometendo superajuste nos dados, com isso essas soluções podem ter baixa capacidade de explicação de dados novos sendo assim pouco informativa para a população de onde a amostra foi obtida.

Uma melhor forma de avaliar o desempenho dos modelos e assim comparar as metodologias é usando o grupo de teste. Este grupo só contém instâncias que não foram usadas nem para gerar as soluções nem para selecioná-las. Além disso, este grupo é único para todas as execuções, assim diferenças encontradas nos resultados têm menos influência de viés de seleção.

As acurácias encontradas por MGGP nos grupos de teste não apresentaram uma distribuição normal como ocorreu com a RL. Foi realizado o teste de normalidade Kolmogorov-Smirnov e foi encontrado uma diferença significativa da distribuição normal para IgE, SPT e asma, todos com  $p$ -valor  $< 0,001$ . Assim, a média nas acurácias do grupo de teste não é um bom critério de decisão e foi usado a mediana como medida de tendência central do grupo teste para realização das comparações.

Para a escolha dos melhores modelos obtidos da população na geração final foi testado duas abordagens:

- Encontrar as soluções não dominadas baseado no grupo de validação entre todas as soluções obtidas na população final.
- Encontrar as soluções não dominadas baseado no grupo de validação entre as soluções não dominadas em relação ao grupo de treinamento.

Com o intuito de escolher qual melhor abordagem, foi realizada uma execução em cada um dos 30 grupos de treinamento da MGGP para os desfechos asma, SPT e IgE usando cada uma das abordagens. Os resultados são mostrados na Figura 33, 34 e 35. Pode-se observar que para a maioria das faixas de complexidade não foi encontrada

diferenças entre as duas abordagens. Porém, para asma na Figura 33 com complexidade maior que 50 números de terminais, a abordagem usando toda a população final foi significativamente maior do que usando apenas as soluções não dominadas em relação ao grupo de treinamento. Contudo, quando se observa o desfecho IgE na Figura 35, as soluções com complexidade maior que 50 número de terminais apresentou uma significativamente maior acurácia na abordagem das soluções não dominadas do que na abordagem de toda a população.

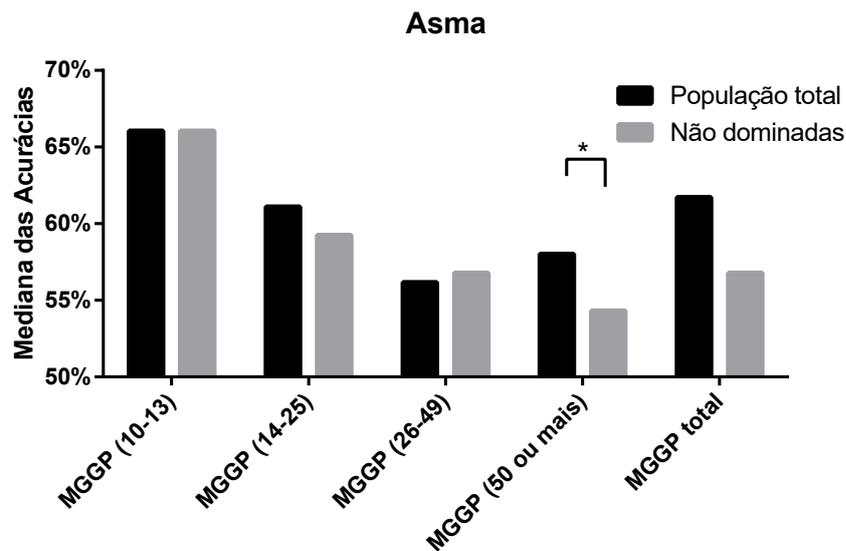


Figura 33 – Mediana da acurácias no grupo de teste das melhores soluções obtidas por MGGP para asma referente a 30 execuções usando a abordagem de escolher as melhores soluções não dominadas no grupo de validação entre todas as soluções e a abordagem de escolher as soluções não dominadas baseadas no grupo de validação entre as soluções não dominadas no grupo de treinamento. Os valores entre parênteses dos grupos representam as faixas de complexidade em número de terminais. \* Representa diferença significativa ( $p$ -valor  $< 0,05$ ) entre as abordagens usando teste Mann-Whitney.

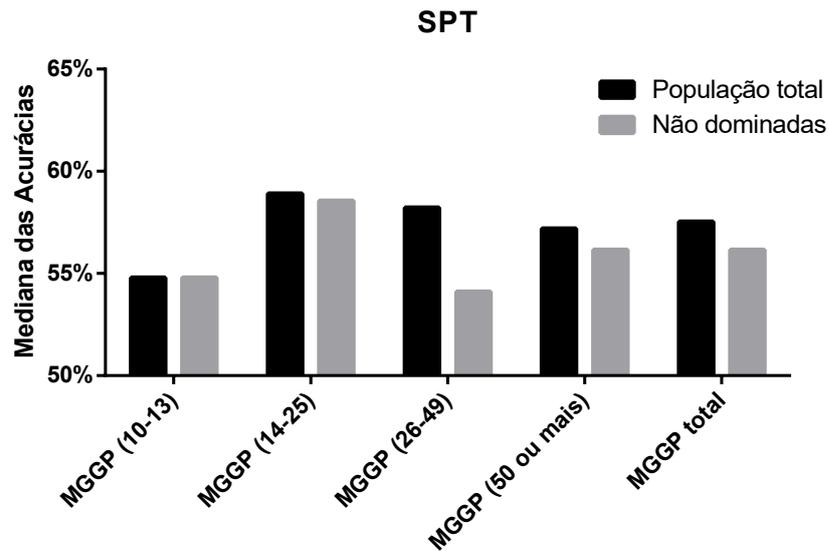


Figura 34 – Mediana da acurácias no grupo de teste das melhores soluções obtidas por MGGP para SPT referente a 30 execuções usando a abordagem de escolher as melhores soluções não dominadas no grupo de validação entre todas as soluções e a abordagem de escolher as soluções não dominadas baseadas no grupo de validação entre as soluções não dominadas no grupo de treinamento. Os valores entre parênteses dos grupos representam as faixas de complexidade em número de terminais.

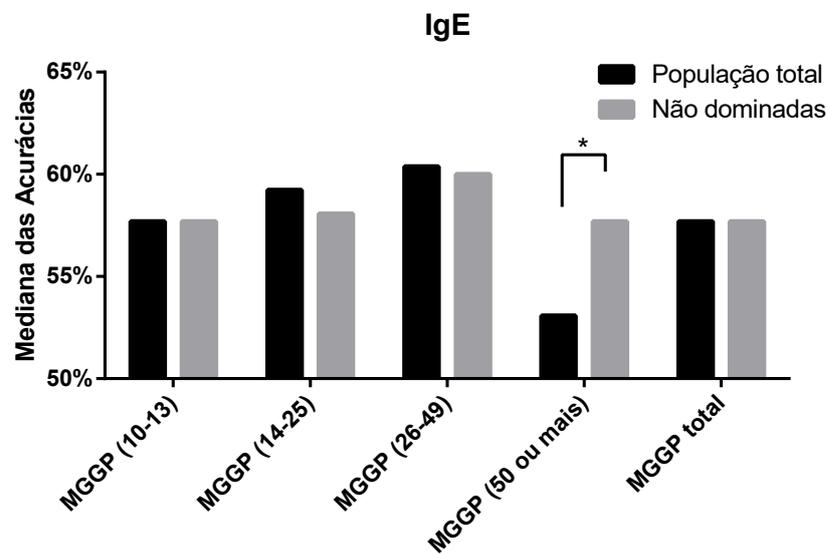


Figura 35 – Mediana da acurácias no grupo de teste das melhores soluções obtidas por MGGP para IgE referente a 30 execuções usando a abordagem de escolher as melhores soluções não dominadas no grupo de validação entre todas as soluções e a abordagem de escolher as soluções não dominadas baseadas no grupo de validação entre as soluções não dominadas no grupo de treinamento. Os valores entre parênteses dos grupos representa as faixas de complexidade em número de terminais. \* Representam diferença significativa ( $p$ -valor  $< 0,05$ ) entre as abordagens usando teste Mann-Whitney.

Apesar das abordagens apresentarem quase equilíbrio nos resultados, a abordagem que utiliza toda a população apresentou a vantagem de gerar maior diversidade de soluções e por isso esta abordagem foi escolhida para realização das subseqüentes análises.

#### 5.4.4 Resultados Asma

Os resultados obtidos por MGGP em todos os desfechos apresentaram muitas soluções que se repetiram em mais de uma execução, o que raramente aconteceu no GGP. A Figura 36 mostra a comparação por meio do erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma nos 30 grupos de treinamentos. MGGP e C4.5 para asma obtiveram significativamente menores erros do que RL. Além do baixo erro obtidos pelo MGGP e C4.5 em relação ao RL, o MGGP foi capaz de achar soluções com baixo erro em diversos níveis de complexidade.

Bornelöv et al. (2013), estudando a aplicação de seleção de característica pelo método de Monte Carlos no estudo da asma em crianças, encontraram 62% de acurácia usando informações ambientais e genéticas. O fato de que MGGP obteve valores equivalentes para asma usando somente dados ambientais e melhores resultados do que RL, mostra o potencial de MGGP para encontrar boas soluções quando aplicado a estudos epidemiológicos em patologias complexas. A Tabela 7 mostra resultados adicionais indicando que existe equilíbrio entre a sensibilidade e especificidade para todas as técnicas.

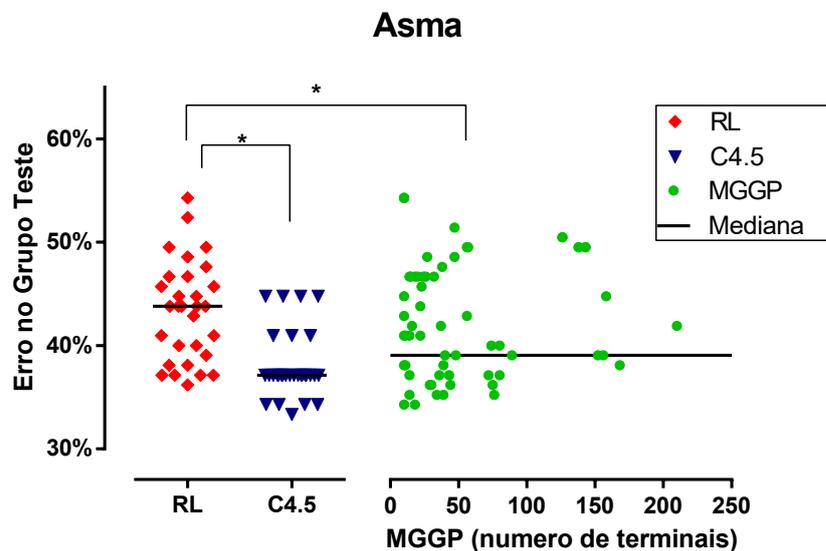


Figura 36 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma. \* Representa diferença significativa (p-valor < 0,05) entre os grupos usando teste Mann-Whitney.

A Figura 37 exibe os comparações da acurácia no grupo de teste das soluções encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho asma.

Tabela 7 – Resumo dos resultados referentes a asma - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	379,8	376	204,8	201	65,4	64,7
validação	77,8	90,9	54,3	67,4	53,5	62,6
teste	13,1	46,4	34,6	10,9	54,4	57,3
	C4.5					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	326,7	374,1	206,7	254,1	56,3	64,4
validação	76,1	93	52,2	69,1	52,4	64,1
teste	13,2	51,9	29,1	10,8	54,9	64,1
	MGGP					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	398,6	390,5	190,2	182	68,6	67,2
validação	89,7	95,2	50	55,6	61,7	65,5
teste	12,4	48,6	32,4	11,1	53,7	60

A acurácia do MGGP foi significativamente maior do que a RL nas complexidades de 10 até 13, de 14 até 25 e de 50 ou mais.

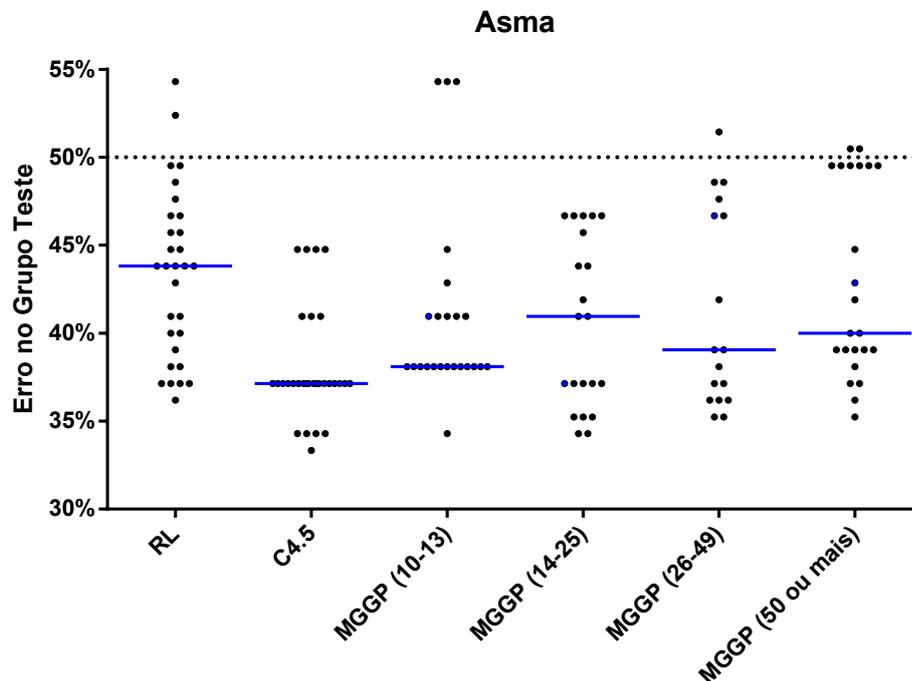


Figura 37 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma.

As principais soluções obtidas por MGGP para asma são mostradas na Tabela 8. Pode-se notar que uma característica importante para a asma que aparece em muitas relações encontradas é a idade. A asma é uma condição heterogênea com diferentes

fenótipos e expressões clínicas. Um fenótipo comum de asma é o fenótipo transitório de sibilância que não é comumente associado com história familiar de asma ou com atopia. Para esse fenótipo, os sintomas tendem a regredir aos 3-5 anos de idade (STEIN; MARTINEZ, 2004), e a alta prevalência deste fenótipo pode explicar essa relação com baixa idade.

Tabela 8 – Exemplos de relações encontradas pelo MGGP para Asma. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “{1}” significa positivo para asma, “{0}” negativo para asma. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo.

Todos indivíduos		Grupo teste		Acurácia (%)	Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)			
2,71(2,20; 3,33)	6,93(3,14; 15,30)		68,5	10	if((troca de roupa de cama = 1) ou (idade = 0)){1}else{0}	
2,74(2,24; 3,35)	3,80(2,01; 7,55)		66,0	10	if((gato = 1) ou (idade = 0)){1}else{0}	
2,70(2,18; 3,3)	3,25(1,64; 6,42)		63,0	10	if((desordem psicológica materna = 1) ou (idade = 0)){1}else{0}	
2,33(1,91; 2,85)	2,36(1,25; 4,43)		60,5	10	if((cachorro != 1) e (desordem psicológica materna = 0)){0}else{1}	
3,25(2,65; 3,98)	3,23(1,69; 6,14)		64,2	14	if((idade <= 0) ou ((gato = 1) e (fator nutricional3 <= 0))){1}else{0}	
3,23(2,63; 3,97)	2,92(1,53; 5,56)		63,0	14	if(((idade > 0) ou (fator nutricional2 = 2)) e (gato != 1)){0}else{1}	
3,47(2,81; 4,28)	2,56(1,34; 4,88)		61,1	18	if(((fator nutricional2 < 2) e (idade = 0))){1}else{0} ou ((fator nutricional2 < 2) e (idade = 0))	
3,69(3,00; 4,53)	2,50(1,32; 4,73)		61,1	22	if(((idade != 0) ou ((desordem psicológica materna = 0) e (cachorro != 1) e (VZV = 1))) e (gato != 1)){0}else{1}	
3,56(2,87; 4,40)	1,73(0,93; 3,23)		56,8	22	if(((HSV != 1) e ((troca de roupa de cama != 0) e (idade = 1))) xor ((cachorro != 0) ou (idade <= 0))){1}else{0}	
4,24(3,44; 5,22)	2,78(1,46; 5,28)		62,3	31	if(gato = 1){1}else{(if((A.humbricoides != 0) ou (fator nutricional2 < 2) e (cachorro != 0) ou ((VZV != 1) ou (desordem psicológica materna = 1) e (idade < 1)))){1}else{0}}	
14,98(11,70; 19,16)	3,77(1,91; 7,46)		64,8	231	<b>muito grande para mostrar</b>	

Asma

Algumas relações menos complexas comumente encontradas foram:

- A baixa idade ou frequente troca de roupa de cama causando asma, indicando que maior cuidado com limpeza pode estar associado com o aumento da asma.
- A presença de gato em casa pode estar aumentando a asma.
- A suspeita de transtorno psicológico da mãe também pode estar relacionada com o aumento da chance de ser asmático.

A presença de gato na casa e sua associação com asma tem apresentado resultados contraditórios na literatura, enquanto alguns trabalhos acham uma associação positivas com a asma (HEISSENHUBER et al., 2003; POLK et al., 2004) outros trabalhos encontram associação negativa (CELEDÓN et al., 2002; PERZANOWSKI et al., 2002b). Um dos motivos para tamanha discordâncias entre os trabalhos é que a presença de gato pode aumentar os sintomas da asma, com isso, é comum pais com crianças asmáticas evitarem ter gatos, o que poderia causar uma associação negativa em boa parte dos estudos.

A suspeita de transtornos psiquiátricos maternos tem sido descrito na literatura como um possível fator de risco para asma (MCQUAID; KOPEL; NASSAU, 2001; FEITOSA et al., 2011). Contudo, McQuaid, Kopel e Nassau (2001) sugerem que além dos transtornos psicológicos materno estarem aumentando as chances da criança ser asmática, o fato da criança ser asmática também pode estar aumentando a chance da mãe apresentar desordens psicológicas. Estudos longitudinais (WEIL et al., 1999; CALAM et al., 2005) mostraram que a associação permanece quando a asma precede o desenvolvimento da desordem psicológica, indicando que de fato a desordem psicológica materna é um importante fator de risco para o desenvolvimento da asma.

A asma é uma condição heterogênea, sendo a apresentação final de diversas vias causais (STEIN; MARTINEZ, 2004). Assim, é necessário estudar separadamente seus diversos fenótipos para seu melhor entendimento. Apesar de difícil discernir entre os diferentes fenótipos da asma, um importante fenótipo que pode ser estudado no presente trabalho foi a asma atópica. A Figura 38 mostra a comparação através do erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma atópica nos 30 grupos de treinamentos. O MGGP para asma atópica obteve significativamente maiores erros do que RL e C4.5. Uma possível explicação para o baixo desempenho na asma atópica é o pequeno número de instâncias usadas para executar o MGGP na asma atópica, pois para estudar este fenótipo todas as instâncias usadas no estudo são positivas para IgE o que representa 37,9% das instâncias (uma população original de N=448 instâncias). Isso indica que o baixo número de instâncias teve mais forte efeito prejudicial no desempenho da MGGP do que nas outras técnicas RL e C4.5. A Tabela 9 mostra resultados adicionais

Tabela 9 – Resumo dos resultados referentes a asma atópica - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	142,5	137,2	70,8	65,5	68,5	66
validação	24,3	31	21	27,7	46,7	59,6
teste	7,1	21,7	7,3	3,9	64,8	74,8
	C4.5					
	treinamento	122,2	125,9	82,1	85,8	58,8
validação	29	30,8	21,2	23	55,7	59,2
teste	6,8	17,5	11,5	4,2	61,5	60,2
	MGGP					
	treinamento	160,1	148,1	59,9	47,9	76,9
validação	36,1	33,8	18,2	15,9	69,3	65
teste	6,1	18,6	10,4	4,9	55,1	64,2

que mostram um baixo número de falso negativos no grupo de teste, o que pode está prejudicando a avaliação dos modelos gerados.

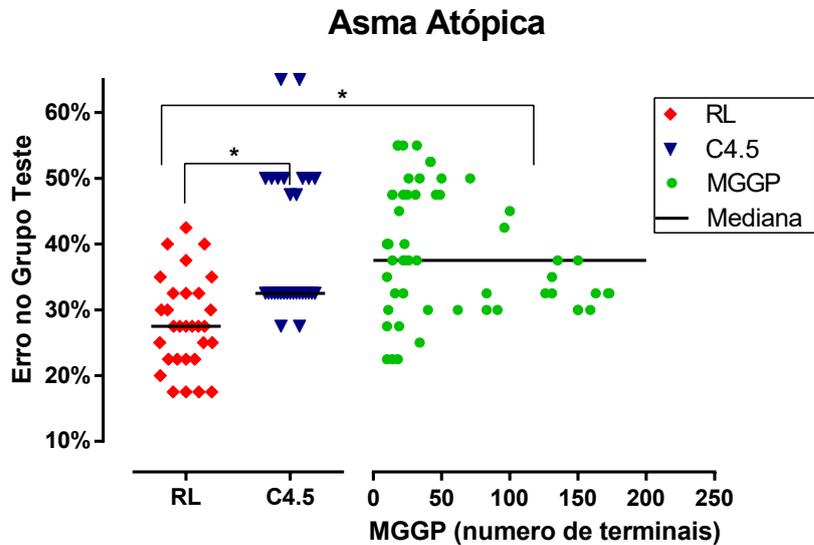


Figura 38 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma atópica. \* Representa diferença significativa ( $p$ -valor  $< 0,05$ ) entre os grupos usando teste Mann-Whitney.

A Figura 39 mostra comparações da acurácia no grupo de teste das soluções encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho asma atópica. A acurácia do MGGP foi significativamente inferior do que a RL em todas as complexidades.

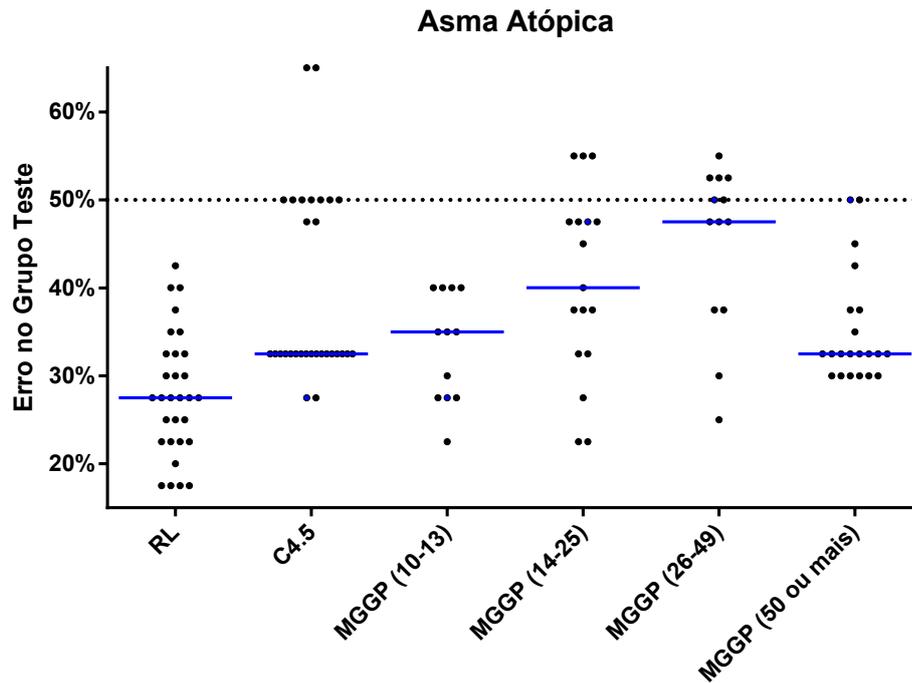


Figura 39 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma atópica.

As principais soluções obtidas por MGGP para asma atópica são mostradas na Tabela 8. Para a asma atópica a maioria das execuções encontraram resultados ruins, contudo 3 das 30 execuções encontraram bons resultados.

Tabela 10 – Exemplos de relações encontradas pelo MGGP para asma atópica. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “{1}” significa positivo para asma atópica, “{0}” negativo para asma atópica. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo

Todos indivíduos		Grupo teste		Acurácia (%)	Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)			
2,94(2,01; 4,12)	6,27(1,95; 20,22)	6,27(1,95; 20,22)	70,7	10	if((gato != 1) e (idade > 0)){ 0 }else{ 1 }	
2,81(2,01; 3,94)	1,77(0,62; 5,10)	1,77(0,62; 5,10)	56,9	14	if((troca de roupa de cama == 1) xor ((EBV != 0) xor (desordem psicológica materna == 0))){ 1 }else{ 0 }	
4,17(2,91; 5,96)	0,57(0,20; 1,62)	0,57(0,20; 1,62)	43,1	18	if((fator nutricional2 != 0) xor (((EBV == 1) xor (troca de roupa de cama == 1)) xor (desordem psicológica materna == 1))){ 1 }else{ 0 }	
5,54(3,87; 7,95)	9,12(2,66; 31,22)	9,12(2,66; 31,22)	74,1	22	if((((troca de roupa de cama != 1) xor (fator nutricional2 == 0)) ou (número de irmãos >= 2)) xor ((EBV == 1) e (desordem psicológica materna == 0))){ 1 }else{ 0 }	
7,07(4,89; 10,22)	3,61(1,22; 10,66)	3,61(1,22; 10,66)	65,5	26	if((desordem psicológica materna == 0) xor (número de irmãos <= 2) e ((HAV == 0) e (((EBV == 1) e (troca de roupa de cama == 1) ou (fator nutricional2 <= 0))){ 0 }else{ 1 }	
11,75(7,76; 17,79)	1,15(0,41; 3,26)	1,15(0,41; 3,26)	51,7	46	if((desordem psicológica materna == 0) xor ((número de irmãos != 3) e ((HAV != 1) e (((Índice econômico > 0) e (asma nos pais != 1)) e ((idade <= 1) xor (cachorro != 1)) xor (moscas != 1)) ou ((EBV == 1) xor (troca de roupa de cama != 1)) xor (fator nutricional2 == 0))){ 0 }else{ 1 }	
16,42(10,86; 24,82)	3,87(1,26; 11,88)	3,87(1,26; 11,88)	65,5	91	if((((fator nutricional1 < 2) xor ((fator nutricional2 <= 1) e (desordem psicológica materna != 1) xor (idade < 2))) e (fator nutricional1 != 2) ou (número de irmãos == 3))) e ((número de irmãos == 1) ou ((número de irmãos == 1) e ((número de irmãos == 2) xor (índice de massa corpórea >= 1)) e (mofo == 0) ou (cachorro != 1) e (desordem psicológica materna != 0)))) xor (((idade == 0) e (fator nutricional2 <= 1)) e ((EBV != 0) e (número de irmãos != 3))){ 1 }else{ if((((idade == 0) xor (rede de esgoto != 0)) e (creche != 1)) e ((Índice econômico >= 1) e (número de irmãos <= 2))){ 1 }else{ 0 } }	
33,75(21,28; 53,52)	3,72(1,24; 11,17)	3,72(1,24; 11,17)	65,5	173	<b>muito grande para mostrar</b>	

Algumas variáveis encontradas que apresentaram associação com asma atópica foram:

- asma nos pais aumentando asma atópica.
- baixo número de irmãos aumentando asma.
- fator nutricional 2 (dieta rica em leite e derivados) aumentando asma atópica.
- frequente troca de roupa de cama aumentando asma atópica.
- HAV e EBV

Os resultados estão de acordo com o esperado quando comparado aos achados obtidos na literatura. Stein e Martinez (2004) sugerem que asma atópica tem forte influência genética. Strachan (1989) mostrou que número de irmãos mais velhos está associado com alergia e sugeriu que essa associação se deve ao maior cuidado com a higiene que se tem com os primeiros filhos e menor número de infecções.

Matricardi et al. (2000) estudando cadetes italianos mostraram que a presença de HAV protege contra atopia. O EBV apresenta em seu genoma uma sequência homóloga ao da IL-10 (WINGATE et al., 2009) e também existe indícios de que este pode imunomodular a resposta do hospedeiro (MARSHALL et al., 2007), assim é esperado que reduza a atopia. Existem evidências na literatura que mostram que o EBV reduz SPT (VEIGA et al., 2011; ALCANTARA-NEVES et al., 2012).

Cunha et al. (2010) mostraram que no Brasil a proporção de asma atribuível a atopia é de 24,5% quando considerando atopia sendo a presença de IgE específica. Isso indica que uma parte da asma na presença de atopia não tem relação com atopia, ou seja, parte da população classificada como asma atópica pode ser de asma não atópica.

Outro fenótipo que foi possível estudar neste trabalho foi a asma não atópica. A Figura 40 mostra a comparação através do erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para asma não atópica nos 30 grupos de treinamentos. O MGGP para asma não atópica não obteve diferenças significativas em relação a RL e o desempenho do C4.5 foi muito ruim, significativamente pior que RL e MGGP. A Tabela 11 mostra resultados adicionais indicando que o baixo desempenho do C4.5 se deve a baixa sensibilidade causada pelo excesso de falsos positivos.

A Figura 41 exhibe as comparações da acurácia no grupo de teste das soluções encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho asma não atópica. A acurácia do MGGP não apresentou diferenças significativas em relação à RL. O desempenho de C4.5 foi significativamente inferior que RL e todas as complexidades da MGGP.

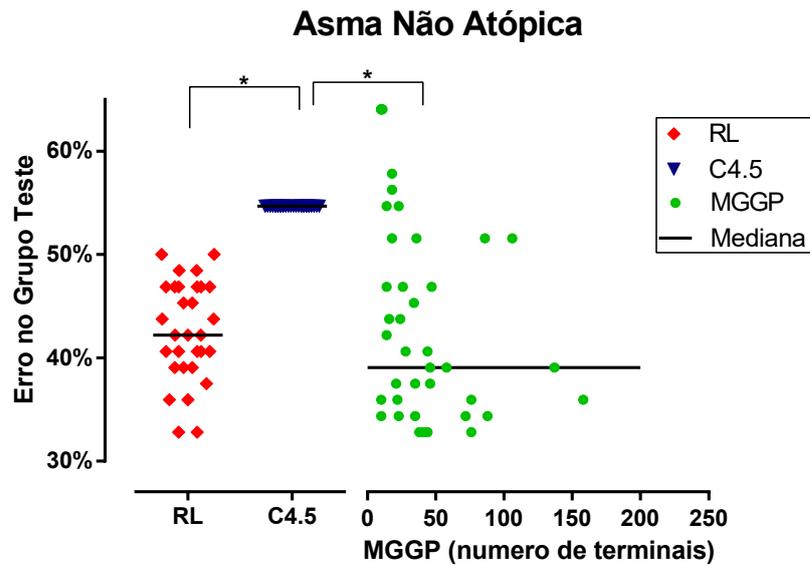


Figura 40 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma não atópica. \* Representa diferença significativa ( $p$ -valor  $< 0,05$ ) entre os grupos usando teste Mann-Whitney.

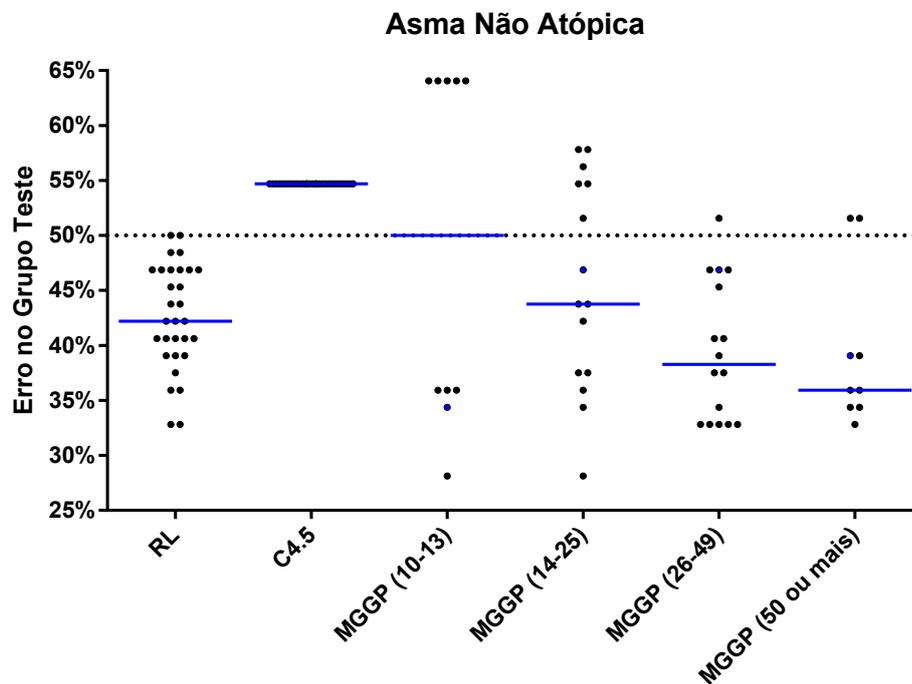


Figura 41 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a asma não atópica.

As principais soluções obtidas por MGGP para asma não atópica são mostradas na Tabela 12. A idade foi importante em muitas relações nas soluções encontradas. Isso

Tabela 11 – Resumo dos resultados referentes a asma não atópica - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	267,1	259	113,8	105,7	71,6	69,5
validação	47,2	60,9	32,3	46	50,6	65,4
teste	6,4	30,4	21,6	5,6	53,1	58,4
	C4.5					
treinamento	206,8	241,6	131,2	166	55,5	64,8
validação	51,7	60,4	32,8	41,5	55,5	64,8
teste	3	26	26	9	25	50
	MGGP					
treinamento	282,4	247,2	125,8	90,6	75,7	66,3
validação	66,8	59,5	33,5	26,3	71,8	63,9
teste	5,8	30,6	21,4	6,2	48,3	58,8

se deve ao importante fenótipo de sibilância transitória que não é comumente associado com atopia. Os sintomas de sibilância transitória tendem à regredir aos 3-5 anos de idade (STEIN; MARTINEZ, 2004), explicando assim sua associação com a idade.

Tabela 12 – Exemplos de relações encontradas pelo MGGP para asma não atópica. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “ {1} ” significa positivo para asma não atópica, “ {0} ” negativo para asma não atópica. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo.

Todos indivíduos		Grupo teste		Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)	Relações Importantes		
2,88(2,19; 3,79)	0,61(0,28; 1,36)	55,7	if((idade > 0) e (desordem psicológica materna != 1)){ 0 }else{ 1 }		
3,47(2,64; 4,57)	2,96(1,25; 6,99)	38,5	if((creche == 1) ou (idade != 0)){ 0 }else{ 1 }		
3,00(2,35; 3,90)	2,78(1,26; 6,15)	37,5	if(gato != 1) e (idade > 0)){ 0 }else{ 1 }		
3,09(2,38; 4,01)	1,08(0,50; 2,33)	49,0	if((idade != 0) e ((desordem psicológica materna != 1) ou (fator nutricional1 <= 0))){ 0 }else{ 1 }		
3,42(2,64; 4,43)	1,26(0,58; 2,74)	47,1	if(creche != 1){ if(idade <= 0){ 1 }else{ 0 } }else{ if(mofo == 0){ 0 }else{ 1 } }		
3,88(3,00; 5,02)	2,08(0,93; 4,64)	41,3	if((idade > 0) ou (fator nutricional2 >= 2)){ if(((desordem psicológica materna != 1) ou (EBV != 1)) ou (fator nutricional1 <= 0)){ 0 }else{ 1 } }		
6,08(4,63; 7,98)	1,93(0,86; 4,30)	42,3	if(((gato != 0) xor (VZV != 1)) e ((HAV != 0) ou (idade == 0))){ 1 }else { if(((fator nutricional2 > 0) xor ((agua tratada != 1) ou (fator nutricional1 == 1))) ou (desordem psicológica materna != 1)){ 0 }else{ 1 } }		
9,16(6,91; 12,16)	3,89(1,72; 8,77)	33,6	if((creche != 0) xor ((fator nutricional3 != 0) e (fator nutricional4 != 0))){ if(fator nutricional2 <= 1) e ((cachorro == 1) xor (pavimentação == 1) e (VZV == 0))) ou ((asma nos pais != 0) e (desordem psicológica materna != 0))){ 1 }else{ 0 }		
10,77(8,06; 14,39)	1,51(0,68; 3,36)	45,1	if(((fator nutricional2 != 2) e (gato == 1)) xor ((idade == 0) ou ((número de irmãos != 3) e (asma nos pais != 0)) xor (fator nutricional2 == 2)) e (((sexo != 0) ou ((número de irmãos > 1) xor (idade == 1) ou (IMC >= 1))) ou ((fator nutricional1 < 2) xor (asma nos pais != 0)))) xor (desordem psicológica materna != 1) e ((idade != 2) ou (número de irmãos >= 2))))){ if(creche != 1) ou (fator nutricional2 < 2)){ 1 }else{ if(número de irmãos != 0) { 0 }else{ 1 } } }else{ 0 }		
15,24(11,21; 20,71)	1,61(0,73; 3,52)	44,2	<b>muito grande para mostrar</b>		

Outro fator que foi muito encontrado nas melhores soluções para asma não atópica foram as características nutricionais, principalmente os fatores nutricionais 1 e 2 que se referem a hábitos nutricionais 1 são abundantes em frutas, hortaliças, leguminosas, cereais e peixes. Este padrão alimentar é rico em antioxidantes e ácidos graxos poli-insaturados, o que alguns trabalhos têm mostrado proteger contra asma (LITONJUA, 2008; DEVEREUX; SEATON, 2005; D'INNOCENZO et al., 2011).

#### 5.4.5 Resultados IgE

A Figura 42 mostra a comparação do erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para IgE nos 30 grupos de treinamentos. O MGGP para IgE obteve significativamente menores erros do que RL e C4.5. Além do baixo erro obtido pelo MGGP em relação ao RL e C4.5, o MGGP foi capaz de achar soluções com baixo erro em diversos níveis de complexidade. A Tabela 13 mostra resultados adicionais.

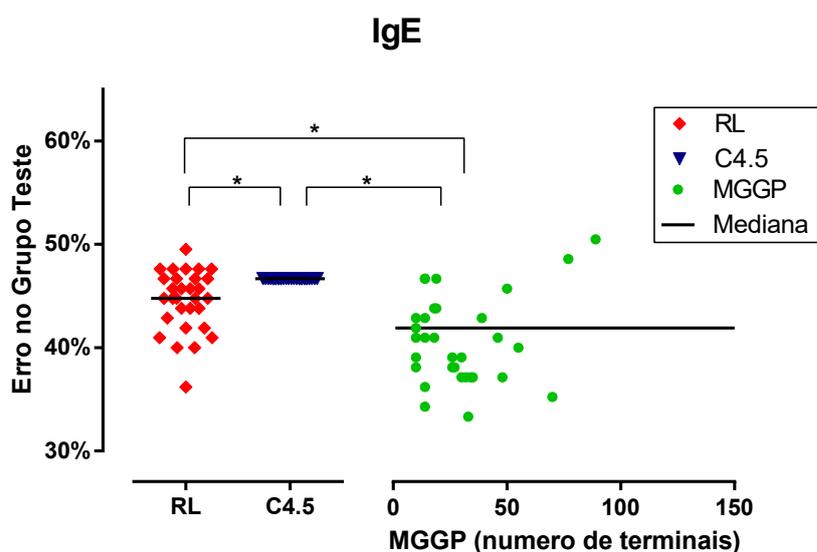


Figura 42 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a IgE. \* Representa diferença significativa ( $p$ -valor  $< 0,05$ ) entre os grupos usando teste Mann-Whitney.

A Figura 43 mostra a comparações da acurácia no grupo de teste das soluções encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho IgE. A acurácia do MGGP foi significativamente maior do que a RL nas complexidades de 10 até 13, de 14 até 25 e de 26 até 49.

As principais soluções obtidas por MGGP para IgE são mostradas na Tabela 14.

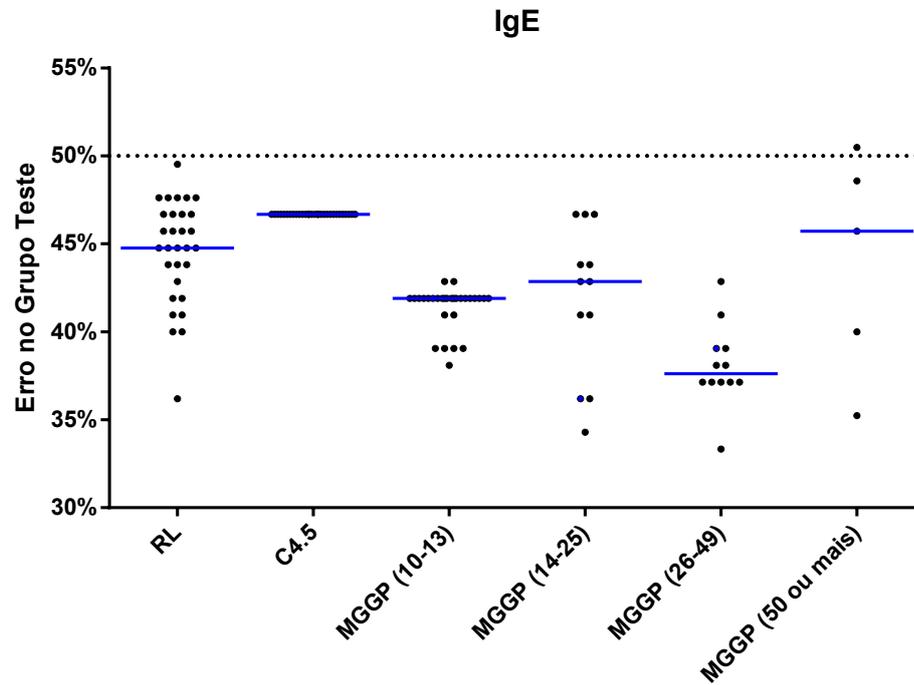


Figura 43 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a IgE.

Tabela 13 – Resumo dos resultados referentes a IgE - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	286,5	274,3	187,3	175,1	62,1	39,8
validação	62,9	64,1	51,3	52,5	54,5	55,6
teste	18,7	39,5	25,5	21,3	46,7	60,8
	C4.5					
treinamento	283,1	244	217,6	178,5	61,3	42,6
validação	70,8	61	54,4	44,6	61,3	52,9
teste	23	33	32	17	57,5	50,8
	MGGP					
treinamento	283,1	279,1	182,4	178,4	61,3	60,5
validação	68,4	68,2	47,2	47	59,3	59
teste	20,9	41	24	19	52,3	63

Tabela 14 – Exemplos de relações encontradas pelo MGGP para IgE. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “{1}” significa positivo para IgE, “{0}” negativo para IgE. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo.

Todos indivíduos		Grupo teste		Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)	Acurácia (%)		
1.81(1.45; 2.26)	2.28(1.12; 4.63)	60.0	60.0	10	if((sexo != 1) ou (gato != 0)){ 0 }else{ 1 }
2.01(1.61; 2.51)	1.87(0.93; 3.75)	57.7	57.7	10	if((sexo != 0) e (rede de esgoto != 1)){ 1 }else{ 0 }
1.65(1.31; 2.08)	1.81(0.88; 3.72)	56.9	56.9	10	if((fator nutricional1 != 1) e (sexo != 0)){ 1 }else{ 0 }
2.18(1.74; 2.74)	2.37(1.11; 5.06)	59.2	59.2	14	if(((Tgondi = 0) e (sexo != 0)) e (rede de esgoto != 1)){ 1 }else{ 0 }
2.02(1.62; 2.53)	1.98(0.99; 3.98)	58.5	58.5	14	if(((sexo != 1) e (gato = 0)) ou (rede de esgoto = 1)){ 0 }else{ 1 }
2.31(1.84; 2.89)	1.75(0.87; 3.52)	56.9	56.9	18	if(((fator nutricional1 = 1) ou (sexo = 1)) e ((rede de esgoto = 1) xor (Tgondi = 0))){ 1 }else{ 0 }
2.41(1.93; 3.02)	2.13(1.05; 4.31)	59.2	59.2	22	if(((Tgondi = 1) ou (rede de esgoto = 1)) ou ((sexo != 1) e (fator nutricional2 <= 0) xor (rede de esgoto = 0))){ 0 }else{ 1 }
3.55(2.80; 4.49)	2.92(1.43; 5.96)	63.1	63.1	46	if(((Tgondi != 1) ou ((fator nutricional4 = 1) xor (pavimentação != 1)) xor (rede de esgoto != 1)) ou ((sexo != 1) e (((idade <= 1) e (fator nutricional2 < 1)) xor (rede de esgoto = 0)) xor ((fator nutricional1 <= 0) xor (Trichiura = 1)) e (número de irmãos > 2))){ 0 }else{ 1 }
3.63(2.88; 4.57)	2.14(1.06; 4.35)	59.2	59.2	58	if((Tgondi = 0) xor ((fator nutricional4 <= 0) e (idade < 2))) { if((sexo = 1) xor (Trichiura = 1)) ou ((rede de esgoto != 0) xor (fator nutricional2 <= 0))}{ if(((IMC = 1) xor (HAV = 0)) ou (fator nutricional1 >= 1)) xor (rede de esgoto != 1)}{ 0 }else{ 1 } }else { 0 } }else{ if(sexo != 0){ if(VZV = 0){ if(VZV = 0){ 1 }else{ 0 } } }else{ 0 } }

Os modelos gerados pelo MGGP para explicar IgE mostraram que muitas soluções apresentaram o gênero masculino um importante fator associado ao aumento da IgE. A ausência das infecções *T.trichiura* e *T. gondii*, bem como a presença de rede de esgoto são características que aparecem em muitas soluções como fatores que aumentam a IgE. Estando de acordo com o hipótese da higiene.

Um modelo que pode ajudar no entendimento desta patologia é: “if(((fator nutricional = 1) ou (sexo = 1)) e ((rede de esgoto = 1) xor (Tgondi = 0))) { 1 } else { 0 }”. Este modelo indica que quando uma pessoa tem níveis moderados de consumo de peixe, fruta, cereais ou é do sexo masculino, e também mostra a ausência de infecções por *T. gondii* ou exclusivo tem de sistema de esgoto, representa o aumento da chance de ser positivo para IgE.

Muitos fenômenos biológicos não têm um comportamento linear. As células imunes como linfócitos, quando estimuladas, têm sua resposta aumentada. Contudo, o excesso de estimulação leva a anergia ou apoptose destas células, reduzindo assim a resposta. Esse tipo de comportamento não é capturado corretamente pelas técnica RL.

O modelo descrito mostra que ser do gênero masculino, valores intermediários do padrão nutricional 1, ausência de *T. gondii*, e presença de rede de esgoto são características que aumentam IgE. Contudo a presença de todos estes fatores de risco juntos reduziria as chances de ter IgE. Este comportamento fica mais evidente quando se observa modelos mais complexos na Tabela 14.

#### 5.4.6 Resultados SPT

A pessoa atópica tem sido definida como aquela que é positiva para SPT ou apresenta anticorpos IgE específicos contra alérgenos comuns. Em geral, ambas as medidas são consistentes quando avalia o mesmo alérgeno na mesma pessoa, particularmente na Europa e nos EUA (WEINMAYR et al., 2010). No entanto, há muitas evidências de dissociação entre a positividade para SPT e IgE mesmo quando o mesmo alérgeno é usado em ambos os testes, particularmente em países de baixa e média renda, principalmente apresentando pessoas positivas para IgE e negativas para SPT (WEINMAYR et al., 2010; JULGE; VASAR; BJÖRKSTÉN, 2001; PERZANOWSKI et al., 2002a).

A Figura 44 mostra a comparação via erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para SPT nos 30 grupos de treinamentos. O MGGP para SPT não encontrou diferenças significativas no erro do grupo de teste em relação à RL. Contudo, o C4.5 apresentou fraco desempenho, obtendo significativamente maiores erros do que RL e MGGP. A Tabela 15 mostra resultados adicionais indicando que o baixo desempenho do C4.5 pode ser devido à baixa sensibilidade obtida.

A Figura 45 mostra a comparações da acurácia no grupo de teste das soluções

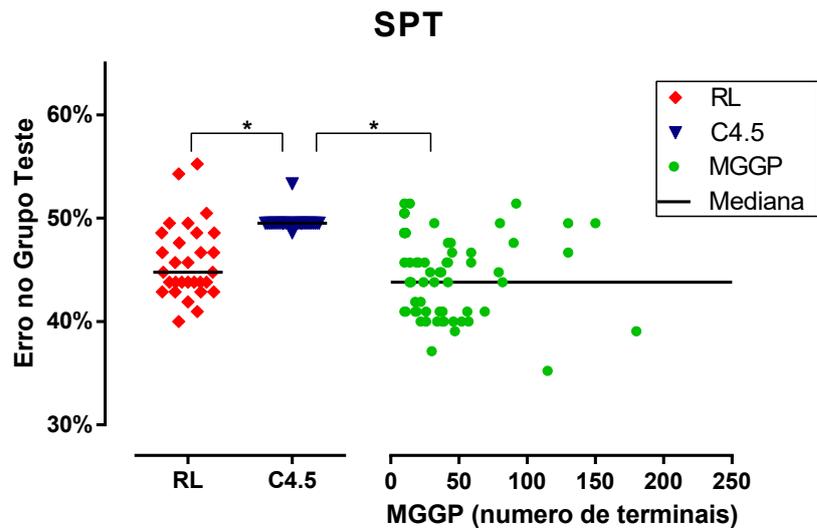


Figura 44 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a SPT. \* Representa diferença significativa ( $p$ -valor  $< 0,05$ ) entre os grupos usando teste Mann-Whitney.

Tabela 15 – Resumo dos resultados referentes a SPT - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	333,3	310,6	214,2	191,5	63,5	59,2
validação	69,8	72,8	58,4	61,4	53,2	55,5
teste	17,7	39,2	33,8	14,3	55,4	53,7
	C4.5					
treinamento	278,5	305,6	219,2	246,3	53,1	58,2
validação	68,7	76,6	54,6	62,5	52,4	58,4
teste	15,3	37,6	35,4	16,7	47,7	51,6
	MGGP					
treinamento	365,8	348,9	175,6	158,8	69,7	66,5
validação	78,5	81	50,4	52,9	59,7	61,6
teste	17,3	42,3	30,7	14,7	54,1	57,9

encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho SPT. Apesar dos dados de IgE apresentarem melhores desempenhos nas soluções de baixa complexidade, o SPT obteve um melhor desempenho em soluções com mais alta complexidade do que o IgE.

As principais soluções obtidas por MGGP para SPT são mostradas na Tabela 16.

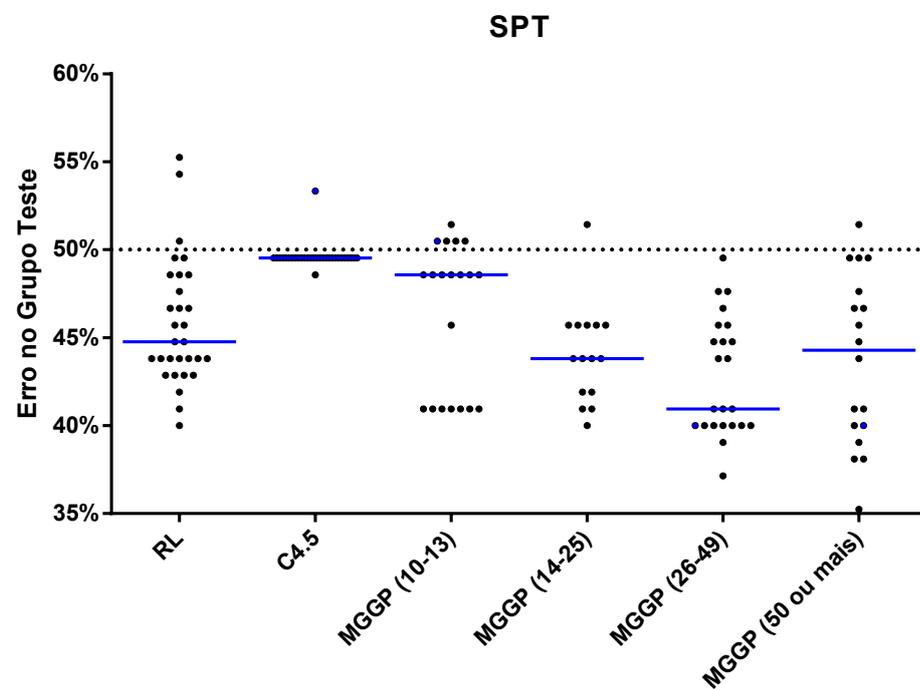


Figura 45 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a SPT.

Tabela 16 – Exemplos de relações encontradas pelo MGGP para SPT. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “{1}” significa positivo para SPT, “{0}” negativo para SPT. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo.

Todos indivíduos		Grupo teste		Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)	Acurácia (%)		
2.01(1.63; 2.49)	2.92(1.46; 5.86)	62.3	62.3	10	if((Troca roupa de cama = 0) ou (IMC != 0)){ 0 }else{ 1 }
2.07(1.67; 2.57)	2.44(1.22; 4.86)	60.3	60.3	10	if((Troca roupa de cama != 0) e (Ttrichiura != 1)){ 1 }else{ 0 }
1.96(1.59; 2.42)	1.58(0.81 ;3.06)	55.5	55.5	10	if((fator nutricional4 > 0) e (Troca roupa de cama != 0)){ 1 }else{ 0 }
2.03(1.63; 2.54)	1.52(0.77; 2.99)	54.8	54.8	10	if((HSV = 0) e (Troca roupa de cama != 0)){ 1 }else{ 0 }
2.41(1.95; 2.98)	2.06(1.06; 3.98)	58.9	58.9	14	if(((HSV = 0) ou (fator nutricional3 >= 1)) e (Troca roupa de cama != 0)){ 1 }else{ 0 }
2.66(2.15; 3.28)	2.45(1.26; 4.77)	60.9	60.9	18	if(((HSV = 0) ou (creche = 1)) e ((fator nutricional4 != 1) ou (Troca roupa de cama != 0))){ 1 }else{ 0 }
2.43(1.97; 3.01)	2.23(1.14; 4.38)	60.0	60.0	20	if((HSV != 0) e (fator nutricional3 = 0)){ (if(fator nutricional1 != 1) { 0 }else{ 1 }) }else{ (if(Troca roupa de cama != 0){ 1 }else{ 0 }) }
3.41(2.75; 4.23)	2.31(1.19; 4.48)	60.3	60.3	39	if((fator nutricional4 < 1) xor ((fator nutricional2 >= 1) ou (número de irmãos = 1) e (Fly at Home = 0))) e ((Mother Psychological disorder != 1) ou (número de irmãos >= 1)) xor (Troca roupa de cama = 0))){ (if((HSV != 1) ou (Tgondi != 1)) { 1 }else{ 0 }) }else{ 0 }
5.92(4.72; 7.42)	2.35(1.14; 4.85)	58.9	58.9	124	<b>Muito grande para mostrar</b>

SPT

Um tipo de relação muito encontrada nas soluções para SPT é a presença de uma infecção junto com uma característica ambiental que usualmente é descrita como protetora contra atopia. Trabalhos anteriores mostraram que infecções virais crônicas protegem contra SPT sem afetar a presença de IgE específica (VEIGA et al., 2011; ALCANTARA-NEVES et al., 2012). Assim essas soluções encontradas para SPT indicam que a proteção dos vírus contra SPT pode ser dependente do ambiente. Contudo, mais estudos são necessários para testar essa hipótese.

#### 5.4.7 Resultados SPT com IgE como possível exposição

O estudo do SPT na presença de IgE como possível variável de exposição modifica completamente a interpretação dos modelos gerados. A IgE e o SPT estão na mesma via causal da alergia, particularmente fatores de risco causam IgE que causa SPT. Assim, o IgE é uma variável de efeito intermediário entre muitos fatores de risco e o SPT, desta forma não poderia aparecer nos modelos SPT quando se quer estudar os fatores de risco para alergia. Contudo, quando se quer entender como e quais fatores contribuem para dissociação da IgE com o SPT, a presença da IgE nos modelos pode evidenciar esses fatores.

A Figura 46 mostra a comparação através do erro no grupo teste das soluções obtidas por RL, C4.5 e MGGP para SPT com IgE como possível exposição nos 30 grupos de treinamentos. O MGGP para SPT com IgE como possível exposição obteve significativamente maiores erros do que RL e C4.5. A Tabela 17 mostra resultados adicionais indicando que todas as técnicas obtiveram bons desempenho com baixo número de falsos positivos e negativos.

A Figura 47 mostra a comparações da acurácia no grupo de teste das soluções encontradas por RL, C4.5 e MGGP separados por faixa de complexidade do desfecho SPT com IgE como possível exposição.

Apesar do inferior desempenho do MGGP em relação ao C4.5 e RL, o MGGP foi capaz de encontrar boas soluções nos mais diversos níveis de complexidade. As principais soluções obtidas por MGGP para IgE são mostradas na Tabela 18.

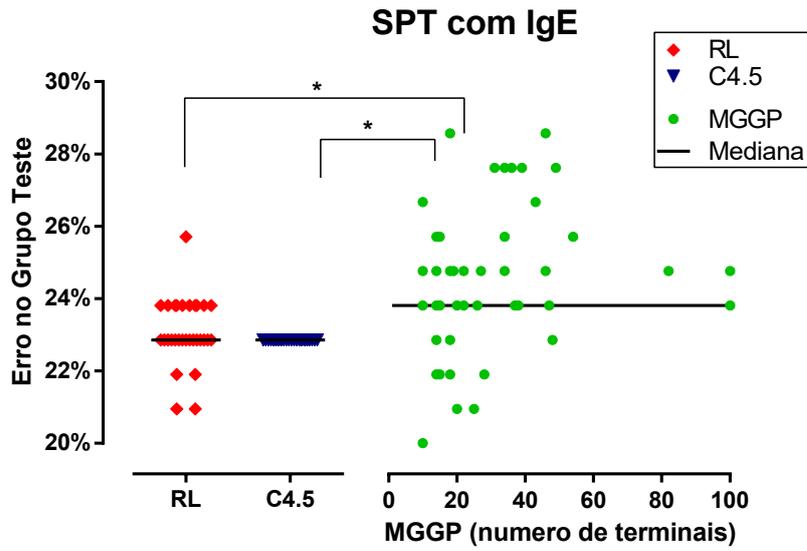


Figura 46 – Erro no grupo teste das soluções obtida de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a SPT com IgE como possível exposição. \* Representa diferença significativa (p-valor < 0,05) entre os grupos usando teste Mann-Whitney.

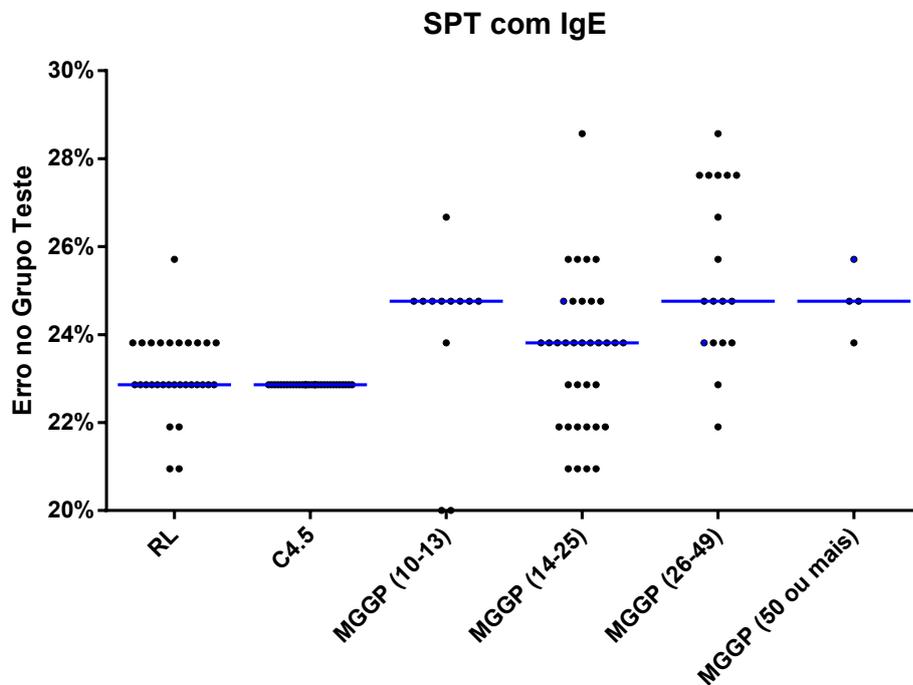


Figura 47 – Comparação das acurácias no grupo teste das soluções separadas por faixas de complexidades obtidas de execuções em 30 grupos de treinamento utilizando as técnicas de RL, C4.5 e MGGP tendo como desfecho a SPT com IgE como possível exposição.

Tabela 17 – Resumo dos resultados referentes a SPT com IgE - Onde vp é média do número de verdadeiros positivos, vn é a média do número de verdadeiros negativos, fp é a média do número de falsos positivos e fn é a média do número de falsos negativos.

	RL					
	vp	vn	fp	fn	sensibilidade	especificidade
treinamento	437,1	431,4	93,4	87,7	83,3	82,2
validação	106,4	106,8	24,4	24,8	81,1	81,4
teste	26,4	54,3	18,7	5,6	82,6	74,4
	C4.5					
treinamento	435,7	427,2	97,6	89,1	83	81,4
validação	108,9	106,8	24,4	22,3	83	81,4
teste	27	54	19	5	84,4	74
	MGGP					
treinamento	442,2	436,7	88	82,5	84,3	83,2
validação	110,4	107,5	23,8	20,9	84,1	81,8
teste	25,5	54	19	6,5	79,6	73,9

Tabela 18 – Exemplos de relações encontradas pelo MGGP para SPT com IgE. A primeira coluna é a razão de chance de toda a base de dados seguido do intervalo de confiança, a segunda coluna é a razão de chance seguido do intervalo de confiança no grupo de teste e a terceira coluna é a acurácia no grupo teste. Onde “{1}” significa positivo para SPT, “{0}” negativo para SPT. Uma variável seguida por “0” significa negativo para esta variável e “1” é positivo.

Todos indivíduos		Grupo teste		Acurácia (%)	Complexidade	Relações Importantes
OR(C.I.95%)	OR(C.I.95%)	OR(C.I.95%)	Acurácia (%)			
16,98(13,08; 22,05)	8,73(4,13; 18,44)		25,3	10	if((IgE != 1) ou (T. trichiura == 1)){ 0 }else{ 1 }	
18,14(13,96; 23,58)	10,85(5,04; 23,38)		23,2	10	if((IgE == 1) e (agua tratada != 0)){ 1 }else{ 0 }	
16,22(12,46; 21,14)	13,88(6,26; 30,78)		21,2	10	if((A.lumbricoides == 0) e (IgE != 0)){ 1 }else{ 0 }	
17,87(13,76; 23,20)	10,12(4,72; 21,70)		23,9	14	if((IMC == 0) ou (T. trichiura != 1) e (IgE == 1)){ 1 }else{ 0 }	
19,29(14,81; 25,14)	11,73(5,40; 25,50)		22,6	18	if((T. trichiura == 1) e ((Índice econômico < 1) e (número de irmãos <= 3))) ou (IgE == 0)){ 0 }else{ 1 }	
25,41(19,27; 33,52)	13,88(6,26; 30,78)		21,2	20	if((T. trichiura != 0) e (Índice econômico != 1)){ if(fator nutricional2 != 2){ 0 }else{ 1 }	
22,88(17,39; 30,10)	8,26(3,91; 17,42)		26,0	34	}else{ if(IgE == 0){ 0 }else{ 1 } }	
					if(((cachorro == 0) ou ((número de irmãos >= 2) xor ((mosca != 0) ou (VZV == 0)))) e	
					((T. trichiura == 0) ou ((Índice econômico > 0) xor (fator nutricional1 == 0))) e	
					(IgE != 0)){ 1 }else{ 0 }	
32,46(24,21; 43,52)	11,05(5,01; 24,39)		23,9	49	if(IgE == 1){ if((fator nutricional2 < 1) e ((número de irmãos != 1) e (cachorro != 0)))	
					{ 0 }else{ 1 } }else{ if(((número de irmãos > 1) xor (fator nutricional3 == 0)) e	
					((número de irmãos > 1) xor (T. trichiura == 1) e (Troca roupa de cama != 0))){	
					if((fator nutricional4 >= 1) xor (fator nutricional2 >= 1)){ 0 }else{ 1 } }else{ 0 } }	
26,63(20,12; 35,24)	8,73(4,13; 18,44)		25,3	82	if((T. trichiura != 0) e (fator nutricional1 <= 0) ou (fator nutricional2 != 2) e	
					((Tgondi != 1) e (Índice econômico == 0))) ou (IgE != 1){ if((((rede de esgoto != 1) ou	
					(fator nutricional2 != 1) ou (sexo != 0) ou (fator nutricional1 >= 2) ou (idade > 0) e	
					(fator nutricional1 == 1)){ 0 }else{ 1 } }else{ if((número de irmãos < 3) ou	
					(fator nutricional3 <= 0)){ if((cachorro == 0) ou (número de irmãos < 2)){ 1 }else{ 0 }	
					}else{ if(((fator nutricional1 > 1) ou (fator nutricional2 >= 2) ou (pavimentação == 1)	
					{ 1 }else{ 0 } }	
39,32(29,17; 53,00)	12,23(5,55; 26,95)		22,6	100	if((IgE == 1) e (((Tgondi == 0) ou ((Troca roupa de cama == 1) ou ((fator nutricional4 != 0) e	
					(HAV == 0))) e ((IMC != 1) ou (desordem psicológica materna == 1) ou (número de irmãos <= 1))) e	
					((fator nutricional1 != 1) xor (número de irmãos <= 1) ou (T. trichiura != 1) ou	
					(Índice econômico > 0))) e (((número de irmãos == 1) ou (mofa != 0) e (Tgondi != 1)) ou	
					(cachorro == 0) ou (fator nutricional2 != 0)))}{ 1 }else{ if((((fator nutricional2 == 2) xor	
					(fator nutricional1 != 0) ou ((fator nutricional4 == 0) xor (fator nutricional2 <= 0)) ou	
					(número de irmãos == 1) ou (fator nutricional3 != 1)){ 0 }else{	
					if((Troca roupa de cama != 0){ 1 }else{ 0 } }	

As infecções helmínticas *T.trichiura* e *A.lumbricoides* protegeram contra SPT na presença da IgE. Outros trabalhos na literatura já tinham mostrado que infecções helmínticas aumentavam a dissociação da IgE com o SPT (PERZANOWSKI et al., 2002a; BIGGELAAR et al., 2001; COOPER et al., 2008).

A maioria dos estudos epidemiológicos utiliza técnicas que capturam apenas relações lineares entre variáveis preditoras, como por exemplo RL. O algoritmo MGGP encontrou melhores soluções para a asma e IgE do que os algoritmos C4.5 e RL. Isso indica que RL não foi capaz de encontrar as relações complexas entre as variáveis.

O MGGP apresentou muitas qualidades desejáveis para o estudo de problemas epidemiológicos. O MGGP apresenta uma gramática que possibilita o desenvolvimento de soluções mais fáceis de serem entendidas, além de apresentar a capacidade de criar soluções direcionadas as perguntas com possibilidade de inserção do conhecimento do especialista.

A abordagem multiobjetivo aos problemas epidemiológicos possibilitou a obtenção de múltiplas soluções com características diversificadas em complexidade e acurácia, o que ajudou muito em entender os problema em um nível de profundidade que dificilmente seria possível em uma abordagem mono-objetiva.

## 6 CONCLUSÕES

A técnica MGGP apresentou características úteis para entender relações epidemiológicas como uma gramática que possibilita o desenvolvimento de soluções mais fáceis de serem entendidas, além de apresentar a capacidade de criar soluções direcionadas as perguntas com possibilidade de inserção do conhecimento do especialista

A abordagem multiobjetiva aos problemas epidemiológicos possibilitou a obtenção de múltiplas soluções com características diversificadas em complexidade e acurácia, o que ajudou muito em entender os problema em um nível de profundidade que dificilmente seria possível em uma abordagem mono-objetiva. Com relação ao estudo da asma e atopia, o MGGP apresentou as seguintes características:

- O MGGP obteve um desempenho melhor em relação à RL na asma e IgE, contudo foi inferior na asma atópica e no SPT contendo IgE.
- O MGGP foi capaz de encontrar relações que não seriam possíveis através de RL.
- Os resultados obtidos pelo MGGP foram melhores quando o critério de complexidade adotado foi numero de terminais em relação ao critério tamanho máximo da árvore.
- O MGGP foi capaz de gerar soluções diversificadas com diversas relações entre complexidade e acurácia para todos os desfechos estudados.
- A idade foi uma característica importante em muitas relações para asma, principalmente no fenótipo asma não atópica.
- Foram encontrados indícios de que o excesso de estímulos que contribuem para produção de IgE pode levar à supressão de sua produção.
- Fatores nutricionais e desordem psicológica na mãe foram características que estiveram associados com o desenvolvimento da asma e da atopia.
- Foram encontrados indícios corroborados pela literatura de que os helmintos podem aumentar a dissociação da IgE com o SPT, fazendo com que pessoas positivas para IgE não apresentem reatividade cutânea.

## REFERÊNCIAS

- ABBAS, A. *Imunologia celular e molecular 7a edição*. [S.l.]: Elsevier Brasil, 2012.
- ALCÂNTARA-NEVES, N. M. et al. The presence of serum anti-ascaris lumbricoides ige antibodies and of trichuris trichiura infection are risk factors for wheezing and/or atopy in preschool-aged brazilian children. *Respiratory research*, BioMed Central, v. 11, n. 1, p. 114, 2010.
- ALCANTARA-NEVES, N. M. et al. The effect of single and multiple infections on atopy and wheezing in children. *Journal of Allergy and Clinical Immunology*, Elsevier, v. 129, n. 2, p. 359–367, 2012.
- ASHER, M. I. et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: Isaac phases one and three repeat multicountry cross-sectional surveys. *The Lancet*, Elsevier, v. 368, n. 9537, p. 733–743, 2006.
- BAINES, K. J. et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *Journal of Allergy and Clinical Immunology*, Elsevier, v. 133, n. 4, p. 997–1007, 2014.
- BARBOSA, H.; BERNARDINO, H. Genetic Programming in Civil, Structural and Environmental Engineering. *Computational Technology Reviews*, v. 4, p. 115–145, set. 2010.
- BARRETO, M. L. et al. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in northeastern brazil (salvador-scaala study). *BMC pulmonary medicine*, BioMed Central, v. 6, n. 1, p. 15, 2006.
- BARRETO, M. L. et al. Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast brazil: assessment by two cohort studies. *The Lancet*, Elsevier, v. 370, n. 9599, p. 1622–1628, 2007.
- BEASLEY, R.; ASTHMA, T. I. S. of et al. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: Isaac. *The Lancet*, Elsevier, v. 351, n. 9111, p. 1225–1232, 1998.
- BERNARDINO, H. S. *Programação Imunológica Gramatical para Inferência Automática de Modelos e Projeto Ótimo de Estruturas*. Tese (Doutorado) — LNCC, 2012.
- BIGGELAAR, A. H. van den et al. The prevalence of parasite infestation and house dust mite sensitization in gabonese schoolchildren. *International archives of allergy and immunology*, Karger Publishers, v. 126, n. 3, p. 231–238, 2001.
- BILLINGS, S.; VOON, W. A prediction-error and stepwise-regression estimation algorithm for non-linear systems. *International Journal of Control*, Taylor & Francis, v. 44, n. 3, p. 803–822, 1986.
- BOJARCZUKA, C. C.; LOPESB, H. S.; FREITASC, A. A. Data mining with constrained-syntax genetic programming: Applications in medical data set. *algorithms*, Citeseer, v. 6, p. 7, 2001.

BONGARD, J.; LIPSON, H. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 104, n. 24, p. 9943–9948, 2007.

BONITA, R.; BEAGLEHOLE, R.; KJELLSTRÖM, T. *Epidemiologia básica*. [S.l.]: SciELO Brasil, 2008.

BORNELÖV, S. et al. Rule-based models of the interplay between genetic and environmental factors in childhood allergy. *PloS one*, Public Library of Science, v. 8, n. 11, p. e80080, 2013.

BUTLAND, B. K. et al. Childhood asthma in south london: trends in prevalence and use of medical services 1991–2002. *Thorax*, BMJ Publishing Group Ltd and British Thoracic Society, v. 61, n. 5, p. 383–387, 2006.

CALAM, R. et al. Behavior problems antecede the development of wheeze in childhood: a birth cohort study. *American journal of respiratory and critical care medicine*, Am Thoracic Soc, v. 171, n. 4, p. 323–327, 2005.

CAMPBELL, D. et al. Mechanisms of allergic disease—environmental and genetic determinants for the development of allergy. *Clinical & Experimental Allergy*, Wiley Online Library, v. 45, n. 5, p. 844–858, 2015.

CANN, P. L. et al. Indoor environment and children’s health: recent developments in chemical, biological, physical and social aspects. *International journal of hygiene and environmental health*, Elsevier, v. 215, n. 1, p. 1–18, 2011.

CELEDÓN, J. C. et al. Exposure to cat allergen, maternal history of asthma, and wheezing in first 5 years of life. *The Lancet*, Elsevier, v. 360, n. 9335, p. 781–782, 2002.

CHANG, S.-S.; HU, H.-Y. No inverse relationship between helicobacter pylori infection and adult asthma with peptic ulcer disease. *Hepato-gastroenterology*, v. 61, n. 130, p. 529–534, 2013.

CHAWLA, N. V. Data mining for imbalanced datasets: An overview. In: *Data mining and knowledge discovery handbook*. [S.l.]: Springer, 2005. p. 853–867.

CHEN, J.-H. Asthma and child behavioral skills: Does family socioeconomic status matter? *Social Science & Medicine*, Elsevier, v. 115, p. 38–48, 2014.

COOPER, P. J. et al. Reduced risk of atopy among school-age children infected with geohelminth parasites in a rural area of the tropics. *Journal of Allergy and Clinical Immunology*, Elsevier, v. 111, n. 5, p. 995–1000, 2003.

COOPER, P. J. et al. *Ascaris lumbricoides*–induced interleukin-10 is not associated with atopy in schoolchildren in a rural area of the tropics. *Journal of Infectious Diseases*, Oxford University Press, v. 197, n. 9, p. 1333–1340, 2008.

COSTA, G. N. et al. A genome-wide association study of asthma symptoms in latin american children. *BMC genetics*, BioMed Central, v. 16, n. 1, p. 141, 2015.

CRAMER, N. L. A representation for the adaptive generation of simple sequential programs. In: *Proc. of the 1st International Conference on Genetic Algorithms*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1985. p. 183–187.

- CUNHA, S. S. d. et al. Asthma cases in childhood attributed to atopy in tropical area in brazil. *Revista panamericana de salud publica*, SciELO Public Health, v. 28, n. 6, p. 405–411, 2010.
- DEB, K. et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, IEEE, v. 6, n. 2, p. 182–197, 2002.
- DEVEREUX, G. The increase in the prevalence of asthma and allergy: food for thought. *Nature Reviews Immunology*, Nature Publishing Group, v. 6, n. 11, p. 869–874, 2006.
- DEVEREUX, G.; SEATON, A. Diet as a risk factor for atopy and asthma. *Journal of allergy and clinical immunology*, Elsevier, v. 115, n. 6, p. 1109–1117, 2005.
- DICK, S. et al. A systematic review of associations between environmental exposures and development of asthma in children aged up to 9 years. *BMJ open*, British Medical Journal Publishing Group, v. 4, n. 11, p. e006554, 2014.
- D'INNOCENZO, S. et al. Condições socioeconômicas e padrões alimentares de crianças de 4 a 11 anos: estudo scaala-salvador/bahia. *Revista Brasileira de Saúde Materno Infantil*, Instituto de Medicina Integral Prof. Fernando Figueira, v. 11, n. 1, p. 41–49, 2011.
- D'INNOCENZO, S. et al. Dietary pattern, asthma, and atopic and non-atopic wheezing in children and adolescents: Scaala study, salvador, bahia state, brazil. *Cadernos de saude publica*, SciELO Public Health, v. 30, n. 9, p. 1849–1860, 2014.
- DOMÍNGUEZ-ALMENDROS, S.; BENÍTEZ-PAREJO, N.; GONZALEZ-RAMIREZ, A. Logistic regression models. *Allergologia et immunopathologia*, Elsevier, v. 39, n. 5, p. 295–305, 2011.
- EIBEN, A. E.; SMITH, J. E. *Introduction to evolutionary computing*. [S.l.]: Springer Science & Business Media, 2003.
- ENDARA, P. et al. Effect of urban vs. rural residence on the association between atopy and wheeze in latin america: findings from a case-control analysis. *Clinical & Experimental Allergy*, Wiley Online Library, v. 45, n. 2, p. 438–447, 2015.
- ESTRADA-GIL, J. K. et al. Gpdti: A genetic programming decision tree induction method to find epistatic effects in common complex diseases. *Bioinformatics*, Oxford Univ Press, v. 23, n. 13, p. i167–i174, 2007.
- FEITOSA, C. A. et al. Behavior problems and prevalence of asthma symptoms among brazilian children. *Journal of psychosomatic research*, Elsevier, v. 71, n. 3, p. 160–165, 2011.
- FIGUEIREDO, C. A. et al. Evidence for a modulatory effect of il-10 on both th1 and th2 cytokine production: the role of the environment. *Clinical Immunology*, Elsevier, v. 139, n. 1, p. 57–64, 2011.
- FIGUEIREDO, C. A. et al. Does ifn- $\gamma$  play a role on the pathogenesis of non-atopic asthma in latin america children? *Allergy, Asthma & Clinical Immunology*, BioMed Central, v. 8, n. 1, p. 18, 2012.

- FJELBYE, J. et al. Cd1d knockout mice exhibit aggravated contact hypersensitivity responses due to reduced interleukin-10 production predominantly by regulatory b cells. *Experimental dermatology*, Wiley Online Library, v. 24, n. 11, p. 853–856, 2015.
- FLOHR, C.; QUINNELL, R.; BRITTON, J. Do helminth parasites protect against atopy and allergic disease? *Clinical & Experimental Allergy*, Wiley Online Library, v. 39, n. 1, p. 20–32, 2009.
- FRANCO, R. et al. Cost-effectiveness analysis of a state funded programme for control of severe asthma. *BMC Public Health*, BioMed Central, v. 7, n. 1, p. 82, 2007.
- GALOBARDES, B. et al. Childhood wheezing, asthma, allergy, atopy, and lung function: different socioeconomic patterns for different phenotypes. *American journal of epidemiology*, Oxford Univ Press, v. 182, n. 9, p. 763–774, 2015.
- GERN, J. E.; BUSSE, W. W. Relationship of viral infections to wheezing illnesses and asthma. *Nature Reviews Immunology*, Nature Publishing Group, v. 2, n. 2, p. 132–138, 2002.
- GINA. *Global Initiative for Asthma (GINA) Global strategy for asthma management and prevention*. 2015. <<http://www.ginasthma.org/>>.
- GOTTESMAN, I. I.; SHIELDS, J. Genetic theorizing and schizophrenia. *The British Journal of Psychiatry*, RCP, v. 122, n. 566, p. 15–30, 1973.
- GOWERS, A. M. et al. Does outdoor air pollution induce new cases of asthma? biological plausibility and evidence; a review. *Respirology*, Wiley Online Library, v. 17, n. 6, p. 887–898, 2012.
- GUO, H.; NANDI, A. K. Breast cancer diagnosis using genetic programming generated feature. *Pattern Recognition*, Elsevier, v. 39, n. 5, p. 980–987, 2006.
- HEISSENHUBER, A. et al. Health impacts of second-hand exposure to cat allergen fel d 1 in infants. *Allergy*, Wiley Online Library, v. 58, n. 2, p. 154–157, 2003.
- HOFFMAN, W. A. et al. The sedimentation-concentration method in schistosomiasis mansoni. *Puerto Rico Journal of Public Health and Tropical Medicine*, v. 9, n. 3, 1934.
- HONG, J.-H.; CHO, S.-B. The classification of cancer based on dna microarray data that uses diverse ensemble genetic programming. *Artificial intelligence in Medicine*, Elsevier, v. 36, n. 1, p. 43–58, 2006.
- HOPCROFT, J. E.; ULLMAN, J. D. Formal languages and their relation to automata. Addison-Wesley Longman Publishing Co., Inc., 1969.
- HUMBERT, M. et al. The immunopathology of extrinsic (atopic) and intrinsic (non-atopic) asthma: more similarities than differences. *Immunology today*, Elsevier, v. 20, n. 11, p. 528–533, 1999.
- IDRIS, I. B. et al. Environmental air pollutants as risk factors for asthma among children seen in pediatric clinics in ukmmc, kuala lumpur. *Annals of Global Health*, Elsevier, v. 82, n. 1, p. 202–208, 2016.

- JANSON, C. et al. The effect of infectious burden on the prevalence of atopy and respiratory allergies in iceland, estonia, and sweden. *Journal of allergy and clinical immunology*, Elsevier, v. 120, n. 3, p. 673–679, 2007.
- JOHANSSON, S. et al. Revised nomenclature for allergy for global use: Report of the nomenclature review committee of the world allergy organization, october 2003. *Journal of allergy and clinical immunology*, Elsevier, v. 113, n. 5, p. 832–836, 2004.
- JULGE, K.; VASAR, M.; BJÖRKSTÉN, B. Development of allergy and ige antibodies during the first five years of life in estonian children. *Clinical & Experimental Allergy*, Wiley Online Library, v. 31, n. 12, p. 1854–1861, 2001.
- KABIR, Z. et al. Prevalence of symptoms of severe asthma and allergies in irish school children: An isaac protocol study, 1995–2007. *International journal of environmental research and public health*, Molecular Diversity Preservation International, v. 8, n. 8, p. 3192–3201, 2011.
- KAMRADT, T.; GÖGGEL, R.; ERB, K. J. Induction, exacerbation and inhibition of allergic and autoimmune diseases by infection. *Trends in immunology*, Elsevier, v. 26, n. 5, p. 260–267, 2005.
- KATZ, N. et al. A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. *Revista do Instituto de Medicina Tropical de São Paulo*, v. 14, n. 6, p. 397–400, 1972.
- KOZA, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. [S.l.]: The MIT Press, 1992.
- KOZA, J. R. Human-competitive results produced by genetic programming. *Genetic Programming and Evolvable Machines*, Springer, v. 11, n. 3-4, p. 251–284, 2010.
- KOZA, J. R. et al. Reuse, parameterized reuse, and hierarchical reuse of substructures in evolving electrical circuits using genetic programming. In: SPRINGER. *International Conference on Evolvable Systems*. [S.l.], 1996. p. 312–326.
- LAST, J. M.; ABRAMSON, J. H.; FREIDMAN, G. D. *A dictionary of epidemiology*. [S.l.]: Oxford Univ Press, 2001. v. 141.
- LEE, P. H. Resampling methods improve the predictive power of modeling in class-imbalanced datasets. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 11, n. 9, p. 9776–9789, 2014.
- LEONARDI-BEE, J.; PRITCHARD, D.; BRITTON, J. Asthma and current intestinal parasite infection: systematic review and meta-analysis. *American journal of respiratory and critical care medicine*, Am Thoracic Soc, v. 174, n. 5, p. 514–523, 2006.
- LITONJUA, A. A. Dietary factors and the development of asthma. *Immunology and allergy clinics of North America*, Elsevier, v. 28, n. 3, p. 603–629, 2008.
- LOHN, J. D.; HORNBY, G. S.; LINDEN, D. S. An evolved antenna for deployment on nasa's space technology 5 mission. In: *Genetic Programming Theory and Practice II*. [S.l.]: Springer, 2005. p. 301–315.

- LUSTGARTEN, J. L. et al. Learning parsimonious classification rules from gene expression data using bayesian networks with local structure. *Data*, Multidisciplinary Digital Publishing Institute, v. 2, n. 1, p. 5, 2017.
- MARI, J. J.; WILLIAMS, P. A validity study of a psychiatric screening questionnaire (srq-20) in primary care in the city of sao paulo. *The British Journal of Psychiatry*, RCP, v. 148, n. 1, p. 23–26, 1986.
- MARSHALL, N. A. et al. Cd4+ t-cell responses to epstein–barr virus (ebv) latent membrane protein 1 in infectious mononucleosis and ebv-associated non-hodgkin lymphoma: Th1 in active disease but tr1 in remission. *British journal of haematology*, Wiley Online Library, v. 139, n. 1, p. 81–89, 2007.
- MATOS, S. M. A. d. et al. Validation of a food frequency questionnaire for children and adolescents aged 4 to 11 years living in salvador, bahia. 2012.
- MATRICARDI, P. M. et al. Exposure to foodborne and orofecal microbes versus airborne viruses in relation to atopy and allergic asthma: epidemiological study. *Bmj*, British Medical Journal Publishing Group, v. 320, n. 7232, p. 412–417, 2000.
- MCQUAID, E. L.; KOPEL, S. J.; NASSAU, J. H. Behavioral adjustment in children with asthma: a meta-analysis. *Journal of Developmental & Behavioral Pediatrics*, LWW, v. 22, n. 6, p. 430–439, 2001.
- MENDONÇA, L. R. et al. Toxocara seropositivity, atopy and wheezing in children living in poor neighbourhoods in urban latin american. *PLoS Negl Trop Dis*, Public Library of Science, v. 6, n. 11, p. e1886, 2012.
- MITRA, A. P. et al. The use of genetic programming in the analysis of quantitative gene expression profiles for identification of nodal status in bladder cancer. *BMC cancer*, BioMed Central, v. 6, n. 1, p. 1, 2006.
- MOORE, J. H.; WHITE, B. C. Genome-wide genetic analysis using genetic programming: The critical need for expert knowledge. In: *Genetic Programming Theory and Practice IV*. [S.l.]: Springer, 2007. p. 11–28.
- MURPHY, K. *Imunobiologia de Janeway-8*. [S.l.]: Artmed Editora, 2014.
- MUTTIL, N.; CHAU, K.-W. Neural network and genetic programming for modelling coastal algal blooms. *International Journal of Environment and Pollution*, Inderscience Publishers, v. 28, n. 3-4, p. 223–238, 2006.
- NILSSON, C. et al. Does early ebv infection protect against ige sensitization? *Journal of allergy and clinical immunology*, Elsevier, v. 116, n. 2, p. 438–444, 2005.
- NUNKESSER, R. et al. Detecting high-order interactions of single nucleotide polymorphisms using genetic programming. *Bioinformatics*, Oxford Univ Press, v. 23, n. 24, p. 3280–3288, 2007.
- PAUL, T. K.; IBA, H. Prediction of cancer class with majority voting genetic programming classifier using gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, IEEE Computer Society Press, v. 6, n. 2, p. 353–367, 2009.

- PERZANOWSKI, M. S. et al. Atopy, asthma, and antibodies to ascaris among rural and urban children in kenya. *The Journal of pediatrics*, Elsevier, v. 140, n. 5, p. 582–588, 2002.
- PERZANOWSKI, M. S. et al. Effect of cat and dog ownership on sensitization and development of asthma among preteenage children. *American journal of respiratory and critical care medicine*, Am Thoracic Soc, v. 166, n. 5, p. 696–702, 2002.
- POLK, S. et al. A prospective study of fel d1 and der p1 exposure in infancy and childhood wheezing. *American journal of respiratory and critical care medicine*, Am Thoracic Soc, v. 170, n. 3, p. 273–278, 2004.
- PONTE, J. C. et al. A study on the immunological basis of the dissociation between type i-hypersensitivity skin reactions to blomia tropicalis antigens and serum anti-b. tropicalis ige antibodies. *BMC immunology*, BioMed Central, v. 12, n. 1, p. 34, 2011.
- QUINLAN, R. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.
- RABINOVITCH, N. et al. The response of children with asthma to ambient particulate is modified by tobacco smoke exposure. *American journal of respiratory and critical care medicine*, Am Thoracic Soc, v. 184, n. 12, p. 1350–1357, 2011.
- RAMEZANKHANI, A. et al. The impact of oversampling with smote on the performance of 3 classifiers in prediction of type 2 diabetes. *Medical decision making*, SAGE Publications, v. 36, n. 1, p. 137–144, 2016.
- RIFFO-VASQUEZ, Y. et al. Effect of mycobacterium tuberculosis chaperonins on bronchial eosinophilia and hyper-responsiveness in a murine model of allergic inflammation. *Clinical & Experimental Allergy*, Wiley Online Library, v. 34, n. 5, p. 712–719, 2004.
- RITCHIE, M. D. et al. Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases. *BMC bioinformatics*, BioMed Central, v. 4, n. 1, p. 28, 2003.
- RODRIGUES, L. et al. Early infection with trichuris trichiura and allergen skin test reactivity in later childhood. *Clinical & Experimental Allergy*, Wiley Online Library, v. 38, n. 11, p. 1769–1777, 2008.
- RODRIGUEZ, A. et al. Lifestyle domains as determinants of wheeze prevalence in urban and rural schoolchildren in ecuador: cross sectional analysis. *Environmental Health*, BioMed Central, v. 14, n. 1, p. 15, 2015.
- SANZ, J. et al. A new survival status prediction system for severe trauma patients based on a multiple classifier system. *Computer methods and programs in biomedicine*, Elsevier, v. 142, p. 1–8, 2017.
- SILVA, R. de C. R. et al. Dietary patterns and wheezing in the midst of nutritional transition: a study in brazil. *Pediatric allergy, immunology, and pulmonology*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 26, n. 1, p. 18–24, 2013.
- SIMPSON, J. L. et al. Inflammatory subtypes in asthma: assessment and identification using induced sputum. *Respirology*, Wiley Online Library, v. 11, n. 1, p. 54–61, 2006.

- SOLÉ, D. et al. Prevalence of symptoms of asthma, rhinitis, and atopic eczema among brazilian children and adolescents identified by the international study of asthma and allergies in childhood (isaac): phase 3. *Jornal de pediatria*, SciELO Brasil, v. 82, n. 5, p. 341–346, 2006.
- SPECTOR, L. et al. Genetic programming for finite algebras. In: ACM. *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. [S.l.], 2008. p. 1291–1298.
- SRINIVAS, N.; DEB, K. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, MIT Press, v. 2, n. 3, p. 221–248, 1994.
- STEIN, R. T.; MARTINEZ, F. D. Asthma phenotypes in childhood: lessons from an epidemiological approach. *Paediatric respiratory reviews*, Elsevier, v. 5, n. 2, p. 155–161, 2004.
- STRACHAN, D. P. Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, BMJ Group, v. 299, n. 6710, p. 1259, 1989.
- SZE-TO, H.-Y. et al. {GP-Pi}: Using Genetic Programming with Penalization and Initialization on Genome-Wide Association Study. *Artificial Intelligence and Soft Computing - 12th International Conference, {ICAISC} 2013, Zakopane, Poland, June 9-13, 2013, Proceedings, Part {II}*, v. 7895, p. 330–341, 2013.
- TAN, K. C. et al. Evolutionary computing for knowledge discovery in medical diagnosis. *Artificial Intelligence in Medicine*, Elsevier, v. 27, n. 2, p. 129–154, 2003.
- THOMAS, D. J.; SMITH, J. M. Joseph raphson, frs. *Notes and Records of the Royal Society of London*, JSTOR, v. 44, n. 2, p. 151–167, 1990.
- TORRES, R. d. S. et al. A genetic programming framework for content-based image retrieval. *Pattern Recognition*, Elsevier, v. 42, n. 2, p. 283–292, 2009.
- TOULOUSE, M. Automatic quantum computer programming: A genetic programming approach. *Genetic Programming and Evolvable Machines*, Springer, v. 7, n. 1, p. 125–126, 2006.
- TSE, K.; HORNER, A. A. Allergen tolerance versus the allergic march: the hygiene hypothesis revisited. *Current allergy and asthma reports*, Springer, v. 8, n. 6, p. 475–483, 2008.
- UY, N. Q. et al. Semantically-based crossover in genetic programming: application to real-valued symbolic regression. *Genetic Programming and Evolvable Machines*, Springer, v. 12, n. 2, p. 91–119, 2011.
- VASUDEV, C. *Graph theory with applications*. [S.l.]: New Age International, 2006.
- VEIGA, R. V. et al. Chronic virus infections suppress atopy but not asthma in a set of children from a large latin american city: a cross-section study. *BMC pulmonary medicine*, BioMed Central, v. 11, n. 1, p. 24, 2011.
- VEIGA, R. V. et al. Using grammar-based genetic programming to determine characteristics of multiple infections and environmental factors in the development of allergies and asthma. In: IEEE. *Evolutionary Computation (CEC), 2015 IEEE Congress on*. [S.l.], 2015. p. 1604–1611.

- WEBB, G. I. Decision tree grafting from the all-tests-but-one partition. In: *IJCAI*. [S.l.: s.n.], 1999. v. 2, p. 702–707.
- WEHRMEISTER, F. C. et al. Time trend of asthma in children and adolescents in brazil, 1998-2008. *Revista de saude publica*, SciELO Brasil, v. 46, n. 2, p. 242–250, 2012.
- WEIL, C. M. et al. The relationship between psychosocial factors and asthma morbidity in inner-city children with asthma. *Pediatrics*, Am Acad Pediatrics, v. 104, n. 6, p. 1274–1280, 1999.
- WEINMAYR, G. et al. International variations in associations of allergic markers and diseases in children: Isaac phase two. *Allergy*, Wiley Online Library, v. 65, n. 6, p. 766–775, 2010.
- WENZEL, S. E. Asthma: defining of the persistent adult phenotypes. *The Lancet*, Elsevier, v. 368, n. 9537, p. 804–813, 2006.
- WHIGHAM, P. A. et al. Grammatically-based genetic programming. In: *Proceedings of the workshop on genetic programming: from theory to real-world applications*. [S.l.: s.n.], 1995. v. 16, n. 3, p. 33–41.
- WINGATE, P. et al. Regulatory t cell activity in primary and persistent epstein–barr virus infection. *Journal of medical virology*, Wiley Online Library, v. 81, n. 5, p. 870–877, 2009.
- WONG, M. L. et al. Discovering knowledge from medical databases using evolutionary algorithms. *IEEE Engineering in Medicine and Biology Magazine*, IEEE, v. 19, n. 4, p. 45–55, 2000.
- WOOD, L. G.; GIBSON, P. G. Dietary factors lead to innate immune activation in asthma. *Pharmacology & therapeutics*, Elsevier, v. 123, n. 1, p. 37–53, 2009.
- World Health Organization, W. H. O. A user’s guide to the Self Reporting Questionnaire (SRQ). *Geneva: World Health Organization*, p. 1–84, 1994.
- YANG, J. et al. Induction of regulatory dendritic cells by lactobacillus paracasei l9 prevents allergic sensitization to bovine  $\beta$ -lactoglobulin in mice. *J Microbiol Biotechnol*, v. 25, p. 1687–1696, 2015.
- YU, J. et al. Feature selection and molecular classification of cancer using genetic programming. *Neoplasia*, Elsevier, v. 9, n. 4, p. 292–IN3, 2007.

## A Resultados adicionais

Tabela 19 – Matriz de covariância

	HSV	VZV	EBV	HAV	T. Gondii	H. Pilory	A. Lumbricoides	T. Trichiura
HSV	0,248	0,016	0,028	0,026	0,019	0,012	0,017	0,011
VZV	0,016	0,248	0,009	0,017	0,012	0,023	0,004	0,003
EBV	0,028	0,009	0,102	0,009	0,008	0,012	0,011	0,004
HAV	0,026	0,017	0,009	0,138	0,014	0,011	0,022	0,024
T. Gondii	0,019	0,012	0,008	0,014	0,150	0,005	0,015	0,012
H. Pilory	0,012	0,023	0,012	0,011	0,005	0,201	0,012	0,007
A. Lumbricoides	0,017	0,004	0,011	0,022	0,015	0,012	0,135	0,040
T. Trichiura	0,011	0,003	0,004	0,024	0,012	0,007	0,040	0,097

Tabela 20 – Mostra comparação dos grupos estudados e excluídos no modelo de infecções virais. Onde n é o número de indivíduos positivos, % é o percentual de indivíduos positivos, N é o número total de indivíduos que apresentam dados para a variável, sd é o desvio padrão.

Variáveis de estudo	População estudada n(%) / 1.296	População excluída n(%) / N	Qui-quadrado ou teste T
N=1445			
HSV	712(54,9)	47(64,4) / 73	0,114
VZV	588(45,4)	31(48,4) / 64	0,631
EBV	1149(88,7)	65(94,2) / 69	0,152
HAV	224(17,3)	15(22,7) / 66	0,257
Asma	295(22,8)	31(20,8) / 149	0,588
SPT	391(30,2)	30(32,6) / 92	0,623
IgE	489(37,7)	24(35,8) / 67	0,753
gênero (homens)	697(53,8)	75(50,3) / 149	0,425
idade			
média (sd)	7,23(1,66)	7,28(1,67)	0,8
Percentis 25/50/75	5,9/7,0/8,7	5,8/6,8/9,0	
Asma nos pais (sim)	173(13,3)	24(18,6) / 129	0,099
fumante na casa (sim)	355(27,4)	43(30,3) / 142	0,465
Rede de esgoto (presença)	1073(82,8)	122(81,9) / 149	0,78
Troca roupa de cama (>= 1 por semana)	573(44,2)	70(47,0) / 149	0,52
número de irmãos			
média(sd)	1,79(1,75)	1,72(1,72)	0,656
Percentis 25/50/75	1/1/2	0/1/2	
gato em casa (sim)	230(17,7)	35(23,5) / 149	0,086
cachorro em casa (sim)	513(39,6)	66(44,3) / 149	0,266
mofo ou umidade (presença)	841(64,9)	93(62,4) / 149	0,549
Escolaridade materna			0,285
primário incompleto	177(13,7)	12(9,4) / 127	
primário completo	444(34,3)	48(37,8) / 127	
Ginásio completo	289(22,3)	23(18,1) / 127	
segundo grau completo	386(29,8)	44(34,6) / 127	
Creche (sim)	208(16,0)	29(19,5) / 149	0,287

**B Artigo publicado**

# Using grammar-based genetic programming to determine characteristics of multiple infections and environmental factors in the development of allergies and asthma

Rafael V. Veiga  
Universidade Federal de Juiz de Fora (UFJF)  
Juiz de Fora, MG, Brazil  
Email: rafael.veiga@uol.com.br

João Marcos de Freitas  
UFJF  
Juiz de Fora, MG, Brazil  
Email: marelo94jf@hotmail.com

Heder S. Bernardino  
UFJF  
Juiz de Fora, MG, Brazil  
Email: heder@ice.ufjf.br

Helio J.C. Barbosa  
Laboratório Nacional de Computação Científica (LNCC/MCTI)  
Petrópolis, RJ, Brazil.  
Universidade Federal de Juiz de Fora (UFJF)  
Juiz de Fora, MG, Brazil  
Email: hcbm@lncc.br

Neuza M. Alcântara-Neves  
Universidade Federal da Bahia (UFBA)  
Salvador, BA, Brazil  
Email: neuza@ufba.br

**Abstract**—In recent decades asthma and allergies had great increase worldwide, being currently a serious global health problem. The causes of these disorders are unknown, but the most accepted hypothesis is that improving hygiene and reducing infections may be the main cause of this increase. Both asthma and allergies are complex diseases with strong environmental influence, so the use of versatile tools such as genetic programming can be important in the understanding of those conditions. We applied genetic programming to data obtained from 1296 children. Data related to chronic viral infections and environmental factors were used to classify in asthmatic and non-asthmatic, IgE and SPT in order to assess allergy. For asthma, viral infections were not relevant while for IgE and SPT they were. The use of genetic programming is shown to be a powerful tool to help understand those conditions.

**Keywords**—genetic programming, classifier, asthma, allergy

## I. INTRODUCTION

Asthma is among the most common chronic diseases worldwide, causing high morbidity [1]. Asthma is a heterogeneous condition and is the final presentation of different pathways. It is associated with diverse environmental factors as well as genetic backgrounds [2]. Atopy is considered an important risk factor for asthma in developed societies, therefore by far the majority of asthma in Latin America is likely to be non allergic [3]. The prevalence of atopy and asthma has increased in affluent countries over recent decades, and has increased also in cities of non-affluent countries such as in Latin America [1]. Such temporal trends occurring over a relatively short period of time are unlikely to be explained by changes in genetic susceptibility and are most likely explained by changes in environmental exposures such as those associated with the adoption of a modern or “westernized” lifestyle [2]. A widely accepted explanation is the hygiene hypothesis [4],

which is based on epidemiologic observations suggesting that populations living in conditions of improved hygiene have reduced exposure to childhood infections which may lead to unbalanced development of the immune system [5], [6].

There is a large body of evidence from experimental animal models and epidemiologic studies of human populations supporting a potentially protective role of exposures to a wide variety of pathogens protecting against atopy and asthma [7], [8], [9], [10], [11], [6], although the knowledge of which infections are more related to asthma and atopy, the role of co-infections and the influence of the environment is still poorly known. Both asthma and atopy are complex and multifactorial phenomena, so the development of models that can explain and predict these conditions is of fundamental importance for their understanding, and provide the basis to develop new hypotheses related to these pathologies.

Most of the models used in epidemiology are based on statistical methods such as multivariate regression and correlation analysis. While generally very powerful, these approaches assume that the variables are independent and that the data can be modeled using linear combinations of these variables. When the relationships are nonlinear and the variables are interdependent (or conditionally dependent) the performance of statistics suffer. It is in those situations where machine learning techniques tend to shine. Many biological systems are fundamentally nonlinear and their parameters conditionally dependent [12]. Unlike statistics, though, machine learning methods can employ Boolean logic (AND, OR, NOT), absolute conditionality (IF, THEN, ELSE), conditional probabilities (the probability of X given Y) which are easier to understand. Recently, machine learning techniques have become popular research tools for medical researchers who seek to identify and exploit patterns and relationships to understand biological phe-

nomena, especially in epidemiology [13], [14], genetics [15] and cancer [16], [17].

Among the techniques of machine learning, genetic programming (GP) appears promising in developing models of asthma and atopy, because it is able to generate models which are easy to understand, to extract useful information, and is able to identify relationships without assuming independence and linear relations between predictor variables, being more in accordance with our understanding of those disorders. This work is a reanalysis of a previous work on the role of chronic viral infections in the development of asthma and atopy [7], which analyzed the same data using standard statistics techniques such as multivariate logistic regression. Here we intend to use GP to generate models for a better understanding of asthma and atopy, and the role of chronic viral infections and environmental factors in the development of these pathologies.

## II. PROBLEM STATEMENT

### A. Data Sets

The data was obtained by an study conducted in the city of Salvador, in northeastern Brazil, with a population of 2.5 million. The general study design has been reported elsewhere [18], [19]. In short, the study population included 1445 unrelated children between 4 and 11 years old, who were recruited in infancy for a prospective study investigating the effect of a citywide sanitation program on childhood diarrhoea morbidity and parasite infections. Data were collected from children, born between 1994 and 2001, who lived in sentinel neighborhoods in the city. The data used in this work were obtained in a survey in 2005 using data collected by questionnaire on asthma by using a Portuguese-adapted phase II International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire and blood samples. Asthma definition was based on the questionnaire and the children were classified as having asthma if parents reported wheezing in the previous 12 months and at least one of the following: (i) diagnosis of asthma ever; (ii) wheezing with exercise in the last 12 months; (iii) 4 wheezing episodes in the last 12 months; (iv) waking up at night because of wheezing in the last 12 months. To evaluate allergy we used the skin prick test(SPT) and the presence of specific IgE in blood for common allergens.

The infections studied were hepatitis A virus(HAV), herpes simplex (HSV), herpes zoster (HZV) and Epstein-Barr (EBV), and they were determined by the presence of IgG antibody against these infections in the children serum. The methodology of evaluated SPT, IgE and serology for viral infections has been described elsewhere [7], [8]. Environmental and socio-economic variables were obtained by means of questionnaire and inspection. Ethical approval for the study was obtained from the Brazilian National Ethical Committee, and written informed consent was obtained from the legal guardian of each child.

### B. Models construction

Three models were built: IgE, SPT and Asthma models, and all input variables were present in the models. Variables were chosen that have neither too high nor too low frequencies, and which can cover several aspects that may be directly or

TABLE I: Variables used in Models

Variables <i>N</i> = 1296	Type	frequency %
target variables		
IgE > 0.70Ku/L	boolean	37.7%
SPT(positives)	boolean	30.2%
Asthma (positives)	boolean	22.8%
input variables		
HSV (positives)	boolean	54.9%
VZV (positives)	boolean	45.4%
EBV (positives)	boolean	88.7%
HAV (positives)	boolean	17.3%
Gender (males)	boolean	53.8%
Age	categorical	
4 and 5		25.7%
6 and 7		41.0%
8 to 11		33.3%
Parental asthma (presence)	boolean	13.3%
Sibling number	categorical	
none		18.6%
1		35.4%
2		23.6%
3 or more		22.4%
Mother's education	categorical	
Illiterate		13.7%
Primary complete		34.3%
Secondary incomplete		22.3%
Secondary complete		29.8%
Daycare ever (yes)	boolean	16.0%
Smoke at home (presence)	boolean	27.4%
Sewage disposal system (presence)	boolean	82.8%
Change bed linen $\geq$ 1 per week	boolean	44.2%
Cat at home (presence)	boolean	17.7%
Dog at home (presence)	boolean	39.6%
Mold/moisture at home (presence)	boolean	64.9%

indirectly related with these diseases. As individual characteristics we choose gender, age, and parental asthma representing ancestry factors. As environmental factors we choose variables like smoking at home, cat at home, sewage disposal system, and other variables, and finally, as representing socio-economic factors, we choose mother's education. The full list of variables used is shown in Table I where, for instance, the last line means that the variable "Mold/moisture at home" is of type "boolean" and is present in "64.9%" of the samples.

## III. GENETIC PROGRAMMING

Genetic programming (GP) is a special type of genetic algorithm (GA) which creates computational artifacts (for instance, computer programs written in a given language) to perform a given task by specifying only what the artifact should achieve. Although GP as it is known nowadays starts with Cramer [20], it was the work by Koza [21] that defined and popularized the method which was subsequently known as "standard GP". In GP, the candidate solutions are referred to as "programs", a high-level structure able to represent a large class of computational artifacts, such as a standard computer program, a numerical function or a classifier in symbolic form.

GP techniques evolve programs, since a population of candidate solutions is improved following the same steps of a GA. The steps of these techniques are presented in Algorithm 1 where "createInitialPopulation" creates the initial population, "evaluatePopulation" finds out how well the candidate solutions perform, "selectFittest" selects the best solutions with respect to their fitness, "crossover" combines each pair of parents generating new candidate solutions which are then mutated in "mutation", and "replace" generates a new population by combining candidate solutions from the

current(parent) and the offspring populations [22].

---

**Algorithm 1:** Algorithm of a GP [22].

---

```

1 begin
2    $population \leftarrow \text{createInitialPopulation}();$ 
3   while a stop condition is not met do
4      $\text{evaluatePopulation}(population);$ 
5      $selected \leftarrow \text{selectFittest}(population);$ 
6      $offspring \leftarrow \text{crossover}(selected);$ 
7      $\text{mutation}(offspring);$ 
8      $\text{replace}(offspring, population);$ 

```

---

Typically, GP differs from standard GA (i) in its representation of the candidate solutions and, consequently, (ii) in the definition of the move operators: crossover and mutation. The different possible ways of representing a candidate solution (such as trees, graphs, etc.) can be used to classify the GP variants. Here, a grammar-based GP (GGP) is used to generate a classifier modeled in “if-then-else” form to relate chronic viral infections and environmental factors with the occurrence (or not) of asthma.

#### A. Grammar Guided GP

Grammar guided GP [23], [24], or grammar-based GP, uses grammars as a way to constrain the representation of the candidate solutions. Grammars can be used to create structures which belong to a specific language, and a formal grammar  $G$  can be defined as [25]

$$G = \{N, \Sigma, R, S\}, \quad (1)$$

where  $N$  is a finite set of nonterminals (each nonterminal is formally delimited by  $\langle$  and  $\rangle$ ),  $\Sigma$  is a finite set of terminals or token symbols which are items that can appear in the language (such as constants, variables, and functions),  $S \in N$  is the start symbol, and  $R$  is a finite set of rules (or productions) which are as

$$(\Sigma \cup N)^* N (\Sigma \cup N)^* ::= (\Sigma \cup N)^*, \quad (2)$$

where  $*$  is the Kleene star operator, and  $\cup$  denotes set union.

Typically, GGP techniques use context-free grammars, a type of grammar in which the left-hand side of each production rule consists of a single nonterminal symbol, that is,

$$N ::= (\Sigma \cup N)^*. \quad (3)$$

The candidate programs in GGP are represented by derivation trees, in which the internal nodes are the nonterminals of the grammar and the leaf nodes are symbols which appear in the language (terminals).

GGP uses a grammar to guide the allowed representation of the candidate programs. The grammar is used in the generation of the initial population and in the genetic operators. When a subtree is selected in the crossover operator, the selection of the other candidate solution’s subtree must consider only subtrees with the same non-terminal. Also, in the mutation operation, the randomly selected subtree is replaced by another with the same non-terminal as root. Examples of crossover and mutation operators are illustrated in Figure 1.

## IV. COMPUTATIONAL EXPERIMENTS

We applied here GGP to the inference of a model to classify subjects with respect to the occurrence of asthma, SPT, and IgE. The dataset is composed by 1296 registers with no missing data (extracted from the original one, which was composed by 1445 registers). We applied the chi-square test in all studied variables comparing the data included and excluded of this study to show that no bias was generated; no significant difference was observed ( $p\text{-value} < 0.05$ ). The frequencies of data values are presented in Table I. It is important to highlight that, as can be seen in Table I, there is a predominance of negative cases in the target variables, which can generate classification issues because of the imbalanced data. Different weights were adopted for the classification errors in the fitness function to analyze the effect of the imbalanced data. Assuming that the pair  $(p_0, p_1)$  is used to calculate the fitnesses, where  $p_0$  is the penalization value when a negative value is incorrectly classified and  $p_1$  is the penalization value otherwise, the following three cases were tested:

- $(1, 1)$  for asthma, SPT and IgE –labeled as UW (unitary weights);
- The adopted weights were calculated in accordance with the proportion of values observed in data, that is,  $(0.80297, 1.32515)$  for IgE,  $(0.71602, 1.65728)$  for SPT, and  $(0.64735, 2.19661)$  for asthma –labeled as PW (proportional weights);
- $(1, 10)$  for asthma, SPT and IgE, labeled as OTW (one-ten weights);

We performed 30 independent runs for each experiment (IgE, SPT, and asthma) and weighted case (UW, PW, and OTW). The data was randomly divided between training (90%) and validation (10%) data: the first group is used to evolve the candidate models while the second one is used to select the model with the smallest generalization error. The search technique is a GGP with randomly generated initial population, tournament selection with 2 individuals, elitism equal to 10% of the population, population size equal to 500 individuals, and maximum tree depth equal to 10. The adopted stop criterion was the maximum allowed number of generations, equal to 1000. Finally, the adopted context-free grammar is composed by the following rules ( $R$ ):

$$\begin{aligned}
\langle expr \rangle & ::= \langle expr \rangle \langle expr \rangle \langle bool \rangle \text{ if - else } | \\
& \quad \langle binary \rangle \\
\langle bool \rangle & ::= \langle binvariable \rangle \langle binary \rangle \langle compbin \rangle | \\
& \quad \text{idade}_{categorica} \langle age \rangle \langle compcat \rangle | \\
& \quad n_i r m a o_{cat} \langle nsiblings \rangle \langle compcat \rangle | \\
& \quad \text{estudo}_{class} \langle study \rangle \langle compcat \rangle | \\
& \quad \langle bool \rangle \langle bool \rangle \langle log \rangle \\
\langle compcat \rangle & ::= < | <= | == | >= | > | != = \\
\langle compbin \rangle & ::= == | != = \\
\langle log \rangle & ::= \text{and} | \text{or} \\
\langle age \rangle & ::= 1 | 2 | 3 \\
\langle nsiblings \rangle & ::= 0 | 1 | 2 | 3 | 4 \\
\langle study \rangle & ::= 1 | 2 | 3 | 4
\end{aligned}$$

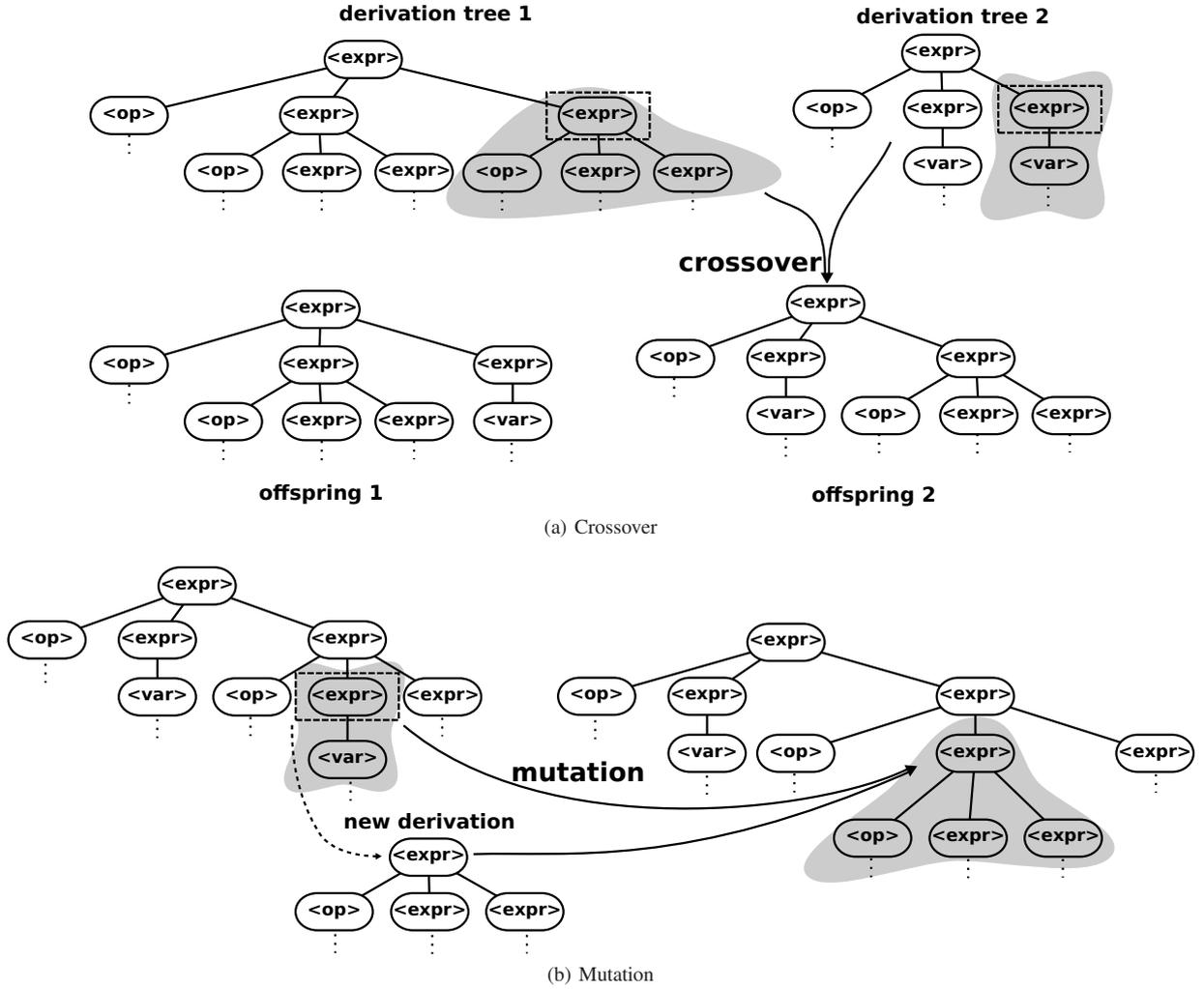


Fig. 1: Crossover and mutation operators of GGP [22].

$\langle binvariable \rangle ::= HSV \mid VZV \mid EBV \mid HAV \mid gender \mid$   
 $parental \ asthma \mid daycare \mid$   
 $smoke \ at \ home \mid sewage \mid bed \ linen \mid$   
 $cat \mid dog \mid mold$

$\langle binary \rangle ::= 0 \mid 1$

where the symbol “|” is used to delimit multiple derivation possibilities.

#### A. Analysis of the Results

The accuracies of the obtained models considering UW, PW, and OTW are presented, respectively, in Tables II, III, and IV. Although different weights were used during the evolution of each case, the results presented in the above mentioned tables were calculated without weights. This work is a reanalysis of previously published work [7], which employed the same database using the technique of multivariate logistic regression. Here we have developed a multivariate logistic regression from the same database. With the obtained model we generated a classification model, by means of a step function and performing a 10-fold cross-validation, that was

used to assess the relative performance of the models obtained by GGP. This model will be labelled here as SLR.

The results as well as the obtained models are discussed in the following.

1) *IgE Model*: The model of IgE showed lower predictive power (Tables II, III, and IV), which was expected because of the stochastic nature of this antibody. In agreement with the previously published work [7], the IgE model showed lower predictive power when compared to the SPT and asthma models. The SLR model achieved 63.7757% of correct results with respect to the training data, and 61.9531% when the validation data is considered. The best model obtained by GGP provides the correct results in 69.23% of the training data and 71.54% in the validation data, which is better than what was given by SLR. Much of the epidemiological studies used total IgE in allergic study as the only measure of atopy, but to increase the specificity, we evaluate instead specific IgE to common allergens and used a cut-off of  $0.70Ku/L$  instead of the usual value which is  $0.35Ku/L$ . The reason for wanting to increase the specificity of IgE measure is that helminth infections stimulate specific IgE against these parasites [26] and recent work has shown cross-reactivity with specific IgE

to common allergens [27]. Two models have reached the best validation classification of 69.23080%, and we chose one of them using the parsimony criterion. Algorithm 2 presents the model which was obtained by GGP, and its simplified version can be found in Algorithm 3.

**Algorithm 2:** Model of IgE with the best validation classification.

```

1 if (sewage disposal != 1) and (VZV == 1) then
2   if (sewage disposal == 0) and (parental asthma !=
3     0) then
4     1;
5   else
6     if age >= "4 and 5" then
7       0;
8     else
9       if age > "8 to 11" then
10        if sibling number == "3 or more" then
11          if sibling number == "3 or more"
12            then
13              if mother education >= 3 then
14                1;
15              else
16                0;
17            else
18              if sibling number > 3 then
19                0;
20              else
21                1;
22            else
23              if age < "4 and 5" then
24                0;
25              else
26                1;
27          else
28            0;

```

**Algorithm 3:** Simplified model of IgE with the best validation classification.

```

1 if (sewage disposal = 0) and (VZV == 1) then
2   if parental asthma == 1 then
3     1;
4   else
5     0;
6 else
7   0;

```

The variables used in this model were sewage disposal system, VZV, parental asthma, age, sibling number, and mother's education. Only the infection of VZV is important to the model, the immune response against the virus is usually IgG instead of IgE, so in principle we would not expect

to find relationships. However, one of the immune mechanisms proposed to explain the hygiene hypothesis is through TH1/TH2 balance of the immune system [28]. Briefly, viral infections stimulate TH1 cytokine production of interferon- $\gamma$  which inhibits the production of cytokines TH2 that in turn stimulate the production of IgE. We were able to find that the infection of VZV is an important factor for IgE positive state, which arises with genetic predisposition (parental asthma) and a low hygiene (sewage disposal) environment. This finding is absent from previous studies [7], [8].

In the other model we find HAV as an important variable. HAV is the only viral infection of the study that has orofecal transmission; this variable represents low hygiene and higher likelihood of having other infections not studied here, such as helminth, that have the same route of transmission, which is known to be associated with increased IgE [26].

2) *SPT Model:* The SPT model showed the lowest standard deviation of the error (Tables II, III, and IV). SLR showed a lower predictive power than GGP: 70,16266% for the training group, and 70.0% for the validation group, against 75.38460% and 75.38460%, respectively, for GGP. Briefly, the SPT is a measure of the size of the allergic response to allergens in the skin. SPT is highly related to IgE; the occurrence of a skin reaction of the type measured by SPT is required to have IgE antibodies. IgE binds to cells of the immune system, like mast cells and eosinophilthus, and when several IgE molecules bound to antigen, it generates a response in the cell that releases factors that cause inflammatory skin reaction that SPT measures. Thus both SPT and IgE may represent atopy. Despite this, recent epidemiological studies have indicated that there is a dissociation between IgE and SPT, being assigned to this dissociation the effect of protection that infections may have on allergy [29], [7]. Two models have reached the best validation classification of 75.38460%, and again we used the parsimony criterion to choose the one presented in Algorithms 4 and 5.

We have found in the model for SPT (Algorithm 5) that VZV is important, but the effect in the model was only in older children that have smoking at home. This presence of VZV in the model leads to absence of SPT, which may indicate the dissociation between IgE and SPT.

3) *Asthma Model:* The International Study of Asthma and Allergies in Childhood Phase II study showed that a higher fraction of recent asthma was attributable to atopy in affluent (41%) compared with nonaffluent (20%) countries [30]. Therefore by far the majority of asthma observed in Latin America is likely to be nonatopic [31], and little is know about risk factors and the mechanism of nonopic asthma. Despite this, the asthma model showed the lowest classification error, reaching about 83,84620% of correct classification (Tables II, III, and IV), and the best model is presented in Algorithms 6 and 7. Again, the method of GGP presents a better predictive power when compared to SLR: 77,61987% for the training group and 76,9531% for the validation group, against 83.84620% and 83.84620%, respectively, for GGP. As can be seen in Algorithm 7, no infection was important for the asthma model, as observed in previous work [7].

From all models of IgE, SPT, and asthma, we observed that some of the variables were more frequent: age, mother's education, sewage disposal, and sibling number, that other

**Algorithm 4:** Model of SPT with the best validation classification.

```

1 if age > "6 and 7" then
2   if (age == "6 and 7") or (smoke at home == 1)
   then
3     if VZV != 0 then
4       if smoke at home == 0 then
5         | 1;
6       else
7         | 0;
8     else
9       if mother education <= 1 then
10      if mother education <= 4 then
11      if (mother education == 3) and
12      (sibling number > 1) then
13      if age >= 2 then
14      | 1;
15      else
16      | 0;
17      else
18      | 0;
19      else
20      | 0;
21      else
22      if (mother education < 1) or ((sibling
23      number < 3) or (mother education ==
24      4)) then
25      | 0;
26      else
27      | 1;
28 else
29 | 1;
30 else
31 | 0;

```

**Algorithm 5:** Simplified model of SPT with the best validation classification.

```

1 if age > "6 and 7" then
2   if smoke at home == 1 then
3     if VZV == 1 then
4       | 0;
5     else
6       if mother education <= 1 then
7         | 0;
8       else
9         if sibling number < 3 then
10        | 0;
11        else
12        | 1;
13     else
14     | 1;
15 else
16 | 0;

```

studies also found as important for this pathologies [32].

Interestingly no viral infection was important for the best model, which is consistent with previous findings [7].

## V. CONCLUSION

Grammar based genetic programming (GGP) has been applied in the epidemiological study of asthma and allergies which have complex relationships with environmental factors. The use of GGP has led to different results when compared with previous studies performed using conventional statistic techniques showing that for both IgE and SPT, viral infections play an important role that was previously identified only for SPT. As compared with asthma, using genetic programming reinforced previous findings using conventional statistics that viral infections have no importance in development of asthma in this population. Due to its flexibility, good performance, and readability of its solutions, it is therefore expected that GGP can be applied to other epidemiological studies where complex relationships may be present.

## ACKNOWLEDGMENT

The authors would like to thank the reviewers for their suggestions and comments, and the support from CNPq (grant 310778/2013-1), CAPES, Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora (PGMC/UFJF), and SCAALA (Social Change, Asthma and Allergy in Latin America) initiative, funded by the Wellcome Trust, Grant No. 072405/Z/03/Z for providing data for this study.

## REFERENCES

- [1] R. Beasley, "Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC," *The Lancet*, vol. 351, no. 9111, pp. 1225–1232, Apr. 1998.
- [2] G. B. Marks, "Environmental factors and gene-environment interactions in the aetiology of asthma," *Clinical and experimental pharmacology & physiology*, vol. 33, no. 3, pp. 285–9, Mar. 2006.
- [3] P. J. Cooper, L. C. Rodrigues, a. a. Cruz, and M. L. Barreto, "Asthma in Latin America: a public health challenge and research opportunity," *Allergy*, vol. 64, no. 1, pp. 5–17, Jan. 2009.
- [4] D. P. Strachan, "Hay fever, hygiene, and household size," *BMJ (Clinical research ed.)*, vol. 299, no. 6710, pp. 1259–60, Nov. 1989.
- [5] D. Daley, "The evolution of the hygiene hypothesis: the role of early-life exposures to viruses and microbes and their relationship to asthma and allergic diseases," *Current opinion in allergy and clinical immunology*, vol. 14, no. 5, pp. 390–6, Oct. 2014.
- [6] T. Kamradt, R. Göggel, and K. J. Erb, "Induction, exacerbation and inhibition of allergic and autoimmune diseases by infection," *Trends in immunology*, vol. 26, no. 5, pp. 260–7, May 2005.
- [7] R. V. Veiga, S. S. Cunha, V. C. C. Dattoli, A. C. Cruz, P. J. Cooper, L. C. Rodrigues, M. L. Barreto, and N. M. Alcântara-Neves, "Chronic virus infections suppress atopy but not asthma in a set of children from a large Latin American city: a cross-section study," *BMC pulmonary medicine*, vol. 11, no. 1, p. 24, Jan. 2011.
- [8] N. M. Alcântara-Neves, R. V. Veiga, V. C. C. Dattoli, R. L. Fiaccone, R. Esquivel, A. A. Cruz, P. J. Cooper, L. C. Rodrigues, and M. L. Barreto, "The effect of single and multiple infections on atopy and wheezing in children," *The Journal of allergy and clinical immunology*, vol. 129, no. 2, pp. 359–67, 367.e1–3, Mar. 2012.

TABLE II: Accuracy of the obtained models considering UW.

	group	max	median	mean	sd	min
IgE model	training	69.23080	62.30770	61.56410	4.12756	48.46150
	validation	70.00000	63.46155	63.38461	3.35969	56.15380
SPT model	training	75.38460	68.84615	68.48718	3.24151	59.23080
	validation	76.15380	69.23080	69.56410	3.29271	60.00000
Asthma model	training	83.07690	76.92310	76.61539	3.90981	66.92310
	validation	83.84620	77.69230	77.51282	3.75044	69.23080

TABLE III: Accuracy of the obtained models considering PW.

	group	max	median	mean	sd	min
IgE model	training	69.23080	61.53850	61.17948	4.57424	48.46150
	validation	71.53850	63.07690	62.84615	3.41720	57.69230
SPT model	training	75.38460	68.46150	68.33332	3.64226	59.23080
	validation	75.38460	68.46150	68.69230	3.45283	59.23080
Asthma model	training	83.84620	77.69230	77.12821	3.95307	69.23080
	validation	83.84620	78.07690	77.46154	3.76926	69.23080

TABLE IV: Accuracy of the obtained models considering OTW.

	group	max	median	mean	sd	min
IgE model	training	69.23080	61.53850	61.43590	4.67571	48.46150
	validation	69.23080	63.07690	63.20513	3.61414	56.92310
SPT model	training	75.38460	68.46150	68.53845	3.60999	59.23080
	validation	75.38460	68.84615	69.25640	3.40124	60.76920
Asthma model	training	83.84620	77.69230	77.12821	3.95307	69.23080
	validation	83.84620	77.69230	77.20513	3.86195	69.23080

- [9] L. R. Mendonça, R. V. Veiga, V. C. C. Dattoli, C. A. Figueiredo, R. Fiaccone, J. Santos, A. A. Cruz, L. C. Rodrigues, P. J. Cooper, L. C. Pontes-de Carvalho, M. L. Barreto, and N. M. Alcântara-Neves, "Toxocara seropositivity, atopy and wheezing in children living in poor neighbourhoods in urban Latin American." *PLoS neglected tropical diseases*, vol. 6, no. 11, p. e1886, Jan. 2012.
- [10] L. C. Rodrigues, P. J. Newcombe, S. S. Cunha, N. M. Alcântara-Neves, B. Genser, a. a. Cruz, S. M. Simoes, R. Fiaccone, L. Amorim, P. J. Cooper, and M. L. Barreto, "Early infection with *Trichuris trichiura* and allergen skin test reactivity in later childhood." *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, vol. 38, no. 11, pp. 1769–77, Nov. 2008.
- [11] Y. Riffo-Vasquez and D. Spina, "Effect of Mycobacterium tuberculosis chaperonins on bronchial eosinophilia and hyper-responsiveness in a murine model of allergic inflammation," *Clinical & ...*, pp. 712–719, 2004.
- [12] J. A. Cruz and D. S. Wishart, "Applications of Machine Learning in Cancer Prediction and Prognosis," *Cancer Informatics*, vol. 2, pp. 59–77, Feb. 2006.
- [13] Y. Mansiaux and F. Carrat, "Detection of independent associations in a large epidemiologic dataset: a comparison of random forests, boosted regression trees, conventional and penalized logistic regression for identifying independent factors associated with H1N1pdm influenza infectio," *BMC medical research methodology*, vol. 14, no. 1, p. 99, Jan. 2014.
- [14] A. Ramezankhani, O. Pournik, J. Shahrabi, D. Khalili, F. Azizi, and F. Hadaegh, "Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study." *Diabetes research and clinical practice*, vol. 105, no. 3, pp. 391–8, Sep. 2014.
- [15] S. Bornelöv, A. Sääf, E. Melén, A. Bergström, B. Torabi Moghadam, V. Pulkkinen, N. Acevedo, C. Orsmark Pietras, M. Ege, C. Braun-Fahrlander, J. Riedler, G. Doekes, M. Kabesch, M. van Hage, J. Kere, A. Scheynius, C. Söderhäll, G. Pershagen, and J. Komorowski, "Rule-based models of the interplay between genetic and environmental factors in childhood allergy." *PloS one*, vol. 8, no. 11, p. e80080, Jan. 2013.
- [16] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods." *Artificial intelligence in medicine*, vol. 34, no. 2, pp. 113–27, Jun. 2005.
- [17] H. Shin and Y. Nam, "A coupling approach of a predictor and a descriptor for breast cancer prognosis." *BMC medical genomics*, vol. 7 Suppl 1, no. Suppl 1, p. S4, Jan. 2014.
- [18] M. L. Barreto, S. S. Cunha, N. Alcântara-Neves, L. P. Carvalho, A. a. Cruz, R. T. Stein, B. Genser, P. J. Cooper, and L. C. Rodrigues, "Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study)." *BMC pulmonary medicine*, vol. 6, p. 15, Jan. 2006.
- [19] M. L. Barreto, B. Genser, A. Strina, M. G. Teixeira, A. M. O. Assis, R. F. Rego, C. A. Teles, M. S. Prado, S. M. A. Matos, D. N. Santos, L. A. dos Santos, and S. Cairncross, "Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies." *Lancet*, vol. 370, no. 9599, pp. 1622–8, Nov. 2007.
- [20] N. L. Cramer, "A representation for the adaptive generation of simple sequential programs," in *Proc. of the 1st International Conference on Genetic Algorithms*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1985, pp. 183–187.
- [21] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. The MIT Press, December 1992.
- [22] H. J. C. Barbosa and H. S. Bernardino, *Computational Technology Reviews*. Civil-Comp, 2011, vol. 4, ch. Genetic Programming in Civil, Structural and Environmental Engineering, pp. 115–145.
- [23] P. Whigham, "Grammatically-based genetic programming," in *Proc. of the Workshop on Genetic Programming: From Theory to Real-World Applications*, J. P. Rosca, Ed., Tahoe City, California, USA, 1995, pp. 33–41.
- [24] P. A. Whigham, "A schema theorem for context-free grammars," in *Proc. of the IEEE International Conference on Evolutionary Computation*, vol. 1, 1995, p. 178.
- [25] N. Chomsky, *Syntactic Structures*. Mouton de Gruyter, 2002.
- [26] J. D. Turner, H. Faulkner, J. Kamgno, M. W. Kennedy, J. Behnke, M. Boussinesq, and J. E. Bradley, "Allergen-specific IgE and IgG4 are markers of resistance and susceptibility in a human intestinal nematode infection." *Microbes and infection / Institut Pasteur*, vol. 7, no. 7-8, pp. 990–6, Jun. 2005.
- [27] N. Acevedo, J. Sánchez, A. Erler, and D. Mercado, "IgE cross-reactivity between *Ascaris* and domestic mite allergens: the role of tropomyosin

**Algorithm 6:** Asthma Model with the best validation classification.

```

1 if age <= "6 and 7" then
2   if mother education >= 3 then
3     if dog != 1 then
4       if mold == 0 then
5         1;
6       else
7         if smoke at home == 0 then
8           1;
9         else
10          0;
11      else
12        if age == "8 to 11" then
13          0;
14        else
15          if sibling number <= 0 then
16            if age < "8 to 11" then
17              1;
18            else
19              0;
20          else
21            if age <= "6 and 7" then
22              0;
23            else
24              1;
25      else
26        0;
27 else
28   if sibling number >= 2 then
29     if daycare == 1 then
30       0;
31     else
32       1;
33   else
34     if age == "8 to 11" then
35       if (sibling number >= 0) and (cat != 1) then
36         0;
37       else
38         if sibling number > 0 then
39           0;
40         else
41           1;
42     else
43       if (sibling number < 0) and (parental asthma == 0) then
44         1;
45       else
46         if (sibling number == 2) or (dog == 1) then
47           0;
48         else
49           if age <= "4 to 5" then
50             if sibling number > 1 then
51               0;
52             else
53               1;
54           else
55             if sibling number < 1 then
56               1;
57             else
58               0;

```

**Algorithm 7:** Simplified asthma model with the best validation classification.

```

1 if age <= "6 and 7" then
2   if mother education >= 3 then
3     if dog == 0 then
4       if mold == 0 then
5         1;
6       else
7         if smoke at home == 0 then
8           1;
9         else
10          0;
11      else
12        0;
13     else
14       0;
15 else
16   if sibling number >= 2 then
17     if daycare == 1 then
18       0;
19     else
20       1;
21   else
22     if cat == 0 then
23       0;
24     else
25       if sibling number > 0 then
26         0;
27       else
28         1;

```

and the nematode polyprotein ABA-1," *Allergy*, vol. 64, no. 11, pp. 1635–43, Nov. 2009.

- [28] S. Romagnani, "The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both?" *Immunology*, vol. 112, no. 3, pp. 352–63, Jul. 2004.
- [29] J. a. C. Ponte, S. B. Junqueira, R. V. Veiga, M. L. Barreto, L. C. Pontes-de Carvalho, and N. M. Alcântara-Neves, "A study on the immunological basis of the dissociation between type I-hypersensitivity skin reactions to *Blomia tropicalis* antigens and serum anti-*B. tropicalis* IgE antibodies." *BMC immunology*, vol. 12, no. 1, p. 34, Jan. 2011.
- [30] G. Weinmayr, S. K. Weiland, B. Björkstén, B. Brunekreef, G. Büchele, W. O. C. Cookson, L. Garcia-Marcos, M. Gotua, C. Gratzou, M. van Hage, E. von Mutius, M.-A. Riiikjäv, P. Rzehak, R. T. Stein, D. P. Strachan, J. Tsanakas, K. Wickens, and G. W. Wong, "Atopic sensitization and the international variation of asthma symptom prevalence in children." *American journal of respiratory and critical care medicine*, vol. 176, no. 6, pp. 565–74, Sep. 2007.
- [31] S. S. Cunha, M. L. Barreto, R. L. Fiaccone, P. J. Cooper, N. M. Alcântara-Neves, S. d. M. a. Simões, A. A. Cruz, and L. C. Rodrigues, "Asthma cases in childhood attributed to atopy in tropical area in Brazil." *Revista panamericana de salud pública*, no. 6, pp. 405–11, Dec. 2010.
- [32] S. S. Cunha, M. Pujades-Rodriguez, M. L. Barreto, B. Genser, and L. C. Rodrigues, "Ecological study of socio-economic indicators and prevalence of asthma in schoolchildren in urban Brazil," *BMC Public Health*, vol. 7, no. 205, pp. 1–6, 2007.

## C Questionário aplicado

# Questionário sobre problemas respiratórios.



**Universidade Federal da Bahia**  
**Instituto de Saúde Coletiva**

Rua Padre Feijó, no. 29 / 4o. andar, Salvador - Bahia, Brasil 40-110-170  
# (071) 245-0151 / 245-0544 FAX (071) 237-5856 e-mail: lis@ufba.br

REG: \_\_\_\_\_

MICRO-ÁREA: \_\_\_\_\_

1. Nome completo da criança: \_\_\_\_\_
  2. Endereço completo: \_\_\_\_\_
  3. Referência da residência: \_\_\_\_\_
  4. Nome completo da mãe: \_\_\_\_\_
- Apelido: \_\_\_\_\_

## Notas para preenchimento do questionário.

1. faça a pergunta como está escrito;
2. não dizer as respostas “sim” ou “não”, deixar a pessoa responder primeiro;
3. se a pessoa disser que não entendeu a pergunta, repita a pergunta novamente;
4. em perguntas sobre 12 meses, se a pessoa não entender pergunte “de mês de maio (exemplo) ano passado até agora, quer dizer, este mês de maio deste ano”;
5. se após repetir a pessoa continuar não entendendo, não marque a questão, deixe em branco, e dizer “e ´doença de pele´ e vamos pular a pergunta”;
6. preste atenção nas notas que seguem as perguntas escritas em negrito, itálico e entre parênteses, por exemplo “(assinale apenas uma opção abaixo)”; ou no **rodapé** das páginas;
7. A versão final deve ser escrita com **CANETA AZUL**, por isso primeiro ouça atentamente a resposta antes de escrever. Se tiver que mudar algo anotar de forma legível e com **LETRA DE IMPRENSA**;
8. **OCUPAÇÃO** é no que a pessoa trabalha agora. Por exemplo, uma pessoa que é pedreiro como profissão (quer dizer tem algum curso/formação como pedreiro) atualmente pode estar trabalhando como cobrador de ônibus. Cobrador de ônibus é a ocupação e é o que deve ser anotado no questionário.
9. **COMEÇAR DIZENDO**: “algumas perguntas que vou fazer podem ser repetidas...”
10. Nas perguntas que tem “seu filho(a)” pode dizer também “ele(a)” ou citar o nome da criança.

O espaço abaixo é para você anotar observações


Data de preenchimento do questionário: \_\_\_\_/\_\_\_\_/ 2005

Nome do(a) entrevistador(a): \_\_\_\_\_ Código: \_\_\_\_\_

## Características demográficas

1. SEXO DA CRIANÇA: menino o-1 menina o-0 QP1

2. DATA DE NASCIMENTO DA CRIANÇA: \_\_\_\_/\_\_\_\_/\_\_\_\_ QP2  
 (ver documento) Viu documento? Sim o-1 Dia Mês Ano  
 Não o-0 QP3

3. DATA DE NASCIMENTO DA MÃE: \_\_\_\_/\_\_\_\_/\_\_\_\_ QP4

4. TELEFONE DE CONTATO (da mãe ou outra pessoas): \_\_\_\_\_ NOME: \_\_\_\_\_

5. ATÉ QUANDO A MÃE ESTUDOU? (assinale apenas uma opção abaixo)

o-1. não sabe ler nem escrever  
 o-2. Primário incompleto  
 o-3. Primário completo QP5  
 o-4. ginásio incompleto (estudou entre a 5ª e 8ª série)  
 o-5. ginásio completo (8ª série completa)  
 o-6. 2º grau incompleto  
 o-7. 2º grau completo  
 o-8. superior incompleto  
 o-9. superior completo

6. OCUPAÇÃO PRINCIPAL DO PAI OU PADRASTO: \_\_\_\_\_ Não sabe - o QP6

7. OCUPAÇÃO PRINCIPAL DA MÃE: \_\_\_\_\_ o-Refere **APENAS** dona de casa QP7

8. SE MÃE REFERIR OUTRA OCUPAÇÃO ANOTAR: \_\_\_\_\_ QP8

9. QUEM RESPONDEU A ESTE QUESTIONÁRIO? QP9  
 Mãe o-1  
 Pai o-2  
 Outra pessoa o-3 Quem? • \_\_\_\_\_ QP10

## Condições da moradia e ambientais

☞ *Atenção: as perguntas em cinza são para o entrevistador observar e responder, não pergunte ao responsável*

<p>1. VERIFICAR A PAVIMENTAÇÃO (<i>observar</i>) (<i>assinale apenas 1 resposta</i>)</p> <p>o-0.rua e calçada são pavimentadas</p> <p>o-1.somente a rua é pavimentada</p> <p>o-2.somente a calçada é pavimentada</p> <p>o-3.nem rua nem calçada são pavimentadas</p> <p>o-4.caminho/escada pavimentada</p> <p>o-5.caminho/escada não pavimentado</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP11</div>
<p>2. VERIFICAR O TIPO DE HABITAÇÃO (<i>observar</i>) (<i>assinale apenas 1 resposta</i>)</p> <p>o-0.casa</p> <p>o-1.apartamento</p> <p>o-2.barraco</p> <p>o-3.quarto (<i>quer dizer 1 cômodo e sempre de tijolos, se barraco anotar em "2"</i>)</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP12</div>
<p>3. HÁ QUANTO TEMPO A SENHORA MORA NESTA HABITAÇÃO?</p> <p>o-0. Há menos de 6 meses</p> <p>o-1. De 6 meses até menos de 2 anos</p> <p>o-2. De 2 anos até menos de 5 anos</p> <p>o-3. Há 5 anos ou mais</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP13</div>
<p>4. VERIFICAR O MATERIAL E O ACABAMENTO PREDOMINANTE (MAIS DE 50%) DAS PAREDES INTERNAS DA SALA (<i>cômodo em que acontece a entrevista; observar</i>) (<i>assinale apenas 1 resposta</i>)</p> <p>o-0.blocos com revestimento</p> <p>o-1.blocos sem revestimento</p> <p>o-2.reaproveitamento de madeira, lata ou papelão</p> <p>o-3.taipa</p> <p>o-4.pré-moldada em concreto</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP14</div>
<p>5. ESSA CASA É: (<i>assinale apenas 1 resposta</i>)</p> <p>o-0.própria</p> <p>o-1.alugada</p> <p>o-2.sublocada/cedida</p> <p>o-3.NS/NR</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP15</div>
<p>6. QUANTOS CÔMODOS TEM SUA HABITAÇÃO? (<i>excluir banheiro e varanda</i>): <input style="width: 50px;" type="text"/></p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">OP16</div>
<p>7. QUANTOS CÔMODOS SÃO USADOS PARA DORMITÓRIO? <input style="width: 50px;" type="text"/></p> <p>(<i>Inclui quarto, sala, etc, é onde as pessoas dormem</i>)</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP17</div>
<p>8. DE ONDE VEM A ÁGUA USADA NA SUA HABITAÇÃO? (<i>pode explicar</i>)</p> <p>o-0.água da Embasa (incluindo 'gato') com canalização interna</p> <p>o-1.água da Embasa (incluindo 'gato') exclusivamente com ponto de água externo (<i>pode ser do vizinho</i>)</p> <p>o-2.água de chafariz</p> <p>o-3.Outro (especificar) _____</p>	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">QP1</div> <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto; margin-top: 10px;">QP19</div>

9. PARA ONDE VÃO OS DEJETOS SANITÁRIOS (FEZES E URINA) DA SUA HABITAÇÃO?

- O-0.rede de esgoto  
O-1.rede de drenagem  
O-2.escadaria drenante  
O-3.fossa séptica [fechada (esvaziada periodicamente) ou seguida de sumidouro]  
O-4.fossa séptica [fossa rudimentar,lançamento na rua ou na rede de drenagem]  
O-5.buraco escavado  
O-6.canal ou vala a céu aberto na rua, riacho, mar  
O-7. Não sabe / não respondeu

**QP20**

## Perguntas sobre PROBLEMAS RESPIRATÓRIOS

1. ALGUMA VEZ NA VIDA SEU FILHO/A TEVE CHIADO (OU PIADO) NO PEITO? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0 (Se a resposta foi "NÃO", passe à pergunta 6,7,8)</li> </ul>	QP21
2. NOS ÚLTIMOS 12 MESES SEU FILHO/A TEVE CHIADO (OU PIADO) NO PEITO? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0 (Se a resposta foi "NÃO", passe à pergunta 6,7,8)</li> </ul>	QP22
3. NOS ÚLTIMOS 12 MESES QUANTAS CRISES/ATAQUES DE CHIADO (OU PIADO) NO PEITO SEU FILHO(A) TEVE? <ul style="list-style-type: none"> <li>• Nenhuma o-0</li> <li>• 1 a 3 o-1</li> <li>• 4 a 12 o-2</li> <li>• Mais de 12 o-3</li> </ul>	QP23
4. NOS ÚLTIMOS 12 MESES QUANTAS VEZES SEU FILHO/A ACORDOU À NOITE POR CAUSA DE CHIADO (OU PIADO)? <ul style="list-style-type: none"> <li>• Nunca se acordou com chiado o-0</li> <li>• Menos de uma noite por semana, quer dizer, nem todas as semanas o-1</li> <li>• Uma ou mais de uma noite por semana o-2</li> </ul>	QP24
5. NOS ÚLTIMOS 12 MESES OS CHIADOS (OU PIADO) NO PEITO TÊM SIDO TÃO GRAVES QUE SEU FILHO NÃO CONSEGUE DIZER DUAS PALAVRAS SEGUIDAS SEM QUE TENHA QUE PARAR PARA RESPIRAR? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP25
6. SEU FILHO/A TEVE ASMA ALGUMA VEZ NA VIDA? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP26
7. NOS ÚLTIMOS 12 MESES VOCÊ NOTOU CHIADO (OU PIADO) NO PEITO DE SEU FILHO/A AO RESPIRAR, DURANTE OU DEPOIS DE FAZER EXERCÍCIO (CORRER, JOGAR BOLA, PULAR, ETC) ? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP27
8. NOS ÚLTIMOS 12 MESES SEU FILHO/A TEM APRESENTADO TOSSE SECA À NOITE, QUE NÃO TENHA SIDO A TOSSE POR RESFRIADO OU GRIPE? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP28

## Questionário sobre PROBLEMAS DE NARIZ

1.	<u>ALGUMA VEZ NA VIDA</u> SEU FILHO/A TEVE ESPIRROS, QUANDO <b>NÃO</b> ESTAVA RESFRIADO(A) OU COM GRIPE? <ul style="list-style-type: none"><li>• Sim o-1</li><li>• Não o-0 (Se a resposta foi "NÃO", passe à pergunta 5)</li></ul>	QP29
2.	<u>NOS ÚLTIMOS 12 MESES</u> SEU FILHO/A TEVE ESPIRROS, QUANDO <b>NÃO</b> ESTAVA RESFRIADO OU COM GRIPE? <ul style="list-style-type: none"><li>• Sim o-1</li><li>• Não o-0 (Se a resposta foi "NÃO", passe à pergunta 5)</li></ul>	QP30
3.	<u>NOS ÚLTIMOS 12 MESES</u> , SEU FILHO/A TEM APRESENTADO ESTES PROBLEMAS DE NARIZ ACOMPANHADOS DE COCEIRA E LACRIMEJAMENTO NOS OLHOS? <ul style="list-style-type: none"><li>• Sim o-1</li><li>• Não o-0</li></ul>	QP31
4.	<u>NOS ÚLTIMOS 12 MESES</u> QUANTAS VEZES OS PROBLEMAS DE NARIZ IMPEDIRAM SEU FILHO/A DE FAZER SUAS ATIVIDADES DIÁRIAS? <ul style="list-style-type: none"><li>• Nunca o-0</li><li>• Poucas vezes o-1</li><li>• Mais que pouca vezes o-2</li><li>• Muito freqüentemente o-3</li></ul>	QP32
5.	SEU FILHO/A TEVE <u>ALGUMA VEZ NA VIDA</u> ALERGIA NASAL OU RINITE ALÉRGICA? <ul style="list-style-type: none"><li>• Sim o-1</li><li>• Não o-0</li></ul>	QP33

## Questionário sobre PROBLEMAS DE PELE

1. **ALGUMA VEZ NA VIDA** SEU FILHO/A APRESENTOU IRRITAÇÕES NA PELE QUE IAM E VOLTAVAM E QUE COÇAVAM, DURANDO PELO MENOS SEIS MESES?

QP34

- Sim o-1
- Não o-0 *(Se a resposta foi "NÃO", passe à pergunta 7,8)*

2. **NOS ÚLTIMOS 12 MESES** SEU FILHO/A APRESENTOU ESTAS IRRITAÇÕES NA PELE QUE IAM E VOLTAVAM E QUE COÇAVAM?

QP35

- Sim o-1
- Não o-0 *(Se a resposta foi "NÃO", passe à pergunta 7,8)*

3. **ALGUMA VEZ NA VIDA** ESTAS IRRITAÇÕES NA PELE QUE COÇAVAM APARECERAM EM ALGUM DESTES LUGARES?

QP36

*Dobras dos cotovelos, ou atrás dos joelhos, ou na frente dos tornozelos, ou na dobra entre coxa e nádega, ou ao redor do pescoço, ou nos olhos ou orelhas? (perguntar sobre esses lugares mesmo que tenha respondido "em todo o corpo")*

- Sim o-1 *(assinalar se respondeu em pelo menos um dos lugares acima)*
- Não o-0

4. **COM QUE IDADE** APARECERAM NO SEU FILHO/A, PELA PRIMEIRA VEZ, ESTAS IRRITAÇÕES NA PELE QUE COÇAVAM?

QP37

- Antes dos 2 anos o-0
- 2 a 4 anos o-1
- 5 anos ou mais o-2

5. **NOS ÚLTIMOS 12 MESES** ESTAS IRRITAÇÕES NA PELE **CHEGARAM A DESAPARECER** COMPLETAMENTE?

QP38

- Sim o-1
- Não o-0

6. **NOS ÚLTIMOS 12 MESES** QUANTAS VEZES SEU FILHO/A TEVE QUE SE **LEVANTAR DA CAMA À NOITE** PORQUE ESTAS IRRITAÇÕES NA PELE COÇAVAM?

QP39

- Nunca nos 12 últimos meses o-0
- Menos de uma noite por semana, quer dizer, nem todas as semanas o-1
- Uma ou mais vezes por semana o-2

7. SEU FILHO/A TEVE ALGUMA VEZ NA VIDA **ECZEMA OU DERMATITE ATÓPICA**?   
*(se não souber responder repita a pergunta, se continuar não sabendo, deixar em branco e dizer "vamos pular")*

QP40

- Sim o-1
- Não o-0

8. **NOS ÚLTIMOS 12 MESES**, SEU FILHO(A) SOFREU DE **PELE SECA** EM TODO O CORPO?

QP41

- Sim o-1
- Não o-0

## Chiado e falta de ar

1. NOS ÚLTIMOS 12 MESES SEU FILHO/A TEVE CANSAÇO?

- Sim o-1
- Não o-0

QP42

2. ALGUMA VEZ NA VIDA SEU FILHO/A ACORDOU COM FALTA DE AR?

- Sim o-1
- Não o-0

QP43

**ATENÇÃO:** faça as perguntas abaixo APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS RESPIRATÓRIOS” (página 5).

Fez as perguntas abaixo? • Sim o-1

- Não (quer dizer pulou as perguntas abaixo) o-0

QP44

3. NOS ÚLTIMOS 12 MESES O QUE TEM PIORADO O CHIADO DE SEU FILHO/A?

(dizer “vou citar alguns itens e você me diz se tem piorado o chiado”; marcar tantas opções quanto necessário)

- o-Mudanças no tempo QP45
- o-Pólen (explicar dizendo “pozinho da flor”) QP46
- o-Estado Nervoso QP47
- o-Fumaças QP48
- o-Poeira de casa QP49
- o-Animais domésticos QP50
- o-Resfriados/gripe QP51
- o-Fumaça de cigarro QP52
- o-Comidas ou bebidas QP53 (indicar quais): \_\_\_\_\_ QP54
- o-Sabonetes, sprays ou detergentes QP55
- o-Outras coisas por favor, enumerar): QP56 \_\_\_\_\_ QP57
- o-Todas opções acima negativas QP58

## Tratamentos para doenças respiratórias

**ATENÇÃO:** faça as TODAS AS perguntas abaixo (1 E 2) APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS RESPIRATÓRIOS” (página 5).

Fez as perguntas abaixo: • Sim o-1

QP59

• Não (quer dizer que pulou as perguntas abaixo) o-0

1. NOS ÚLTIMOS 12 MESES SEU FILHO/A USOU ALGUM REMÉDIO (COMPRIMIDOS, XAROPES) PARA CHIADO OU PIADO OU ASMA?

- Sim o-1 (*Se respondeu "SIM", escreva o nome dos medicamentos abaixo*)
- Não o-0

QP60

**Notas 1-** *Pode assinalar mais de uma opção; 2-pregunta se tomou remédio, depois nome e depois sobre frequência*

<u>Remédios de farmácia</u> (dizer “remédios comprados em farmácias, igrejas, catálogos”, e ver caixa)	Só quando tinha piado	Regularmente Todo dia por pelo menos 2 meses ao ano	Outro Escreva ao lado
<u>Nome do remédio</u>			
QP61	.....o-0	.....o-1.....o-9	QP63
QP64	.....o-0	.....o-1.....o-9	QP66
QP67	.....o-0	.....o-1.....o-9	QP69
QP70	.....o-0	.....o-1.....o-9	QP72
QP73	.....o-0	.....o-1.....o-9	QP75
<u>Remédios caseiros (quer dizer “feitos em casa”)</u>	Só quando tinha piado	Regularmente Todo dia por pelo menos 2 meses ao ano	Outro Escreva ao lado
<u>Nome do remédio</u>			
QP76	.....o-0	.....o-1.....o-9	QP78
QP79	.....o-0	.....o-1.....o-9	QP81

2. NOS ÚLTIMOS 12 MESES SEU FILHO(A) USOU ALGUM REMÉDIO (COMPRIMIDOS, XAROPES) PARA CHIADO OU PIADO OU PARA ASMA ANTES, DURANTE OU DEPOIS DE EXERCÍCIO FÍSICO (CORRER, JOGAR BOLA, PULAR, ETC)?

- Sim o-1 (*Se respondeu "SIM", escreva o nome dos medicamentos*)
- Não o-0

QP82

**Notas: 1-** *Pode assinalar mais de uma opção; 2-pregunta se tomou remédio, depois nome e depois sobre frequência*

<u>Remédios de farmácia</u> (dizer “remédios comprados em farmácias, igrejas, catálogos”, e ver caixa)	Só quando tinha piado	Regularmente Todo dia por pelo me- nos 2 meses ao ano	Outro Escreva ao lado
<u>Nome do remédio</u>			
QP83	.....o-0	.....o-1.....o-9	QP85
QP86	.....o-0	.....o-1.....o-9	QP88
QP89	.....o-0	.....o-1.....o-9	QP91
QP92	.....o-0	.....o-1.....o-9	QP94
QP95	.....o-0	.....o-1.....o-9	QP97
<u>Remédios caseiros (dizer “feitos em casa”)</u>	Só quando tinha piado	Regularmente Todo dia por pelo me- nos 2 meses ao ano	Outro Escreva ao lado
<u>Nome do remédio</u>			
QP98	.....o-0	.....o-1.....o-9	QP100
QP101	.....o-0	.....o-1.....o-9	QP103

**ATENÇÃO: faça as perguntas PARA TODAS AS PERGUNTAS abaixo APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS RESPIRATÓRIOS” (página 5).**

Fez as perguntas abaixo? • Sim o-1

QP104

• Não (quer dizer pulou as perguntas abaixo) o-0

3. VOCÊ TEM UM PLANO DE TRATAMENTO POR ESCRITO QUE LHE EXPLICA COMO CUIDAR DA ASMA DE SEU FILHO(A)? *Dizer “por escrito quer dizer se tem um papel do médico(a) ou alguma cartilha ou folheto, com instruções sobre asma, o que é a doença, como tratar, etc.”*

- Sim ..... o-1
- Não ..... o-0

QP105

4. NOS ÚLTIMOS 12 MESES QUANTAS VISITAS SEU FILHO/A FEZ A QUALQUER UM DOS SEGUINTE PROFISSIONAIS DE SAÚDE POR CHIADO OU PIADO NO PEITO OU ASMA?

a) POR UM ATAQUE DE CHIADO OU PIADO NO PEITO OU ASMA?

Agente de Saúde	QP106	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Enfermeiro/a	QP107	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Médico <sup>1</sup>	QP108	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Sala emergência hospital	QP109	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3

b) PARA UMA VISITA DE REVISÃO DE CHIADO OU PIADO NO PEITO OU ASMA?

Agente de Saúde	QP110	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Enfermeiro/a	QP111	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Médico <sup>1</sup>	QP112	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3
Sala emergência hospital	QP113	Nenhuma o-0	1-3 vezes o-1	4-12 vezes o-2	mais de 12 vezes o-3

5. NOS ÚLTIMOS 12 MESES QUANTAS VEZES SEU FILHO/A INTERNOU NO HOSPITAL POR CRISE DE CHIADO OU PIADO NO PEITO OU ASMA?

- Nenhuma ..... o-0
- 1 vez ..... o-1
- vezes.... ..... o-2
- Mais de 2..... o-3

QP114

6. SEU FILHO/A CONSULTOU COM ALGUM DOS SEGUINTE PROFISSIONAIS, POR CHIADO OU PIADO NO PEITO OU ASMA, NOS ÚLTIMOS 12 MESES?

• Homeopata..... Sim o-1 ..... Não o-0

QP115

• Fisioterapeuta..... Sim o-1 ..... Não o-0

QP116

• Psiquiatra ou psicólogo..... Sim o-1 ..... Não o-0

QP117

• Assistente social..... Sim o-1 ..... Não o-0

QP118

• Curandeiro..... Sim o-1 ..... Não o-0

QP119

• Outros: especificar: \_\_\_\_\_

QP120

7. SEU FILHO/A ALGUMA VEZ TOMOU UMA INJEÇÃO DE VACINA PARA A ALERGIA, A FIM DE PREVENIR OU TRATAR A ASMA?

- Sim o-1
- Não o-0

QP121

<sup>1</sup> Refere-se a atendimento regular em ambulatório e pela crise marcou-se uma consulta extra.

<sup>2</sup> Refere-se atendimento de urgência ou consulta não marcada

**ATENÇÃO:** faça a pergunta abaixo APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS RESPIRATÓRIOS” (página 5).

Fez as perguntas abaixo? • Sim o-1

QP122

• Não (quer dizer pulou as perguntas abaixo) o-0

8. NOS ÚLTIMOS 12 MESES QUANTOS DIAS DE COLÉGIO OU CRECHE SEU FILHO/A PERDEU POR CHIADO OU PIADO NO PEITO OU ASMA?

• Nenhum.....o-0

• 1 a 5..... o-1

• 6 a 10..... o-2

• Mais de 10.. o-3

• Não freqüentava creche ou colégio nos 12 meses.....o-4

QP123

## Tratamentos para o nariz

**ATENÇÃO:** faça as perguntas abaixo APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS DE NARIZ” (página 6).

Fez as perguntas abaixo: • Sim o-1

QP124

• Não (quer dizer pulou as perguntas abaixo) o-0

1. NOS ÚLTIMOS 12 MESES SEU FILHO/A UTILIZOU ALGUM REMÉDIO (COMPRIMIDOS, SPRAYS NASAIS, OU OUTRA MEDICAÇÃO) PARA PROBLEMAS DE ALERGIA NASAL?

- Sim o-1
- Não o-0

QP125

**SE RESPONDEU "SIM", ESCREVA O NOME DOS REMÉDIOS; SE NÃO DEIXE EM BRANCO":**  
**(ATENÇÃO DIZER QUE "regularmente" significa todos os dias, durante ao menos dois meses ao ano)**

REMÉDIOS DE FARMÁCIA COM QUE FREQUÊNCIA? *marque com um círculo uma ou ambas opções*  
(dizer “remédios comprados em farmácias, igrejas, catálogos”, e ver receita ou caixa para anotar nome)

QP126

Quando estava irritado o-1 regularmente o-2

QP127

QP128

Quando estava irritado o-1 regularmente o-2

QP129

QP130

Quando estava irritado o-1 regularmente o-2

QP131

QP132

Quando estava irritado o-1 regularmente o-2

QP133

REMÉDIOS CASEIROS (dizer “remédios feitos em casa”)

QP134

Quando estava irritado o-1 regularmente o-2

QP135

QP136

Quando estava irritado o-1 regularmente o-2

QP137

QP138

Quando estava irritado o-1 regularmente o-2

QP139

2. NOS ÚLTIMOS 12 MESES SEU FILHO/A FOI A QUALQUER DOS SEGUINTE PROFISSIONAIS DE SAÚDE POR PROBLEMAS DE NARIZ OU DE RINITE,?

• Farmacêutico... QP140 ..... Sim o-1 Não o-0

• Agente de saúde... QP141 ..... Sim o-1 Não o-0

• Enfermeiro/a... QP142 ..... Sim o-1 Não o-0

• Pediatra ou Médico do Posto... QP143 Sim o-1 Não o-0

• Especialista<sup>3</sup>... QP144 ..... Sim o-1 Não o-0

• Sala de Emergência hospital... QP145 Sim o-1 Não o-0

Outros: especificar: \_\_\_\_\_

QP146

3. NOS ÚLTIMOS 12 MESES SEU FILHO/A TOMOU ALGUMA INJEÇÃO DE VACINA PARA ALERGIA, A FIM DE PREVENIR OU TRATAR OS PROBLEMAS DE NARIZ OU DE RINITE?

- Sim o-1
- Não o-0

QP147

4. NOS ÚLTIMOS 12 MESES SEU FILHO CONSULTOU COM HOMEOPATA, FOI A CURANDEIRO OU A PROFISSIONAIS DE MEDICINAS ALTERNATIVAS, PARA TRATAR OS PROBLEMAS DE NARIZ OU RINITE?

- Sim o-1
- Não o-0

QP148

**ATENÇÃO:** faça a pergunta abaixo APENAS se a pessoa respondeu “sim” a qualquer pergunta do quadro “PROBLEMAS DE NARIZ ” (página 6).

Fez as perguntas abaixo: • Sim o-1

QP149

• Não (quer dizer pulou as perguntas abaixo) o-0

5. NOS ÚLTIMOS 12 MESES QUANTOS DIAS COMPLETOS OU EM PARTE DE COLÉGIO OU CRECHE SEU FILHO/A PERDEU POR PROBLEMAS DE NARIZ OU RINITE?

• Nenhum..... o-0

• 1 a 5..... o-1

• 6 a 10..... o-2

• Mais de 10.. o-3

• Não freqüentava creche ou colégio nos 12 meses.....o-4

QP150

## Tratamentos para a pele

**ATENÇÃO:** faça a pergunta abaixo APENAS se a pessoa respondeu "sim" a qualquer pergunta do quadro "PROBLEMAS DE PELE".

Fez as perguntas abaixo? • Sim o-1

Qp151

• Não (quer dizer pulou as perguntas abaixo) o-0

1. NOS ÚLTIMOS 12 MESES SEU FILHO/A UTILIZOU ALGUM REMÉDIO (POMADA, CREME, COMPRIMIDOS) PARA IRRITAÇÕES NA PELE QUE COÇAVAM, OU POR CAUSA DE ECZEMA?

- Sim o-1 (**SE RESPONDEU "SIM", ESCREVA O NOME DOS REMÉDIOS**)
- Não o-0

QP152

**(ATENÇÃO DIZER QUE "regularmente" significa todos os dias, durante ao menos dois meses ao ano)**  
**Remédios de farmácia (dizer "remédios comprados em farmácias, igrejas, catálogos", ver caixa, e anotar se "não soube")** **Com que frequência?** \_\_\_\_\_  
 (marque com um círculo uma ou ambas opções)

QP153

Quando coçava o-0 regularmente o-1 QP154

QP155

Quando coçava o-0 regularmente o-1 QP156

QP157

Quando coçava o-0 regularmente o-1 QP158

QP159

Quando coçava o-0 regularmente o-1 QP160

**Remédio caseiros (dizer "feitos em casa")**

QP161

Quando coçava o-0 regularmente o-1 QP162

QP163

Quando coçava o-0 regularmente o-1 QP164

2. NOS ÚLTIMOS 12 MESES SEU FILHO/A FOI A QUALQUER DOS SEGUINTE PROFISSIONAIS DE SAÚDE PELAS IRRITAÇÕES NA PELE QUE COÇAVAM OU POR ECZEMA?

- Farmacêutico... QP165 ..... Sim o-1 Não o-0
- Agente de saúde.. QP166 ..... Sim o-1 Não o-0
- Enfermeiro/a..... QP167 ..... Sim o-1 Não o-0
- Pediatra ou Médico do Posto.. QP168 Sim o-1 Não o-0
- Especialista<sup>4</sup>..... QP169 ..... Sim o-1 Não o-0
- Sala de Emergência hospital.. QP170 Sim o-1 Não o-0
- Outros: especificar: QP171 \_\_\_\_\_

3. NOS ÚLTIMOS 12 MESES SEU FILHO/A INTERNOU EM HOSPITAL PELAS IRRITAÇÕES NA PELE QUE COÇAVAM?

- Sim o-1
- Não o-0

QP172

4. NOS ÚLTIMOS 12 MESES QUANTOS DIAS DE COLÉGIO SEU FILHO/A PERDEU PELAS IRRITAÇÕES NA PELE QUE COÇAVAM OU POR ECZEMA?

- Nenhum ..... o-0
- 1 a 5 ..... o-1
- 6 a 10 ..... o-2
- Mais de 10 .... o-3
- Não freqüentava creche ou colégio nos 12 meses.....o-4

QP173

## Os primeiros dias

1.	SEU FILHO(A) TEM ALGUM IRMÃO/Ã GÊMEO? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP174
2.	SEU FILHO(A) FOI AMAMENTADO NO PEITO? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP175
3.	EM CASO AFIRMATIVO, DURANTE QUANTO TEMPO? <ul style="list-style-type: none"> <li>• Menos de 6 meses o-0</li> <li>• 6-12 meses o-1</li> <li>• mais de 1 ano o-3</li> </ul>	QP176
4.	<b>EM CASO AFIRMATIVO</b> , DURANTE QUANTO TEMPO FOI ALIMENTADO SÓ NO PEITO, SEM ACRESCENTAR OUTROS ALIMENTOS, SUCOS, CHÁS OU ÁGUA? <ul style="list-style-type: none"> <li>• menos 2 meses o-0</li> <li>• 2-4 meses..... o-1</li> <li>• 5-6 meses ..... o-2</li> <li>• mais de 6 meses o-3</li> </ul>	QP177
5.	SEU FILHO(A) TEM IRMÃOS/IRMÃS <u>MAIS VELHOS</u> ? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP178
6.	<b>EM CASO AFIRMATIVO</b> , QUANTOS IRMÃOS <u>MAIS VELHOS</u> (MENINOS)? ____	QP179
	QUANTAS IRMÃS <u>MAIS VELHAS</u> (MENINAS)? ____	QP180
7.	SEU FILHO TEM IRMÃOS/IRMÃS <u>MAIS NOVOS</u> ? <ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	QP181
8.	<b>EM CASO AFIRMATIVO</b> , QUANTOS IRMÃOS <u>MAIS NOVOS</u> (MENINOS)? ____	QP182
	QUANTAS IRMÃS <u>MAIS NOVAS</u> (MENINAS)? ____	QP183

9. SEU FILHO(A) FREQUENTA OU FREQUENTOU CRECHE ALGUMA VEZ NA VIDA?

- Sim o-1  QP184
- Não o-0

EM CASO AFIRMATIVO, DESDE QUE IDADE? \_\_\_\_\_  QP185

E DURANTE QUANTO TEMPO  QP186 ANOS/MESES (*passar um círculo em anos ou meses*)  QP187

10. SEU FILHO(A) FREQUENTA OU FREQUENTOU JARDIM DE INFÂNCIA ALGUMA VEZ NA VIDA?

- Sim o-1  QP188
- Não o-0

EM CASO AFIRMATIVO, DESDE QUE IDADE? \_\_\_\_\_  QP189 ANOS/MESES  QP190

E DURANTE QUANTO TEMPO  QP191 ANOS/MESES (*passar um círculo em anos ou meses*)  QP192

11. SEU FILHO(A) JÁ FREQUENTA A ESCOLA?

- Sim o-1  QP193
- Não o-0

12. SEU FILHO(A) NASCEU COM QUANTOS MESES DE GRAVIDEZ? \_\_\_\_\_  QP194

13. SEU FILHO(A) NASCEU DE PARTO NORMAL OU CESARIANA?

- Parto normal o-1
- Cesariana o-0  QP195

## Doenças e vacinas

1. A **MÃE** DA CRIANÇA TEVE ALGUMA DAS SEGUINTE DOENÇAS?  
(*Marcar todas as opções corretas*)

- Asma... **QP196** ..... o Sim-1 Não o-0
- Alergia de nariz ou rinite alérgica.. **QP197** o Sim-1 Não o-0
- Eczema atópico ou dermatite alérgica. **QP198** Sim-1 Não o-0

2. O **PAI** DA CRIANÇA TEVE ALGUMA DAS SEGUINTE DOENÇAS?  
(*Marcar todas as opções corretas*)

- Asma... **QP199** ..... o Sim-1 Não o-0
- Alergia de nariz ou rinite alérgica. **QP200** o Sim-1 Não o-0
- Eczema atópico ou dermatite alérgica **QP201** o Sim-1 Não o-0

3. SEU FILHO/A FOI VACINADO PARA QUALQUER DAS SEGUINTE DOENÇAS?

ANTES DE RESPONDER AS PERGUNTAS ABAIXO O INFORMANTE VIU O CARTÃO DE VACINAÇÃO?

- Sim o-1
- Não o-0

**QP202**

(*Marcar todas as opções corretas*)

**Sarampo** só ou combinado com Rubéola e Caxumba

- Sim o 1... **QP203** • Em caso afirmativo, com que idade? **QP204** meses ▶ Não sabe idade-o 2 **QP205**
- Não o 0

**Tuberculose/BCG**

- Sim o 1. **QP206** • Em caso afirmativo, com que idade? **QP207** meses ▶ Não sabe idade-o 2 **QP208**
- Não o 0

(*menos de 1 mês (dias) assinalar "1"*)

4. SEU FILHO/A TEVE ALGUMA DAS SEGUINTE DOENÇAS?

(*Marcar todas as opções corretas*)

**Sarampo**

- Sim o 1. **QP209** • Em caso afirmativo, com que idade? **QP210** meses ▶ Não sabe idade-o 2
- Não o 0 **QP211**

**Coqueluche**

- Sim o 1. **QP212** • Em caso afirmativo, com que idade? **QP213** meses ▶ Não sabe idade-o 2
- Não o 0 **QP214**

**Tuberculose**

- Sim o 1. **QP215** • Em caso afirmativo, com que idade? **QP216** meses ▶ Não sabe idade-o 2
- Não o 0 **QP217**

( ) cartão não encontrado

QP218

***Anotar vacinações mesmo que estejam registradas em mais de um cartão***

Vacina		Número de doses	Data última dose para cada vacina
QP219		QP220	QP221
QP222		QP223	QP224
QP225		QP226	QP227
QP228		QP229	QP230
QP231		QP232	QP233
QP234		QP235	QP236
QP237		QP238	QP239
QP240		QP241	QP242
QP243		QP244	QP245
QP246		QP247	QP248
QP249		QP250	QP251



7. **INSPEÇÃO** (*Aqui você deve olhar nas paredes e teto da casa*)

- Sim o-1

QP269

- Não o-0

Se sim, em que lugar na casa:

QP270

Lesões típicas

**Mofo em parede:** manchas circulares, negras ou verdes, mancha não homogênea (como tinta salpicada), em geral com halo claro no centro, mancha mais clara em volta, com ou sem relevo, tinta estufada

**Mofo em madeira:** manchas verdes ou negras ou cinzas, com relevo, esfarela quando tocado, madeira estufada

8. QUE TIPO DE FOGÃO SE UTILIZA OU UTILIZAVA PARA COZINHAR?

(*repetir pergunta para cada momento abaixo, exemplo, “teve durante primeiro ano de vida?”*)

	<u>Atualmente/agora</u>		<u>Durante o primeiro ano de vida</u>	
• Elétrico	QP271	Sim o-1 Não o-0	QP272	Sim o-1 Não o-0
• Gás	QP273	Sim o-1 Não o-0	QP274	Sim o-1 Não o-0
• Carvão ou lenha	QP275	Sim o-1 Não o-0	QP276	Sim o-1 Não o-0
• Outros	QP277	Sim o-1 Não o-0	QP278	Sim o-1 Não o-0

**ATENÇÃO:** faça a pergunta **9** abaixo **APENAS** se a pessoa respondeu “sim” a qualquer pergunta DOS QUADROS “PROBLEMAS RESPIRATÓRIOS” (página 5), “PROBLEMAS DE NARIZ” (página 6) OU “PROBLEMAS DE PELE” (página 7).

QP279

Fez as perguntas abaixo: • Sim o-1

• Não (quer dizer pulou as perguntas abaixo) o-0

9. VOCÊ FEZ ALGUMA MUDANÇA NA CASA POR CAUSA DA ASMA DE SEU FILHO(A) OU PELOS PROBLEMAS ALÉRGICOS?

(Marcar as opções que parecerem corretas)

- Se desfez de animais?..... QP280 ..... QP281 Sim o-1 Não o-0 Nunca teve animais o-2  
Em caso afirmativo, que idade tinha o menino/a? \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)
- A pessoa que fuma, parou de fumar ou diminuiu o consumo de cigarro? QP282 Sim o-1 Não o-0 Nunca fumou o-2  
Em caso afirmativo, que idade tinha o menino/a? QP283 \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)
- Trocou tipo de travesseiros? QP284 ..... QP285 ..... Sim o-1 Não o-0  
Em caso afirmativo, que idade tinha o menino/a? \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)
- Trocou tipo de roupa de cama? QP286 ..... QP287 ..... Sim o-1 Não o-0  
Em caso afirmativo, que idade tinha o menino/a? \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)
- Trocou o piso da casa? QP288 ..... QP289 ..... Sim o-1 Não o-0  
Em caso afirmativo, que idade tinha o menino/a? \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)
- Outras mudanças? QP290 ..... QP291 ..... Sim o-1 Não o-0  
Em caso afirmativo, que idade tinha o menino/a? \_\_\_\_\_ anos (menor de 1 ano escreva o número 0 - zero)  
Quais mudanças? \_\_\_\_\_ QP292

10- QUE TIPO DE ROUPA DE CAMA O/A MENINO/A USA OU USAVA?

(repetir pergunta para cada momento, por exemplo, “tem edredon atualmente...?”)

(Marcar uma ou mais das opções corretas)

- Edredon ou colcha ou lençol sintético\* QP293 Atualmente/agora- o Durante o primeiro ano de vida-o QP294
  - Cobertor . QP295 ..... Atualmente/agora- o Durante o primeiro ano de vida-o QP296
  - Outros materiais QP297 ..... Atualmente/agora- o Durante o primeiro ano de vida-o QP298
- Quais? QP299 \_\_\_\_\_

\* Pode explicar como “não é feito de algodão, é como tecido de nylon, de tergal, etc”

## Alimentos

**ATUALMENTE**, COM QUE FREQUÊNCIA MÉDIA SEU FILHO/A COME OU BEBE:

(*citar cada item separadamente e marcar a frequência apropriada a resposta dada*)

Alimento	Nunca ou ocasionalmente	1 ou 2 vezes por semana	3 ou mais vezes por semana	
Carne (galinha, carne de boi, outras carnes)	o-0	o-1	o- 2	<b>QP300</b>
Peixe	o-0	o-1	o- 2	<b>QP301</b>
Frutas	o-0	o-1	o- 2	<b>QP302</b>
Verduras ou legumes frescas	o-0	o-1	o- 2	<b>QP303</b>
Suco de frutas	o-0	o-1	o- 2	<b>QP304</b>
Refrigerantes	o-0	o-1	o- 2	<b>QP305</b>
Manteiga	o-0	o-1	o- 2	<b>QP306</b>
Margarina	o-0	o-1	o- 2	<b>QP307</b>
Leite	o-0	o-1	o- 2	<b>QP308</b>
Ovos	o-0	o-1	o- 2	<b>QP309</b>
Hamburger	o-0	o-1	o- 2	<b>QP310</b>
Comida com azeite de dendê?	o-0	o-1	o- 2	<b>QP311</b>
Acarajé ou abará ou vatapá?	o-0	o-1	o- 2	<b>QP312</b>

A MAIOR PARTE DO TEMPO A CRIANÇA NÃO ESTÁ EM CASA, INFORMAÇÃO REFERE-SE AO QUE O ENTREVISTADO VÊ/CONHECE:

- Sim o-1
- Não o-0

<b>QP313</b>
--------------

**INSETOS**

**INSETOS. QUAL O INSETO OU ANIMAL QUE MAIS TEM APARECIDO EM SUA CASA ?**  
**(dizer “vou citar cada inseto e você me diz se tem”)**

1 – Ratos .....	Sim o-1	Não ou raro o-0	<b>QP314</b>
2 - Baratas			
das pequenas .....	Sim o-1	Não ou raro o-0	<b>QP315</b>
das grandes .....	Sim o-1	Não ou raro o-0	<b>QP316</b>
3 – Moscas .....	Sim o-1	Não ou raro o-0	<b>QP317</b>
4 – Mosquitos/muriçocas .....	Sim o-1	Não ou raro o-0	<b>QP318</b>
5 – Maruin ou Muruin	Sim o-1	Não ou raro o-0	<b>QP319</b>
6 – Outro. Qual? _____			<b>QP320</b>
7 - Nenhum deles .....	o-1		<b>QP321</b>

SE UTILIZA ATUALMENTE/AGORA INSETICIDAS EM CASA PARA MATAR INSETOS? **QP322**

    Sim o-1    Não o-0

SE SIM, QUAL? \_\_\_\_\_ **QP323** \_\_\_\_\_

## Outras perguntas sobre saúde

1. QUANTAS VEZES A SENHORA ESTEVE GRÁVIDA, INCLUINDO PERDAS?	Resposta: _____	<b>QP324</b>
2. NOS ÚLTIMOS 12 MESES A CRIANÇA FOI INTERNADO(A)?	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	<b>QP325</b>
3. POR QUE? RESPOSTA: _____		<b>QP326</b>
4. NOS ÚLTIMOS DOZE MESES SEU FILHO(A) TEVE BRONQUITE?	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> <li>• Não sabe o-2</li> </ul>	<b>QP327</b>
5. DE ZERO ATÉ 3 ANOS DE IDADE, ALGUM MÉDICO(A) DISSE QUE SEU FILHO(A) TEVE BRONQUIOLITE?	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> <li>• Não sabe o-2</li> </ul>	<b>QP328</b>
<i>Se respondeu Sim na pergunta acima, fazer a pergunta abaixo:</i>		
6. DE ZERO ATÉ 3 ANOS DE IDADE, SEU FILHO/A INTERNOU EM HOSPITAL OU FOI ATENDIDO EM SALA DE EMERGÊNCIA POR BRONQUIOLITE?	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> <li>• Não feito a pergunta o-9</li> </ul>	<b>QP329</b>
7. DE ZERO ATÉ 3 ANOS DE IDADE, SEU FILHO(A) TEVE PNEUMONIA OU BRONCOPNEUMONIA?	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> </ul>	<b>QP330</b>
<i>Se respondeu Sim na pergunta acima, fazer a pergunta abaixo:</i>		
8. DE ZERO ATÉ 3 ANOS DE IDADE, SEU FILHO/A INTERNOU EM HOSPITAL OU FOI ATENDIDO EM SALA DE EMERGÊNCIA POR PNEUMONIA OU BRONCOPNEUMONIA??	<ul style="list-style-type: none"> <li>• Sim o-1</li> <li>• Não o-0</li> <li>• Não feito a pergunta o-9</li> </ul>	<b>QP331</b>
9. TEM NOS ÚLTIMOS 12 MESES A CRIANÇA TIDO ALGUMA DAS DOENÇAS DE PELE?		
<i>(Pode marcar mais de uma opção, ou apenas “não refere problema de pele nos últimos 12 meses”)</i>		
	<ul style="list-style-type: none"> <li>• Eczema o-0</li> <li>• Verruga o-1</li> <li>• Escabiose o-2</li> <li>• Psoríase o-3</li> <li>• Outro o-4</li> <li>• Qual? _____</li> </ul>	<b>QP332</b> <b>QP333</b> <b>QP334</b> <b>QP335</b> <b>QP336</b> <b>QP337</b>
	<ul style="list-style-type: none"> <li>• Não refere problema de pele nos últimos 12 meses o-5</li> </ul>	<b>QP338</b>

## Questionário sobre características do leito e local de dormir

1. NA CASA TEM COLCHÃO FORRADO DE PLÁSTICO OU COM CAPA COM MATERIAL ANTI-ALÉRGICO?

- Sim o-1
- Não o-0

QP339

2. NA CASA TEM TRAVESSEIRO FORRADO DE PLÁSTICO OU COM CAPA COM MATERIAL ANTI-ALÉRGICO?

- Sim o-1
- Não o-0

QP340

3. ONDE A CRIANÇA (DA PESQUISA) DORME?

- Cama com colchão o-0
- Colchão sem cama o-1
- Outro ..... o-2: \_\_\_\_\_

QP341

QP342

4. COM QUE FREQUÊNCIA TROCA A ROUPA DE CAMA(OU DE ONDE A CRIANÇA DORME) (*criança da pesquisa*)?

- Mais de uma vez por semana.... o-0
- Toda semana..... o-1
- 15 em 15 dias..... o-2
- Mês em mês..... o-3
- Não se aplica/outro/não sabe.... o-9

QP343

5. COM QUE FREQUÊNCIA REALIZA ESTES CUIDADOS COM O LUGAR ONDE A CRIANÇA DORME (*criança da pesquisa*)?

(*pode marcar mais de uma opção*)

### Frequência

	Todos os dias	Toda semana	Todo mês	De 6/6 meses	1 vez/ano	Não se Aplica/outro/não sabe	
4.1 Põe ao sol.....	o 0	o 1	o 2	o 3	o 4	o 9	QP344
4.2 "Bate" a poeira.....	o 0	o 1	o 2	o 3	o 4	o 9	QP345
4.3 Aspira o pó.....	o 0	o 1	o 2	o 3	o 4	o 9	QP346

6. COM QUE FREQUÊNCIA REALIZA ESTES CUIDADOS COM O (OS) TRAVESSEIRO(S) DO LUGAR ONDE A CRIANÇA DORME (*criança da pesquisa*)?

(*pode marcar mais de uma opção*)

	Frequência						
	Todos os dias	Toda semana	Todo mês	De 6/6 meses	1 vez/ano	Não se Aplica/outro/não sabe	
5.1 Põe ao sol.....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP347
5.2 “Bate” a poeira	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP348
5.3 Aspira o pó.....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP349
5.4 Lava.....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP350

7. COM QUE FREQUÊNCIA REALIZA ESTES CUIDADOS COM O LOCAL (QUARTO) ONDE DORME A CRIANÇA (*da pesquisa*)?

(*pode marcar mais de uma opção*)

CUIDADO COM O LOCAL ONDE DORME	Frequência						
	Todos os dias	Toda semana	Todo mês	De 6/6 meses	Não / Nunca	Não se Aplica/outro/não sabe	
6.1 Varre .....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP351
6.2 Passa pano úmido no chão.....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP352
6.3 Passa pano úmido na parede...	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP353
6.4 Aspira o pó.....	<input type="radio"/> 0	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 9	QP354

8. OS CUIDADOS COM O LOCAL (QUARTO) ONDE DORME A CRIANÇA FORAM ADOTADOS PARA (*marcar apenas uma opção*) (*Se respondeu SIM a uma das opções da pergunta 7 acima*)

• Limpeza de rotina em todos os quartos.....	<input type="radio"/> 0	QP355
• Apenas para os leitos de pessoas alérgicas	<input type="radio"/> 1	QP356
• Outro .....	<input type="radio"/> 2 (especificar)	QP357
• As respostas 0,1,2 acima foram negativas	<input type="radio"/> 9	QP359