



MODELOS BASEADOS EM REDES NEURAIAS ARTIFICIAIS PARA O  
DIAGNÓSTICO EM TRIAGEM DE TUBERCULOSE RESISTENTE E  
MULTIRRESISTENTE NO BRASIL

Luiz Henrique Ramos de Azevedo Évora

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientadores: José Manoel de Seixas

Afranio Lineu Kritski

Rio de Janeiro

Março de 2018

MODELOS BASEADOS EM REDES NEURAIAS ARTIFICIAIS PARA O  
DIAGNÓSTICO EM TRIAGEM DE TUBERCULOSE RESISTENTE E  
MULTIRRESISTENTE NO BRASIL

Luiz Henrique Ramos de Azevedo Évora

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR  
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

---

Prof. José Manoel de Seixas, D.Sc.

---

Prof. Afranio Lineu Kritski, D.Sc.

---

Prof. Carlos Dias Maciel, D.Sc.

---

Prof. Guilherme Loureiro Werneck, D.Sc.

---

Profa. Fernanda Carvalho de Queiroz Mello, D.Sc.

---

Prof. Jurandir Nadal, D.Sc.

---

Prof. Luiz Pereira Caloba, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2018

Évora, Luiz Henrique Ramos de Azevedo

Modelos Baseados em Redes Neurais Artificiais para o Diagnóstico em Triagem de Tuberculose Resistente e Multirresistente no Brasil/Luiz Henrique Ramos de Azevedo Évora. – Rio de Janeiro: UFRJ/COPPE, 2018.

XIV, 99 p.: il.; 29,7cm.

Orientadores: José Manoel de Seixas

Afranio Lineu Kritski

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2018.

Referências Bibliográficas: p. 85 – 95.

1. Tuberculose. I. Seixas, José Manoel de *et al.*
- II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELOS BASEADOS EM REDES NEURAIAS ARTIFICIAIS PARA O  
DIAGNÓSTICO EM TRIAGEM DE TUBERCULOSE RESISTENTE E  
MULTIRRESISTENTE NO BRASIL

Luiz Henrique Ramos de Azevedo Évora

Março/2018

Orientadores: José Manoel de Seixas

Afranio Lineu Kritski

Programa: Engenharia Elétrica

A tuberculose (TB) resistente a drogas (TB-DR) implica em tratamentos mais complexos, acarretando em taxas de mortalidade e morbidade mais elevadas. Neste contexto, propomos o uso de testes de triagem que identifiquem precocemente os pacientes com maior probabilidade de TB-DR, para serem priorizados. Modelos baseados em Redes Neurais Artificiais e Árvores de Classificação e Regressão foram desenvolvidos, incluindo algoritmos de *boosting*. Consideram-se sintomas e variáveis sócio demográficas e clínicas dos pacientes. Foram produzidos escores a nível estadual e os resultados comparados aos de uma abordagem nacional. Modelos de diferentes complexidades foram desenvolvidos para se adequarem aos recursos disponíveis nos locais de aplicação, sendo guiados pela relevância das variáveis e a qualidade dos dados. Observou-se que os modelos estaduais obtiveram uma sensibilidade, em média, maior que 85% na triagem de pacientes para o diagnóstico de TB-DR, contra 82,7% da abordagem nacional, indicando que os escores clínicos locais tendem a capturar melhor as desigualdades operacionais do sistema de saúde.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

NEURAL NETWORK MODELS FOR SCREENING DRUG AND MULTIDRUG  
RESISTANT TUBERCULOSIS IN BRAZIL

Luiz Henrique Ramos de Azevedo Évora

March/2018

Advisors: José Manoel de Seixas

Afranio Lineu Kritski

Department: Electrical Engineering

Drug-Resistant TB (DR-TB) implicates in more complex treatments and leads to higher deceased and morbidity numbers. In this context, we propose the use of screening tests that early identifies patients with higher probability of having DR-TB and prioritize them. Artificial Neural Networks and Classification And Regression Tree models are generated, and a boosting algorithm is applied, considering as input the patient's symptoms and social-demographic variables. Specific scores by each State are produced and the results are compared to a national-wide approach. Models with different complexity levels were developed in order to fit the available resources in each site, being guided by variable relevance and data quality. Models developed by each State achieved an average sensitivity higher than 85% when screening RJ patients considering DR-TB from non DR-TB, against 82.7% using the national approach, indicating that the local clinical scores can better capture operational differences present in the health system.

# Sumário

<b>Lista de Figuras</b>	<b>viii</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	3
1.2 Objetivos . . . . .	6
1.3 Organização do documento . . . . .	7
<b>2 Tuberculose</b>	<b>8</b>
2.1 Tuberculose pulmonar . . . . .	8
2.1.1 Incidência, métodos diagnósticos e tratamento . . . . .	8
2.1.2 Resistência aos medicamentos em TB . . . . .	12
2.1.3 Tuberculose no Brasil . . . . .	14
2.1.4 Modelos Estatísticos . . . . .	18
<b>3 Método</b>	<b>27</b>
3.1 Pré-processamento dos dados . . . . .	27
3.2 Desenvolvimento dos Modelos Computacionais . . . . .	30
3.2.1 Redes Neurais Artificiais . . . . .	30
3.2.1.1 Treinamento do modelo . . . . .	31
3.2.2 Árvores de Classificação . . . . .	33
3.2.2.1 Treinamento do modelo . . . . .	34
3.2.2.2 O <i>boosting</i> e o algoritmo AdaBoost . . . . .	35
3.3 Avaliação do desempenho . . . . .	37

3.4	Análise de relevância . . . . .	42
3.4.1	O método <i>Weights</i> . . . . .	43
3.4.2	O método <i>Relevance mapping</i> . . . . .	43
3.4.3	Índice de relevância consolidado . . . . .	44
3.5	Geração de modelos de menor complexidade . . . . .	44
<b>4</b>	<b>Resultados e Discussão</b>	<b>46</b>
4.1	Base de Dados . . . . .	46
4.2	Análise dos modelos computacionais . . . . .	55
4.3	Análise de relevância das variáveis . . . . .	69
4.4	Estudo de prevalência . . . . .	78
<b>5</b>	<b>Conclusões</b>	<b>81</b>
	<b>Referências Bibliográficas</b>	<b>85</b>
<b>A</b>	<b>Publicações Relativas à Tese</b>	<b>96</b>

# Lista de Figuras

2.1	Mapa que representa visualmente a incidência de TB por país no ano de 2016. Fonte: OMS. Extraído de [25] . . . . .	11
2.2	Mapa que indica os países que estão utilizando o regime terapêutico de curta duração para tratamento de TB-MDR. Fonte: OMS. Extraído de [25] . . . . .	14
2.3	Desfechos operacionais e índices de TB no Brasil por unidade federada, no ano de 2016. Fonte: Ministério da Saúde. Dados extraídos de [12] . . . . .	15
2.4	Evolução temporal do número de casos novos de TB-DR e TB-MDR no Brasil, do ano de 2001 a 2015. Dados extraídos de [39] . . . . .	17
2.5	Mortalidade ocasionada por TB-DR e TB-MDR no ano de 2014, por unidade federativa. Dados extraídos de [40] . . . . .	18
3.1	Ilustração da técnica de validação cruzada pelo método k-fold. . . . .	29
3.2	Ilustração de uma rede neural multicamadas (MLP), mostrando como a informação é processada, assim como a estrutura de um neurônio artificial. . . . .	31
3.3	Representação da evolução do erro de classificação de um treinamento por iteração e escolha do ponto de parada com a melhor generalização. . . . .	33
3.4	Ilustração de uma árvore de classificação, mostrando sua estrutura (raiz, nó, folha e ramos) e o fluxo da informação processada. . . . .	34



3.5	Arquitetura de <i>ensemble</i> e <i>boosting</i> adotado pelo algoritmo AdaBoost. Nele, as amostras que geram mais dificuldade de aprendizado recebem pesos maiores na geração do modelo CART, o qual contém apenas um nó de decisão (modelos de fraco aprendizado). Os resultados de cada modelo são então ponderados e somados para gerarem a classificação final. . . . .	38
3.6	Distribuição de pacientes com diagnósticos positivos e negativos e exemplo de ponto de corte para definição da sensibilidade e especificidade. Regiões de verdadeiros negativos (VN), falsos negativos (FN), verdadeiros positivos (VP) e falsos positivos (FP) são indicadas. . . .	39
3.7	Exemplo de curva ROC para medir a relação entre sensibilidade e falsos positivos, indicando o ponto que representa o índice SP ótimo. .	40
3.8	Exemplo de construção do índice de relevância consolidado a partir de dois métodos para o cálculo de relevância. . . . .	44
4.1	Diagnósticos de TB considerando a complexidade de tratamento. O estudo desenvolve escores clínicos para a classificação dos diagnósticos de TB-DR/MDR dentre todos os outros pacientes e de TB-DS, além de TB-MDR e TB-DR . . . . .	48
4.2	Índice SP obtido pela RNA, a partir da validação cruzada utilizando até 10 neurônios na camada escondida da rede, para o suporte a discriminação entre TB-DR/MDR e não TB-DR/MDR com o modelo RJ. . . . .	61
4.3	Curva de aprendizado da RNA, em termos de erro quadrático médio, assim como a indicação da época referente a parada do treinamento da rede, considerando o suporte ao diagnóstico entre TB-DR/MDR e não TB-DR/MDR no modelo RJ. . . . .	61
4.4	Histograma com os valores de saída da rede neural que apresentou maior índice SP, na discriminação dos diagnósticos de TB-DR/MDR e não TB-DR/MDR, para o modelo RJ. . . . .	62

4.5	A curva ROC para as RNA desenvolvidas para o estado do RJ e pela abordagem nacional, quando propagando todos os dados. Aqui é considerado o diagnóstico para discriminar TB-DR/MDR de não TB-DR/MDR. Os pontos de corte que representam o maior valor do índice SP são indicados, assim como os valores de sensibilidade, especificidade, índice SP e AUC. . . . .	62
4.6	Estrutura da árvore de classificação, gerada utilizando CART, para a discriminação entre TB-DR/MDR e não TB-DR/MDR, considerando o desenvolvimento do modelo do estado do RJ. . . . .	64
4.7	Evolução do desempenho por modelo CART de fraco aprendizado, para dados de treino e teste, utilizando o algoritmo AdaBoost. É considerado o diagnóstico entre TB-DR/MDR e não TB-DR/MDR para o estado do RJ. O peso de cada modelo para o resultados final também é mostrado. . . . .	65
4.8	Estrutura dos 10 modelos CART de fraco aprendizado de maior peso na decisão final obtidos pelo algoritmo AdaBoost. Os modelos são para o diagnóstico de TB-DR contra não TB-DR para o estado do RJ.	66
4.9	Classificação dos pacientes que foram excluídos do estudo por apresentarem tratamento há mais de 30 dias e diagnóstico de não TB. Para isso, é utilizado o modelo RNA para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR, e o resultado para cada abordagem é indicado. . . . .	68
4.10	Lista com todas as variáveis ordenadas pelo índice de relevância consolidado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR. Os resultados obtidos pelos métodos <i>Relevance mapping</i> e <i>Weights</i> são somados para se chegar ao resultado mostrado, sendo que cada respectiva contribuição é ainda indicada como referência. As variáveis selecionadas em cada abordagem (Tight, Medium e Loose) são ressaltadas, refletindo também os critérios de qualidade de dados utilizados. . . . .	71

4.11 Valores de VPP e VPN para triagem de TB-DR relativos a abordagem por estados, de acordo com respectiva prevalência e diferentes estratégias de desenvolvimento de modelos, sendo os resultados para: (a) RJ, (b) CE e (c) RS. A probabilidade de 95% é indicada, assim como as respectivas prevalências de TB-DR e TB-MDR, em relação a base de dados, para cada estado. . . . . 79

# Lista de Tabelas

2.1	Lista com os valores relativos a índices e desfechos operacionais de TB por unidade federativa, no ano de 2016. Fonte: Ministério da Saúde. Dados extraídos de [12] . . . . .	16
4.1	Número de pacientes incluídos no estudo separados por regiões (RJ, CE, RS, outras localizações e todas as regiões exceto RJ) e diagnósticos de TB (Não TB, TB-DS, além de TB-DR e TB-MDR). . . . .	49
4.2	Lista de todas variáveis, e suas respectivas estatísticas em termos de resposta, para o diagnóstico de TB-DR/MDR, considerando a abordagem nacional. Os dados são expressos como o total de pacientes (% do total) ou a média $\pm$ RMS (um valor de raiz média quadrática da variância). Possíveis problemas em qualidade de dados são indicados, como: (1) falta de infraestrutura e/ou problemas relacionados a acessibilidade da informação e (2) validade da resposta do paciente. . . . .	52
4.3	Lista de todas variáveis, e suas respectivas estatísticas em termos de resposta, para o diagnóstico de TB-DR/MDR, considerando pacientes do RJ. Os dados são expressos como o total de pacientes (% do total) ou a média $\pm$ RMS (um valor de raiz média quadrática da variância). Possíveis problemas em qualidade de dados são indicados, como: (1) falta de infraestrutura e/ou problemas relacionados a acessibilidade da informação e (2) validade da resposta do paciente. . . . .	53

4.4	Desempenho dos modelos de RNA e CART por região, obtidos através da validação cruzada, com informações sobre o número (#N) de neurônios na camada escondida ou profundidade da árvore de classificação, e ponto de corte (RNA) considerando o suporte ao diagnóstico entre TB-DR/MDR e não TB-DR/MDR; TB-DR/MDR e TB-DS; e TB-MDR e TB-DR. . . . .	58
4.5	Desempenho alcançado, por diagnóstico, ao se propagar os pacientes do estado de SP pelos modelos RNA desenvolvidos para os estados do RJ, CE e RS. . . . .	67
4.6	Desempenho alcançado, por diagnóstico, ao se desenvolver modelos RNA utilizando os dados dos pacientes do RJ, CE e RS, e excluindo dados de pacientes do estado de SP. . . . .	68
4.7	Análise da discriminação dos pacientes com o diagnóstico de não TB para avaliar as possíveis diferenças de comportamento do modelo com os que tinham menos de 30 dias de tratamento e os sem tratamento. São utilizados os modelos RNA de classificação entre TB-DR/MDR e não TB-DR/MDR para os estados do RJ, CE e RS. . . . .	69
4.8	Desempenho por estratégia de relevância e abordagens (estados e nacional) considerando o índice consolidado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR; TB-DR/MDR e TB-DS; e TB-MDR e TB-DR, utilizando os modelos RNA. O total de neurônios na camada escondida é representado pela coluna #N e o número de variáveis utilizado na coluna #V. . . . .	73
4.9	Lista de variáveis incluídas em cada estratégia de desenvolvimento considerando relevância ( <i>Tight</i> , <i>Medium</i> e <i>Loose</i> ), para as abordagens estaduais e nacional, incluindo quando se exclui os pacientes do RJ. A relevância utilizada é a resultante do índice agregado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR. . . . .	75

4.10 Tabela com os valores de VPP e VPN conforme o aumento da prevalência, considerando a abordagem por estados e cada estratégia de relevância adotada. Aqui os resultados são relativos ao diagnóstico de TB-DR e não TB-DR. . . . .	80
--	----

# Capítulo 1

## Introdução

O reconhecimento de padrões objetiva classificar objetos em um determinado número de categorias ou classes, realizando essa tarefa a partir da observação de suas características [1]. Para isso, a partição entre classes deve levar em consideração a relevância dessas características [2] e a qualidade dos dados a serem analisados [3]. Na área médica, reconhecer os padrões de uma doença a partir de formulários, os quais reúnem informações de pacientes, pode, por exemplo, acelerar o encaminhamento de um doente para o tratamento. No caso de doenças infecciosas, essa ação diminui a probabilidade da contaminação de mais pessoas em uma população.

O processo de tomada de decisão clínica é uma atividade presente no cotidiano da área médica. Independentemente da situação, busca-se tomar sempre a melhor decisão possível. Para isso, o profissional de saúde precisa avaliar um conjunto de informações e, por meio de diferentes critérios, chegar a uma conclusão. Nesse âmbito, é possível identificar diversas situações que são básicas e estão sempre presentes na área médica, como triagem, diagnóstico, tratamento, prognóstico e prevenção. Em cada uma dessas situações, para se obter um bom resultado é preciso realizar um estudo cuidadoso sobre as informações disponíveis para a identificação do problema.

Na etapa de triagem/diagnóstico, a equipe de saúde tem como objetivo a identificação de uma eventual doença ou transtorno de saúde. Os dados disponíveis podem ser provenientes dos sinais e sintomas apresentados, do histórico clínico, de exame físico e de exames complementares. Ainda, pode-se levar em consideração os

indicadores operacionais do sistema de saúde da região onde o paciente se encontra, e que podem influenciar nesse processo, assim como um entendimento prévio sobre as características de uma determinada população, por possibilitar uma interpretação direcionada sobre hábitos culturais que possam influenciar na incidência de uma doença. A partir do diagnóstico, é então realizado o planejamento para o tratamento, além do prognóstico.

Ao realizar a triagem de um paciente, para o tratamento de uma dada doença, é desejável que esse processo seja realizado com uma alta eficiência, gerando uma baixa percentagem de falsos alarmes. Dessa forma, evita-se que pacientes sem a doença sejam encaminhados para exames complementares, que podem muitas vezes serem invasivos, além de acarretarem em maiores custos, ou tratamentos desnecessários e seus possíveis efeitos colaterais. Por outro lado, deve-se também buscar ao máximo encaminhar corretamente um paciente com agravo de saúde para um tratamento que seja necessário, ou mesmo para o seu isolamento, nas unidades hospitalares, em caso de doenças altamente contagiosas. Caso contrário, o problema poderá acabar se alastrando em uma dada população por conta do contágio de outras pessoas pelo paciente triado de forma errônea.

É preciso também considerar que existem unidades de saúde com diferentes níveis de estrutura e disponibilidade de informações. Determinadas variáveis clínicas, que representam características consideradas de menor relevância para um diagnóstico, podem acabar assumindo um papel mais decisivo em detrimento de informações de difícil acesso. Dessa forma, os processos adotados em triagem devem ser capazes de se adaptar às características do ambiente no qual é aplicado. Além disso, a inclusão de certas informações no processo de decisão pode ser revista durante o contato entre o profissional de saúde e o paciente, caso perceba-se que dados fornecidos possam não ser válidos. Essa situação pode acontecer, por exemplo, por desconforto da pessoa em compartilhar alguma informação de caráter privado no momento da triagem e que poderia influenciar no risco associado a uma doença.

A habilidade de realizar um diagnóstico correto exige, não só conhecimento e experiência do profissional de saúde, mas também um elaborado raciocínio clínico.



Além disso, o profissional de saúde tem se defrontado com uma quantidade de informações referentes à área da saúde que dobram a cada três anos, refletindo os avanços que têm sido realizados nas áreas médica e correlatas, como novos métodos de diagnósticos e terapêutica, princípios químicos, inovações da área da biologia molecular e da genética, entre outros [4].

Portanto, considerando esses fatores e as necessidades existentes na etapa de triagem, tem sido proposto o desenvolvimento de Sistemas de Apoio à Decisão (SAD), que utilizam modelos computacionais para auxiliar na interpretação dos dados, no reconhecimento de padrões e no direcionamento para uma tomada de decisão [5]. Ao levar em consideração o volume de variáveis a qual a equipe de saúde é exposta, os SAD se apresentam como uma poderosa ferramenta no entendimento e processamento das informações associadas e das incertezas clínicas correspondentes. Além disso, de forma a garantir sua aplicabilidade, o sistema deve visar ser de fácil utilização e implementação, para proporcionar a agilização do processo, e não acarretar em altos custos.

## 1.1 Motivação

A tuberculose (TB) é a nona causa de mortalidade no mundo e a principal entre as doenças transmissíveis, se posicionando a frente da AIDS/HIV [6]. Aproximadamente um quarto da população mundial está infectada pelo bacilo *Mycobacterium tuberculosis*. Estima-se que 10,4 milhões de pessoas adoeceram com TB em 2016, sendo que apenas 6,3 milhões de casos foram registrados em unidades de saúde. Em relação à mortalidade decorrente da doença, nesse mesmo ano, cerca de 1,6 milhões de pessoas vieram a óbito [6]. Entretanto, a maior parte desses desfechos poderiam ter sido evitados se o diagnóstico tivesse sido realizado de forma rápida e o tratamento apropriado fosse disponibilizado. Por exemplo, em 2015, dentre os casos onde a TB foi detectada, 83% obtiveram sucesso no tratamento. [6]. Entretanto, os resultados são menos alvissareiros nas regiões com elevada proporção de TB-DR (resistente a pelo menos um medicamento), TB-MDR (resistente a rifampicina e a isoniazida) ou TB-XDR (MDR + resistente a fluoroquinolonas e aminoglicosídeos).

Nestas regiões a TB-DR/MDR/XDR tornou-se uma ameaça à saúde pública e renova o senso de urgência em relação a medidas para o controle de TB, em nível global. Esse tipo da doença pode decorrer de tratamentos inadequados, o que inclui o abandono total ao tratamento; interrupção do tratamento; uso irregular da medicação, como doses incorretas ou por não ser levada em consideração a interação medicamentosa; recusa do doente em se tratar ou irregularidade no fornecimento de medicamentos no sistema de saúde. Ainda, um paciente, que desenvolva a TB em sua forma resistente, poderá transmiti-la para outros indivíduos, o que agrava ainda mais a situação. No mundo, em 2016, um total de 600 mil casos de TB-DR/MDR foram estimados, sendo que 82% desses foram referentes a TB-MDR, e dentre os quais 240 mil vieram a falecer [6]. Porém, o número de casos que iniciaram o tratamento foi de apenas 22% desse total. Globalmente, 3,5% dos novos casos de TB e 20,5% de casos com tratamento anterior de TB são estimados como sendo TB-MDR [7].

O diagnóstico da TB pulmonar é realizado com base na anamnese associada ao uso de testes diagnósticos. A baciloscopia e a cultura para a micobactéria são testes comumente indicados para o diagnóstico da TB pulmonar. A baciloscopia direta do escarro é o exame prioritário para os casos suspeitos de TB pulmonar, permitindo identificação do paciente bacilífero, ou seja, com grande potencial de contaminação por contato. Por ser um método simples e seguro, é praticado em todos os serviços de saúde que possuem um laboratório. Como desvantagem, a baciloscopia possui uma baixa sensibilidade (probabilidade de detecção), entre 40% e 80% [8]. Além disso, a baciloscopia não consegue identificar as cepas resistentes ao tratamento, assim como micobactérias atípicas. Já a cultura para a micobactéria é o teste de maior sensibilidade (>85%) para o diagnóstico de TB pulmonar. Porém, a demora em se obter o resultado, de 4 a 8 semanas, é um grande inconveniente, o que a torna pouco utilizada no processo de tomada de decisão clínica [8]. Além disso, a cultura exige infraestrutura maior em biossegurança, dificultando sua implementação em locais de poucos recursos. Desse modo, até o momento, a cultura é indicada a suspeitos de TB que apresentem diagnósticos negativos na baciloscopia, além de

ser indicada para os casos extrapulmonares, HIV positivos e em crianças. Ainda, a diferenciação de cepas sensíveis e resistentes aos medicamentos é possível, sendo com isso recomendada para pacientes com suspeita de resistência. Já a radiografia de tórax é utilizada como exame auxiliar, porém sua grande desvantagem é a logística necessária para se realizar esse procedimento, assim como seu custo.

Desde 2006, seguindo o plano global *STOP TB*, novas abordagens em diagnóstico foram propostas pela OMS, isto é, a utilização de algoritmos clínicos-radiológicos que incluem métodos de teste por cultura líquida ou testes moleculares, como o Xpert MTB/RIF, em regiões de alta prevalência [9]. O teste Xpert MTB/RIF possui um grande potencial para revolucionar o diagnóstico de TB-DR (detecção de resistência a rifampicina/RR). Porém, um resultado que venha a ser positivo para RR com esse teste ainda precisa ser interpretado posteriormente por um especialista para a tomada de decisão terapêutica. Ainda, o risco de TB-MDR deve ser levado em consideração, assim como uma alta taxa de falsos positivos com Xpert MTB/RIF é esperada em unidades de saúde com prevalência de TB-MDR abaixo de 10% (normalmente pacientes suspeitos de TB-DR são atendidos no primeiro ou segundo nível de atendimento básico de saúde, em países com renda baixa ou média) [10].

No Brasil, diversos avanços no controle de TB foram alcançados nos últimos 10 anos; porém, grandes desafios ainda precisam ser superados, incluindo o baixo percentual de detecção de TB-DR e a alta morbidade e mortalidade entre casos de TB-MDR. Em 2015, foram realizados exames de cultura em apenas 23,1% dos novos casos notificados de TB, e somente em 30,7% de casos previamente tratados para TB [11]. Em 2012, foram registrados 812 casos de TB-DR, aumentando para 1.027 em 2015. Desse total, 548 (53,3%) eram de relacionados à TB-DR, e 442 (43,0%) à multirresistência [11]. Já em 2016, esse número aumentou para 1.044 e, dentre os casos de multirresistência, apenas 66,5% evoluíram para a cura e 16,7% abandonaram o tratamento [12]. É ainda observada uma redução significativa no tempo entre a coleta da amostra para diagnóstico e o início do tratamento apropriado para pacientes diagnosticados com TB-DR, quando se utiliza o Xpert (média de 15 dias), em detrimento da cultura/MGIT (média de 40,5 dias). Entretanto, um alto

percentual de tratamentos empíricos iniciados diretamente por meio da abordagem clínica ainda persiste em ambos os métodos [13].

Considerando essas questões operacionais, o alto custo do Xpert MTB/RIF e seu baixo valor preditivo positivo ( $\leq 80\%$ ) em regiões onde a prevalência de TB é menor que 10% entre os suspeitos de TB, testes em triagem têm sido propostos e discutidos com o objetivo de apoiar a identificação de indivíduos com uma maior probabilidade de ter a doença [14].

A estratégia global para erradicação da TB (*Global End Tuberculosis (TB) Strategy* [15], proposta pela OMS, tem como meta a redução de mortes por TB, assim como a sua incidência em todos os países para atingir os níveis observados em países de alta renda. Essa estratégia é baseada em três pilares: tratamento e prevenção integrados centrados no paciente (Pilar 1); políticas arrojadas e sistemas de apoio (Pilar 2); e esforços em pesquisa e inovação intensificados (Pilar 3) [16]. Portanto, existe uma busca por soluções inovadoras, assim como pela diminuição no tempo de diagnóstico de TB-DR e a redução da mortalidade e morbidade [17]. O desenvolvimento de uma solução para triagem direcionada para suspeitos de TB-DR/MDR pode auxiliar a identificar os casos a serem testados pelo Xpert MTB/RIF. Além disso, a possibilidade de auxiliar no entendimento de como cada sintoma ou resposta dada por um paciente pode se relacionar com a doença, assim como características relacionadas à regionalidade, podem ser também de grande valia.

## 1.2 Objetivos

Este trabalho propõe o desenvolvimento de escores clínicos para realizar a triagem para testes moleculares, como o Xpert MTB/RIF, de pacientes com suspeita de TB-DR encaminhados a diferentes centros de referência no Brasil. Dessa forma, visa-se encontrar soluções que apresentem uma alta sensibilidade, para apoiar o encaminhamento dos pacientes, e que seja de fácil aplicação em unidades de saúde. Portanto, modelos computacionais para a triagem de TB-DR/MDR e TB sensível aos medicamentos (TB-DS) são avaliados, considerando em seu desenvolvimento informações características do paciente, seus sintomas, fatores históricos e exames.

Para possibilitar sua aplicação em localidades com uma menor disponibilidade de recursos, o estudo também deve considerar diferentes cenários de operação e o seu potencial de adaptação a esses ambientes. Aspectos como a validade e a acessibilidade das informações utilizadas devem ser levados em consideração. Dessa forma, modelos com níveis de complexidade distintos são testados, onde conjuntos mais restritos de variáveis são selecionados por meio de critérios de relevância e qualidade dos dados, de acordo com diferentes estratégias para cada cenário. Ainda, é levantada a hipótese de que indicadores operacionais do sistema de saúde (taxa de mortalidade, cura, abandono do tratamento, entre outros) possam influenciar na triagem, ou mesmo eventuais características da população em diferentes localidades. Portanto, são desenvolvidos modelos direcionados a estados do Brasil, no caso o RJ, CE e RS por sua representatividade estatística, de forma que se possa comparar seu desempenho ao alcançado por meio de uma abordagem nacional. Modelos baseados em Redes Neurais Artificiais (RNA) e árvores de regressão e classificação (CART) são desenvolvidos e seus resultados comparados, explorando a sua capacidade de extrair correlações não-lineares dos dados e aproximar funções contínuas [18] [19]. Ainda é aplicado um método de *boosting* com modelos CART, no caso o AdaBoost, o qual permite uma tomada de decisão realizada por meio da combinação de respostas fornecidas por diversos modelos [20].

### 1.3 Organização do documento

No capítulo seguinte, a TB e suas características são descritas em uma breve introdução, incluindo os sintomas e os métodos de diagnósticos existentes. Uma revisão da literatura que descreve a aplicação de modelos estatísticos para a geração de escores clínicos em triagem médica é realizada.

No capítulo 3, os métodos aplicados são descritos. Já no capítulo 4, os resultados e as discussões correspondentes são apresentados, e o capítulo 5 inclui as conclusões e perspectivas de trabalhos futuros.

# Capítulo 2

## Tuberculose

Este capítulo apresenta um panorama sobre a relevância da TB em nível de saúde global, além de definir as variações da doença, que incluem as suas formas resistentes aos medicamentos utilizados no tratamento. São descritos os fatores de risco, além dos sintomas correspondentes e as formas de diagnóstico existentes.

Um estudo sobre os diferentes métodos estatísticos mais utilizados para o desenvolvimento de escores clínicos é apresentado, assim como é feita uma revisão da literatura que relata o desenvolvimento de modelos computacionais aplicados na área de medicina e no diagnóstico de TB e TB-DR em triagem.

### 2.1 Tuberculose pulmonar

#### 2.1.1 Incidência, métodos diagnósticos e tratamento

A TB é causada pelo bacilo *Mycobacterium tuberculosis*, que afeta mais frequentemente os pulmões e é transmitida de pessoa a pessoa através do ar. Quando um doente com TB pulmonar tosse, espirra ou cospe, ele emite gotículas contaminadas no ambiente, estas, ao serem inaladas por terceiros, os contamina. Em 2016, 10,4 milhões de pessoas adoeceram com TB, acarretando 1,67 milhões de mortes (o que inclui 374 mil de pessoas também com HIV), sendo ainda a principal causa de morte entre pessoas com HIV positivo, representando 35% do total. Nesse mesmo ano, do total de novos casos notificados de TB, 57% possuíam HIV positivo [6].

Cerca de um quarto da população mundial possui TB latente (não ativa), ou seja, pessoas que estão infectadas com o bacilo *M. tuberculosis* mas (ainda) não estão doentes e também não transmitem a TB para a comunidade [6]. Pessoas infectadas com a bactéria têm 10% de chance em seu período de vida de ficar doente. Entretanto, se o sistema imunológico estiver comprometido, como em pessoas infectadas por HIV, mal nutridas, com diabetes, depressão, ou mesmo fumantes, a chance da doença se manifestar se torna maior. Estudos apontam que pessoas infectadas com HIV possuem de 20 a 30 vezes mais chance de desenvolver TB ativa [6]. Já o uso de tabaco é atribuído como responsável por mais de 20% dos casos de TB [6].

Quando a pessoa desenvolve a TB pulmonar em sua forma ativa, os sintomas (como tosse, febre vespertina, sudorese noturna, ou perda de peso) podem ser leves por muitos meses. Isso pode acarretar na demora em se buscar tratamento e o doente se torna um transmissor da doença enquanto isso, auxiliando na sua proliferação. Estima-se que pessoas com TB ativa podem infectar de 10 a 15 outras pessoas por meio de contato em um período de um ano. Sem o tratamento adequado, em um período de 5 anos, em média 45% de pessoas com HIV negativo e aproximadamente todos com HIV positivo evoluem para o óbito [6].

Após adentrar no organismo humano pela via respiratória, a bactéria pode disseminar-se e instalar-se em qualquer órgão. Com isso, ainda existe o risco de casos (cerca de 10 a 15% do total infectado) em que a TB seja extrapulmonar. As principais formas de TB extrapulmonar são: TB pleural, TB ganglionar periférica, TB meningoencefálica, TB pericárdica e TB óssea [21]. Geralmente, quando se está detectando a TB extrapulmonar, o pulmão pode também já estar afetado [22]

No ambiente hospitalar, caso não sejam adotados procedimentos adequados ao fluxo de pacientes com TB pulmonar, há um aumento de contágio da doença. Salas de espera cheias e/ou partilhadas por muitas especialidades, além de diversas etapas que possam ser criadas para o acesso do paciente ao médico, são um terreno fértil para a disseminação da doença. Medidas podem ser tomadas para evitar essa situação, como atendimento com hora marcada para evitar aglomerações, e salas de espera em locais bem ventilados, se possível ao ar livre [23]. Pacientes com sus-

peita de TB, ou com diagnóstico positivo ainda em período infectante, devem ter prioridade no atendimento, para tornar o tempo despendido na unidade de saúde o menor possível. No âmbito hospitalar, pacientes considerados de risco devem ficar em isolamento respiratório, em quartos individuais, ou em condições excepcionais, ter mais de um paciente por quarto, desde que os pacientes estejam com diagnóstico confirmado e não possuam o tipo resistente da doença [24].

A TB afeta principalmente adultos em seus anos mais produtivos. Entretanto, todas as faixas etárias estão sujeitas a adoecerem com a doença. Um exemplo disso é que no ano de 2016, 6,9% dos casos novos notificados de TB foram relativos a crianças de 0 a 14 anos. Observa-se também que mais de 95% dos casos notificados de TB estão localizados nos países em desenvolvimento [6].

A TB ocorre em todo o mundo. Em 2016, a maior incidência de novos casos de TB ocorreu nas regiões sudeste asiática (45%), africana (25%) e pacífica ocidental (17%); em menores proporções de incidência estão a região do Mediterrâneo oriental (7%), da Europa (3%) e das Américas (3%) [6]. Entretanto, cinco países são responsáveis por 56% dos casos: Índia, Indonésia, China, as Filipinas e o Paquistão. A Figura 2.1 exibe um mapa dos países com maior incidência. É observado também que a incidência de TB diminui em uma média de 1,5% ao ano, desde o ano 2000. Porém, esta taxa precisa chegar a 4-5% para atender as metas definidas para 2020 na "Estratégia Pelo Fim da Tuberculose" da OMS [15].

Os sintomas mais comuns de TB pulmonar ativa são tosse com escarro, e ocasionalmente sangue, dores no peito, sensação de fraqueza, emagrecimento, febre vespertina e sudorese noturna [26]. Vários países ainda se baseiam na baciloscopia para diagnosticar TB, devido às suas características de ser simples, rápido e barato. Dessa forma, técnicos de laboratórios treinados observam amostras de escarro por meio de um microscópio para analisar se a bactéria responsável pela TB está presente, porém, não detecta resistência.

Já a cultura é o método que possui uma sensibilidade alta para o diagnóstico de TB, detectando de 70% a 89% dos casos [27]. Esse exame permite a multiplicação e o isolamento dos bacilos a partir da sementeira da amostra clínica, em meios de



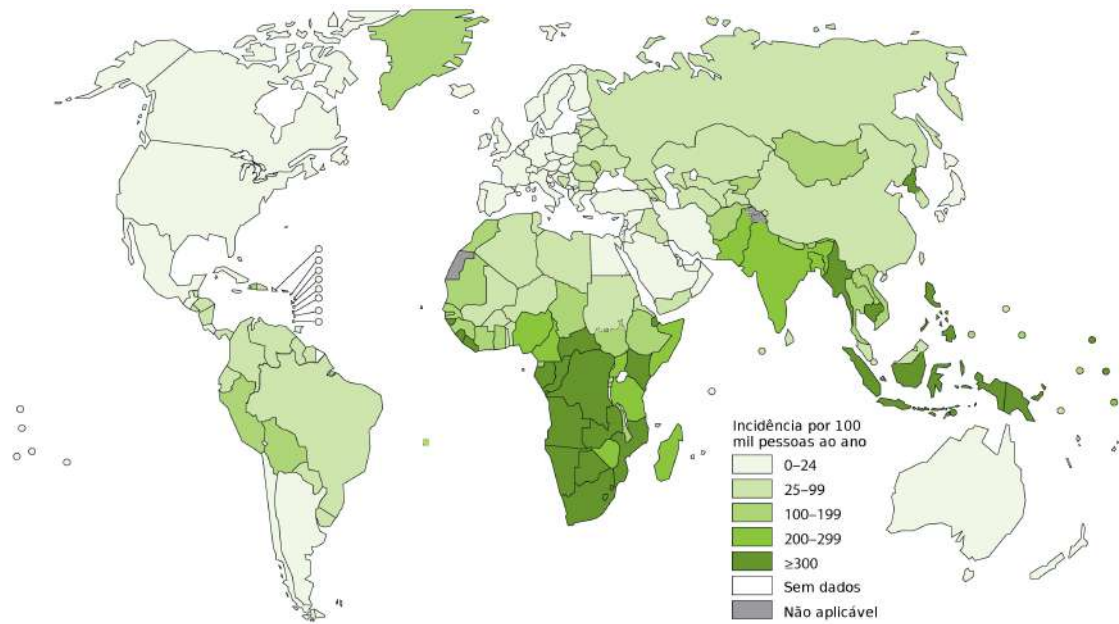


Figura 2.1: Mapa que representa visualmente a incidência de TB por país no ano de 2016. Fonte: OMS. Extraído de [25]

cultura específicos para micobactérias. Através de identificação por meio de testes bioquímicos ou moleculares, e a partir da cultura, é possível realizar a diferenciação de cepas sensíveis e resistentes aos medicamentos disponíveis para o tratamento [28]. A cultura possui o inconveniente da demora em obter seus resultados [8]. Além disso, a cultura exige infra-estrutura maior em biossegurança, que torna difícil sua implementação em locais de poucos recursos [29].

A radiografia de tórax é utilizada como exame auxiliar, apresentando proporções de sensibilidade de 70% a 80% na presença das anormalidades típicas em adultos imunocompetentes, e 95% se considerarmos outras anormalidades, entretanto, em torno de 10% dos pacientes com AIDS a radiografia pode ser normal [30]. Além disso, este exame não auxilia na caracterização de TB resistente.

O uso de testes rápidos como o do Xpert MTB/RIF tem aumentado substancialmente desde 2010, quando a Organização Mundial de Saúde (OMS) recomendou a sua utilização pela primeira vez. Esse teste consegue detectar, simultaneamente, TB ativa e a resistência a rifampicina, droga mais importante no tratamento de TB. O diagnóstico pode ser realizado em até 2 horas, e o teste é agora recomendado pela OMS como o teste diagnóstico inicial para todas as pessoas com sinais e sintomas

de TB. Mais de 100 países estão utilizando esse método, de tal modo que, em 2016, cerca de 6,9 milhões de cartuchos (necessários para a realização do teste) foram adquiridos globalmente.

Ainda, é visto que a TB é particularmente difícil de diagnosticar em crianças, uma vez que, especialmente quando elas possuem menos de 5 anos de idade, existe uma dificuldade em expelir catarro, exigindo a realização de outros métodos para retirar o material necessário para os exames, como aspiração nasofaríngea ou gástrica [31].

A TB é uma doença tratável e possui cura, e quando não há resistência, é possível aplicar um tratamento padrão com duração de 6 meses, onde 4 medicamentos antimicrobiais são administrados por profissionais de saúde em conjunto com informação sobre a doença e a supervisão e suporte ao paciente [32]. Sem um acompanhamento do doente durante esse período, a aderência ao tratamento pode ser comprometida e a doença continuar sendo transmitida. Assim, a grande maioria dos casos de TB podem ser curados com os medicamentos sendo aplicados de forma apropriada. Estima-se que entre os anos 2000 e 2015, 49 milhões de vidas foram salvas por meio do correto diagnóstico e tratamento de TB [6].

### **2.1.2 Resistência aos medicamentos em TB**

Medicamentos anti-TB têm sido utilizados por décadas e casos de resistência a uma ou mais dessas drogas têm sido documentados em todos os países pesquisados. A resistência às drogas surge quando os medicamentos anti-TB são utilizados de forma inapropriada, por prescrição incorreta dos médicos, má qualidade da medicação e interrupção no tratamento [33].

Estudos ainda apontam que as condições de moradia, como falta de esgoto no domicílio, além de alcoolismo associado ao tabagismo, número de tratamentos anteriores, tratamentos irregulares e cavidades pulmonares são importantes fatores para o desenvolvimento de TB-DR [34][35]. Em geral, os fatores de risco relacionados à forma resistente da doença são similares aos da não resistente, devido às suas proximidades nas formas de transmissão e desenvolvimento do agente causador.

A TB-MDR é a forma da doença causada pela micobactéria que apresenta resistência a isoniazida e rifampicina, as duas drogas de primeira linha mais poderosas. A doença é tratável e curável com o uso de drogas de segunda linha; entretanto, as opções são limitadas e exigem uma longa quimioterapia (até dois anos de tratamento) com medicamentos caros e tóxicos [36].

Em alguns casos, a resistência pode ser ainda mais grave como na TB-XDR, sendo causada pela bactéria que não responde aos medicamentos de segunda linha, frequentemente deixando os pacientes sem opção de tratamento. O diagnóstico de TB-MDR ou TB-XDR, assim como no caso onde a TB está associada ao HIV, se torna complexo e acarreta em altos custos.

Em 2016, foram notificados 600 mil novos casos de TB com resistência a rifampicina (o medicamento mais efetivo de primeira linha), dos quais 490 mil eram casos de TB-MDR. A incidência de TB-MDR se encontra, principalmente, em três países - China, Índia e Rússia - os quais representam quase a metade dos casos notificados. Em 2016, um total estimado de 4,1% de novos casos e 19% de casos que já tiveram algum tipo de tratamento foram de TB-MDR e TB-DR, respectivamente. Cerca de 6,2% dos casos de TB-MDR eram de TB-XDR. [6].

No mundo, em somente 54% dos pacientes com TB-MDR, e 30% de TB-XDR, observou-se resultado favorável ao tratamento adotado [6]. Em 2016, a OMS aprovou a utilização de um regime terapêutico de curta duração padronizado para pacientes com TB-MDR que não possuem resistência a medicamentos de segunda linha [6]. Esse regime leva em torno de 9 a 12 meses de tratamento com custos menores que o tratamento convencional para TB-MDR, o qual pode durar até 2 anos. Entretanto, pacientes com TB-XDR ou com alguma resistência a medicamentos de segunda linha não são elegíveis para esta abordagem, precisando de tratamentos mais longos. Como pode ser visto na Figura 2.2, o Brasil ainda não está incluído no conjunto de países que aplicam esse novo regime.

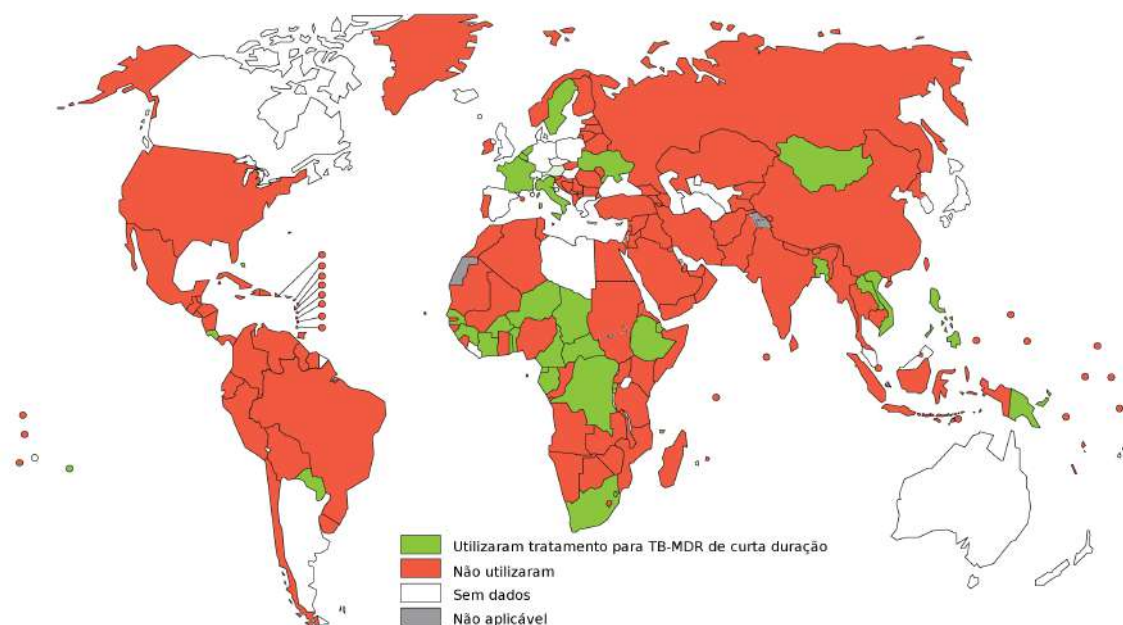


Figura 2.2: Mapa que indica os países que estão utilizando o regime terapêutico de curta duração para tratamento de TB-MDR. Fonte: OMS. Extraído de [25]

### 2.1.3 Tuberculose no Brasil

No Brasil, em 2016, foram diagnosticados 69 mil novos casos de TB, dos quais cerca de 4,6 mil mortes foram notificadas como decorrência da doença [37]. De acordo com a nova classificação da OMS 2016-2020, o Brasil ocupa a vigésima posição na lista dos 30 países com alta carga de TB e a 19ª posição na lista dos 30 países com alta carga de TB/HIV [37]. No país, a TB é a terceira causa de mortes entre doenças infecciosas, sendo a primeira dentre pacientes contaminadas por HIV/AIDS [37]. Entre 2005 e 2014, uma média de 70 mil casos novos, além de 4,4 mil mortes por TB, foram registradas por ano [11].

As diferentes regiões do Brasil são impactadas e lidam com a TB de formas distintas, o que pode ser observado por meio de estatísticas e indicadores operacionais relacionados a doença. Nesse sentido, na Figura 2.3 estão listadas as Unidades Federativas (UF) do Brasil em ordem crescente de incidência de TB, considerando casos relativos ao ano de 2016. Observa-se que os estados com maior incidência por 100 mil habitantes é o Amazonas (67,2) e o RJ (61,2), os quais, nesse critério, se encontram destacados do demais. Ainda, considerando que a média do Brasil é de 30,2 casos, 11 unidades federativas se encontram em um patamar acima, incluindo, os

estados do CE, RS e SP. Com foco nessas regiões, resultados relativos a indicadores operacionais também são destaques. Em relação a proporção de cura de novos casos, RJ (70,6), CE (64) e RS (63,4) se encontram abaixo da média nacional (71,9). Já ao compararmos a proporção de abandono do tratamento, é possível observar que RS (16,8) é o estado que apresenta a maior proporção. Além disso, RJ (12,3), CE (11,4) e SP (10,2) também se encontram com valores acima da média (9,3). Apesar do RJ apresentar números melhores relacionados a proporção de cura de casos novos de TB-MDR, seu coeficiente de mortalidade por TB (5) é o maior do Brasil. Ainda, os estados do CE e RS também não alcançam bons desempenhos em ambos os indicadores. Dessa forma, todos esses fatores apontam para possíveis problemas operacionais nesses estados em relação a TB, colocando-os como potenciais cenários para aplicação de novas abordagens que possam vir a apoiar a cascata diagnóstica e terapêutica da TB e TB-DR. Os valores de todas as taxas e indicadores podem ser vistos na Tabela 2.1.

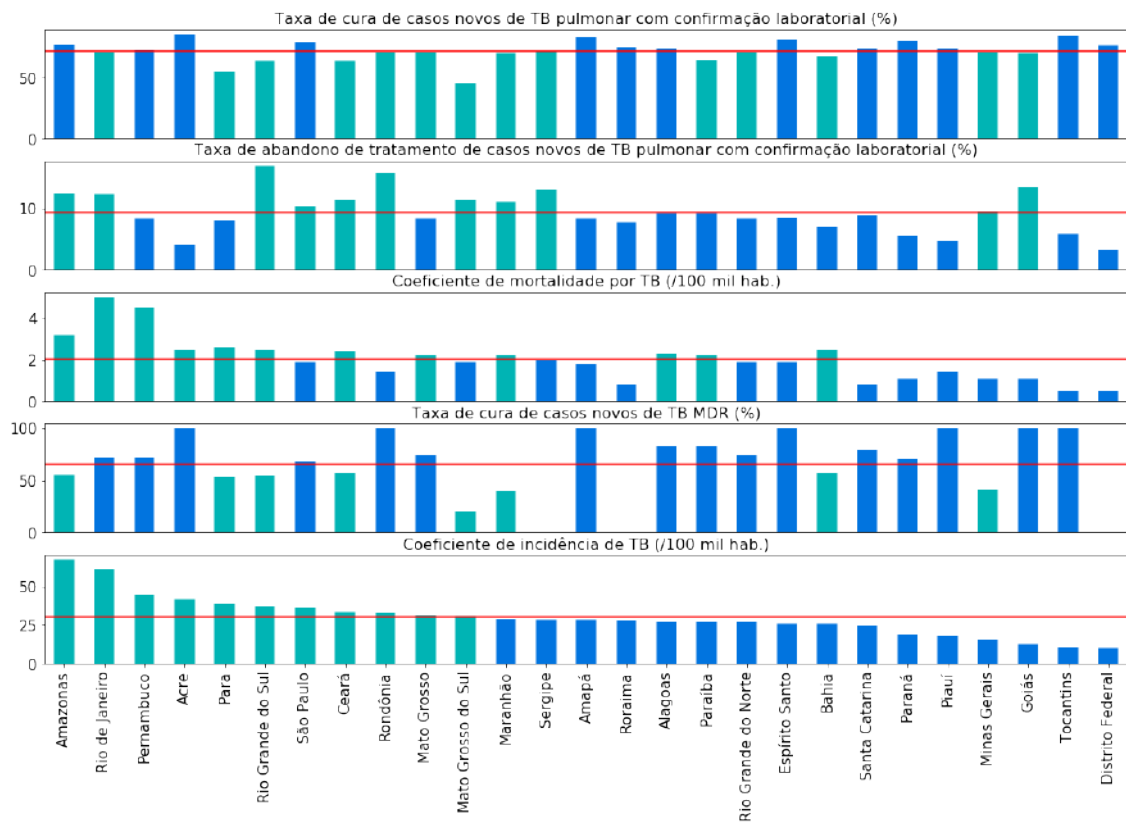


Figura 2.3: Desfechos operacionais e índices de TB no Brasil por unidade federada, no ano de 2016. Fonte: Ministério da Saúde. Dados extraídos de [12]

Tabela 2.1: Lista com os valores relativos a índices e desfechos operacionais de TB por unidade federativa, no ano de 2016. Fonte: Ministério da Saúde. Dados extraídos de [12]

UF	Coefficiente de incidência de TB (/100 mil hab.)	Coefficiente de mortalidade por TB (/100 mil hab.)	Contatos examinados dos casos novos de TB pulmonar com confirmação laboratorial (%)	Casos novos de TB confirmados por critério laboratorial (%)	Sensibilidade entre os casos de retratamento de TB pulmonar com cultura positiva (%)	Casos novos de TB pulmonar que realizaram o TDO (%)	Taxa de cura de casos novos de TB com confirmação laboratorial (%)	Taxa de abandono de tratamento de casos novos de TB com confirmação laboratorial (%)	Taxa de cura de casos novos de TB MDR (%)
Amazonas	67.2	3.2	54.1	67.7	21.1	20.5	77.2	12.5	55.6
Rio de Janeiro	61.2	5.0	39.8	65.8	47.6	35.7	70.6	12.3	72.4
Pernambuco	45.0	4.5	50.6	63.3	20.8	37.6	72.4	8.3	72.0
Acre	41.4	2.5	86.1	84.3	50.0	54.3	84.8	4.2	100.0
Para	39.3	2.6	32.9	73.8	35.6	25.8	55.1	8.0	53.6
Rio Grande do Sul	37.5	2.5	45.7	71.5	57.6	17.2	63.4	16.8	54.5
São Paulo	36.4	1.9	59.2	80.1	62.2	44.7	78.9	10.2	68.4
Ceará	33.5	2.4	46.1	69.5	27.8	33.3	64.0	11.4	57.7
Rondônia	33.1	1.4	27.9	80.1	61.5	22.9	71.1	15.8	100.0
Mato Grosso	31.3	2.2	36.8	39.7	21.4	39.1	70.7	8.4	75.0
Mato Grosso do Sul	30.9	1.9	57.0	69.2	11.9	39.2	45.7	11.4	20.0
Maranhão	29.0	2.2	45.9	62.9	56.1	25.7	70.3	11.0	40.0
Sergipe	28.6	2.0	51.5	76.0	16.7	37.2	71.6	13.1	0.0
Amapá	28.4	1.8	53.8	82.6	0.0	40.5	83.3	8.3	100.0
Roraima	27.6	0.8	56.4	85.1	50.0	51.6	74.8	7.8	0.0
Alagoas	27.2	2.3	47.2	66.7	68.0	29.4	74.3	9.1	83.3
Parabá	27.0	2.2	29.4	75.6	24.3	21.2	64.7	9.2	83.3
Rio Grande do Norte	27.0	1.9	38.8	67.4	66.7	22.4	70.5	8.4	75.0
Espírito Santo	26.2	1.9	63.4	85.0	26.7	29.6	80.8	8.6	100.0
Bahia	26.1	2.5	31.6	70.2	26.5	20.7	67.8	7.0	57.6
Santa Catarina	24.7	0.8	76.6	71.1	40.0	58.3	73.9	8.8	80.0
Paraná	18.9	1.1	75.2	73.9	49.4	74.8	80.6	5.6	71.4
Piauí	18.5	1.4	42.6	69.8	18.2	48.6	73.9	4.8	100.0
Minas Gerais	15.8	1.1	65.1	72.6	40.0	43.0	71.0	9.5	41.2
Goiás	13.0	1.1	77.7	73.3	48.0	37.0	70.3	13.4	100.0
Tocantins	10.7	0.5	48.0	76.4	14.3	68.1	84.0	5.9	100.0
Distrito Federal	10.5	0.5	51.2	71.5	80.0	50.2	76.5	3.3	0.0

O teste rápido Xpert MTB/RIF teve seu primeiro ano de implantação no Brasil entre junho de 2014 a maio de 2015. Nesse período, 146.939 testes para casos novos e de retratamento foram realizados, refletindo em um média de 12.245 por mês [38]. Os equipamentos foram implantados em 25 unidades federativas, com 265.600 mil cartuchos distribuídos entre eles, o que inclui o RJ, CE e RS. A percentagem de erros na obtenção dos resultados ocorridos durante a realização do teste foram de 2,7% em média, onde o RJ apresentou uma proporção de 5,7% [38]. Esses erros são, em sua grande maioria, relativos ao processamento da amostra e/ou manipulação do equipamento, o que reflete a necessidade de se contar com profissionais especializados para operar o equipamento e coletar as amostras.

Quando analisamos o volume de notificações relacionadas a casos novos de TB-DR e TB-MDR, é possível ver, por meio da Figura 2.4, que houve um aumento nos casos notificados a partir de 2010. Apesar de haver, desde 2012, uma diminuição dos casos de TB-MDR, a identificação de casos de TB-DR continua aumentando, podendo ser observado que, por exemplo, de 2014 para 2015, houve um salto de 792 para 1.077 casos, o que corresponde a um aumento de 135%. Esse fato pode ser consequência da implantação do teste rápido Xpert MTB/RIF, já que como consegue diagnosticar resistência a rifampicina ajuda a aumentar o número de notificações de TB-DR, porém não influencia diretamente nos casos de multirresistência.

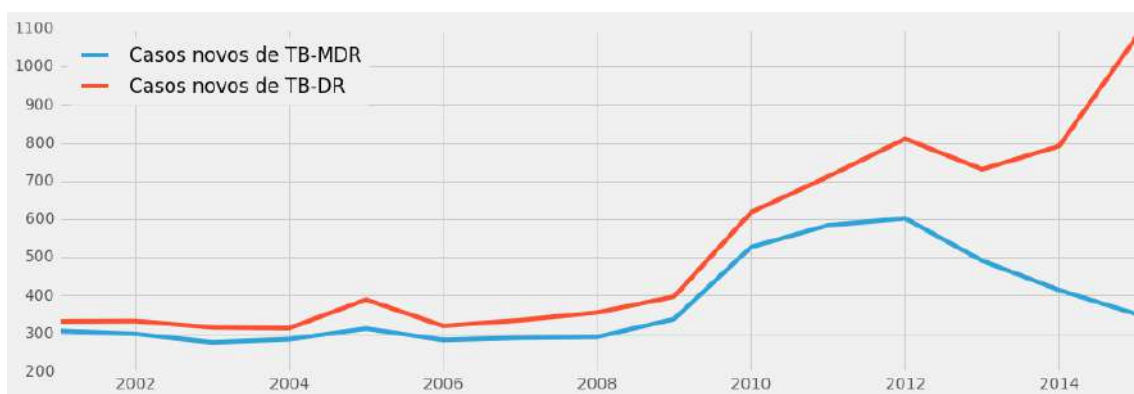


Figura 2.4: Evolução temporal do número de casos novos de TB-DR e TB-MDR no Brasil, do ano de 2001 a 2015. Dados extraídos de [39]

Já na Figura 2.5 é possível analisar a mortalidade por casos de TB-DR e TB-MDR no ano de 2014 para cada unidade federativa, sendo esse mais um indicador

operacional. Em relação a TB-DR, o estado do RJ se encontra em primeiro lugar em relação ao número de óbitos (115), seguido pelo RS (62) e com o Ceará (30) em quarto [40].

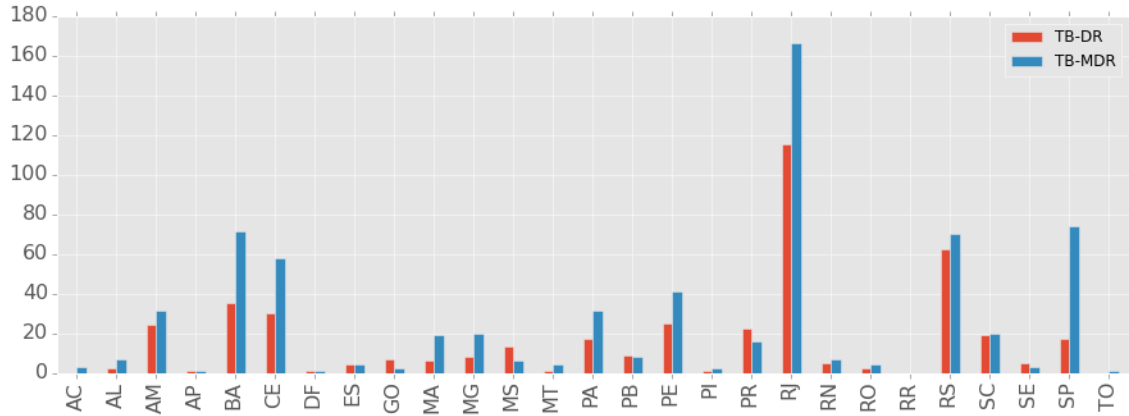


Figura 2.5: Mortalidade ocasionada por TB-DR e TB-MDR no ano de 2014, por unidade federativa. Dados extraídos de [40]

## 2.1.4 Modelos Estatísticos

A utilização de modelos estatísticos visa ser uma solução para apoio a triagem e ao diagnóstico que busca aliar eficiência em termos de alta sensibilidade, com custo-efetividade, e facilidade na interpretação dos resultados. Esses são fatores fundamentais para a aplicabilidade dos modelos, e que pode os tornar de grande utilidade em lugares com alta carga da doença, poucos recursos e/ou erros sistêmicos na triagem. Sua utilização na triagem de pacientes pode ainda auxiliar em diagnósticos mais complexos, como a TB-DR/MDR.

Diversos estudos têm sido realizados com o objetivo de apoiar a triagem médica de pacientes por meio do desenvolvimento de escores clínicos por modelagem estatística. Exemplos de tais aplicações podem ser encontrados na literatura, como no suporte ao diagnóstico de isquemia cardíaca [41] e síndrome coronária aguda em pacientes que chegam ao setor de emergência com dores no peito [42]; diagnóstico de hemorragia gastrointestinal superior aguda [43]; predição de risco de mortalidade por trauma na cabeça [44]; insuficiência pulmonar e incompetência tricúspide [45]; carcinoma gástrico [46]; e predição do desfecho para traumas [47]. Em todos esses exemplos, os escores clínicos aparecem como uma solução rápida, de baixo custo,



e de simples aplicação, que apoiam a tomada de decisão na triagem, evitando exames mais invasivos e acelerando o encaminhamento dos pacientes, o que é essencial quando se trata de doenças infecciosas.

Dessa forma, observa-se que o processo de diagnóstico médico pode ser auxiliado, ou mesmo melhorado, por meio da utilização de modelos estatísticos de forma complementar aos testes clínicos. A partir de bases de dados consistentes, os modelos são vistos como sistemas que conseguem representar o problema relacionado ao diagnóstico e, portanto, auxiliar na tomada de decisão dos profissionais de saúde. Os modelos estatísticos mais comuns de serem utilizados no apoio ao diagnóstico de doenças, inclusive TB, são a regressão logística, redes bayesianas, árvores de decisão e redes neurais artificiais. Cada método possui vantagens e desvantagens em termos de complexidade para a criação do modelo, adaptabilidade aos dados, e clareza em relação as suas regras de decisão.

Por exemplo, o método de regressão logística tem como objetivo produzir, a partir de uma série de variáveis explicativas contínuas e/ou binárias, um modelo que permita a predição de valores de variáveis categóricas, as quais são frequentemente binárias. Essa técnica é bastante utilizada na área da medicina. Entretanto, quando intervenções são aplicadas com esses modelos, os resultados tendem a ser direcionados para um indivíduo com perfil próximo da ocorrência média em uma população, desconsiderando as necessidades especiais de subgrupos de uma comunidade [48]. Além disso, interações estatísticas podem ser difíceis de interpretar, particularmente quando três ou mais variáveis são avaliadas ao mesmo tempo [49].

Em relação às redes bayesianas, elas consistem em representar um conjunto de variáveis aleatórias e suas dependências condicionais, mostrando, desse modo, a relação probabilística entre doenças e sintomas. Assim, é possível identificar as relações de causa e efeito entre as variáveis [50]. Apesar disso, existe certa complexidade em se especificar uma rede bayesiana para aplicações com inúmeras variáveis, já que a técnica requer os valores de todas as probabilidades condicionais e, antecipadamente, os valores das variáveis explicativas.

Já as árvores de decisão representam uma técnica que visa particionar, de forma

recursiva, um conjunto de dados, formando um modelo graficamente estruturado, compacto e que descreve, de forma clara, as regras de decisão para a classificação [51]. Em problemas mais complexos, árvores de decisão maiores tendem a ser necessárias, e podem surgir sub-árvores idênticas em caminhos diferentes. Desse modo, quanto maior o número de decisões a serem tomadas, mais pontos de decisão são necessários percorrer e menos acurados são os resultados.

As redes neurais artificiais (RNA) são utilizadas no reconhecimento de padrões, onde os modelos gerados se inspiram na estrutura e no funcionamento do cérebro humano, baseando-se em interconexões de unidades, chamadas de neurônios. As redes neurais podem ser treinadas de forma supervisionada, onde o conhecimento das respostas do modelo para cada entrada do sistema é necessário, ou não supervisionado, na qual as respostas são completamente desconhecidas do sistema [52]. Essa técnica é uma realidade na área médica, sendo eficiente em diversas áreas da medicina, inclusive no auxílio ao diagnóstico, prognóstico, e no reconhecimento de padrões em exames clínicos e imagens médicas. Uma desvantagem dessa abordagem é o processo necessário para o ajuste dos parâmetros para o desenvolvimento do modelo, o qual depende de métodos heurísticos, e que geram uma solução não replicável para outros problemas.

Existem ainda algoritmos que visam melhorar o desempenho de modelos estatísticos a partir de uma combinação ponderada do resultado de diversos modelos, sendo esses métodos de *ensemble* [53]. Além de combinar os resultados, é possível criar melhores classificadores a partir dos modelos originais. Conforme os modelos são gerados, pode-se favorecer, de forma sequencial, o aprendizado de amostras de dados que originaram erros de classificação, técnica chamada de *boosting* [54]. Diferentes algoritmos que implementam essas técnicas podem ser encontrados, sendo um dos primeiros e mais utilizados o AdaBoost [20]. Entre suas vantagens é a sua simples implementação e por apresentar resultados com boa generalização. Porém, as soluções encontradas podem ser sub-ótimas, sendo sensível a ruído nos dados e casos de exceções (*outliers*).

A eficiência do AdaBoost na melhoria do aprendizado de modelos computacio-

nais foi investigada em diversos estudos, em que utilizam, principalmente, as árvores de decisão, onde alcançam bons resultados. [55, 56]. Como um exemplo, Zhou, et al, compararam AdaBoost com CART e RNA para realizar o diagnóstico de arritmia no coração. O estudo tinha dois objetivos: diferenciar batidas de contração ventricular prematura de batidas normais, assim como discriminar de batidas prematuras atriais. Em todos os casos o conjunto AdaBoost e CART obteve a melhor acurácia [57]. Para diagnosticar doenças cardiovasculares, Tan, et al, exploraram o AdaBoost com CART (modelos de fraco aprendizado), em conjunto com análises de traço em fios de cabelo, para realizar a predição. A base de dados consistia em um total de 124 amostras de fio de cabelo relativos a 100 pessoas saudáveis e 24 doentes. O classificador final foi composto por apenas 4 modelos, alcançando uma acurácia de 95,2%, sensibilidade de 100% e especificidade de 94% [58]. Tan, et al, também aplicaram o mesmo método para a detecção de câncer no pulmão, combinando análises de traço em urina (122 amostras). Como resultado, os classificadores obtidos com AdaBoost e CART obtiveram uma sensibilidade de 100% em ambos os casos testados, além de uma especificidade acima de 93% [59]. Já a predição de crises de esquizofrenia foi estudado por Hrdlicka, et al, visando comparar os resultados obtidos por AdaBoost e com métodos já utilizados. Os dados era obtidos das respostas de pacientes por questionários enviados regularmente aos seus telefones móveis. Apesar da mesma sensibilidade ter sido encontrada (65%), a valor obtido para a especificidade foi superior, de 73% para 84% [60].

Ao se realizar uma revisão da literatura para TB, é possível encontrar diferentes métodos computacionais sendo aplicados e em cenários distintos. Utilizando regressão logística, Solari, et al, construíram um sistema de score para a predição de TB pulmonar em pacientes com sintomas respiratórios atendidos na emergência de um hospital, em Lima, Perú. As informações utilizadas consistiram de sintomas clínicos, histórico médico, dados demográficos e resultados de exames, levando a alcançar uma sensibilidade de 93% e especificidade de 42% [61]. Uma combinação entre regressão logística e modelos CART foi apresentada como solução por Mello, et al, para apoiar o diagnóstico de TB pulmonar com baciloscopia negativa em pacientes não

internados de áreas com recursos limitados. Modelos foram desenvolvidos utilizando os sintomas dos pacientes, sinais físicos e raio-X de tórax, atingindo sensibilidade entre 64% e 71% e especificidade no intervalo de 58% a 76% [62]. Também aplicando CART, Aguiar, et al, reportaram o desenvolvimento de um modelo para predição de TB pulmonar em pacientes hospitalizados, admitidos com suspeita de TB, em área de alta prevalência. Para isso foram considerados dados clínicos, fatores médicos com potencial de predição para o diagnóstico de TB, além de resultados de testes radiológicos, atingindo sensibilidade e especificidade de 60% e 76%, respectivamente [63].

Estudos com comparações entre diferentes métodos com o objetivo de apoiar o diagnóstico de TB também são encontrados. Nesse sentido, Wisnivesky, et al, realizou uma análise sobre a aplicação de regras de decisão na predição do risco de TB pulmonar ativa, em triagem, para reduzir o isolamento de pacientes sem TB [64]. Esta abordagem foi realizada por meio de uma revisão sobre diversos estudos que descreveram resultados em sensibilidade e especificidade alcançados pela modelagem de variáveis clínicas. Assim, os autores observaram que, pelo material analisado, a sensibilidade para identificar pacientes com TB pulmonar ativa variou entre 81% a 100%, com a especificidade entre 19% e 84%. Tais resultados indicam que regras de decisão podem ser utilizadas para identificar, dentre os suspeitos de TB, pacientes com risco muito baixo de infecção. Ainda, é também observado que, quando comparado a outros métodos, os modelos RNA apresentaram os melhores resultados em relação a sensibilidade e especificidade.

Uma comparação entre modelos estatísticos foi realizada por Souza Filho, et al, com o objetivo de apoiar o diagnóstico de pacientes com TB paucibacilar, para os quais os testes de diagnósticos utilizados apresentam baixa sensibilidade [65]. Os resultados apontaram que os modelos de RNA apresentam uma sensibilidade mais alta (100%) do que outras técnicas, como regressão logística multivariada (RLM) (79%) e CART (71%). Em relação à especificidade, RLM apresentou um valor ligeiramente acima da MLP (85% em comparação com 80%), porém, ambas com taxas maiores do que as alcançadas através de CART (65%). Foram utilizados

apenas sinais e sintomas dos pacientes para a geração dos modelos.

Dessa forma, muitos estudos aplicam RNA para apoiar o diagnóstico de TB. Para a predição de TB pulmonar ativa na cidade de Búfalo, Nova York, El-Solh, et al, utilizaram dados clínicos e de exames radiográficos, constituindo 21 variáveis, como idade, presença de diabetes, HIV, dor no peito, emagrecimento, tosse, febre, sudorese, falta de ar, além dos resultados de exames. Com isso, os autores alcançaram uma sensibilidade de 100% e especificidade de 72% [66]. Na Turquia, Er, et al, também abordaram o diagnóstico de TB ao modelar uma RNA com duas camadas escondidas de neurônios utilizando dados de pacientes em epícrise, o que incluiu 38 variáveis [67]. Informações sobre presença de tosse, dor no peito, temperatura do corpo, dispnéia, fraqueza, catarro, hábito de fumar e testes laboratoriais foram utilizadas, o que ao final apresentou uma acurácia de 93,93%. Com a mesma base de dados, algoritmos genéticos foram aplicados por Elveren, et al, para treinar uma rede neural com uma estrutura que apresenta duas camadas escondidas, também baseado em dados de epícrise, que incluíram resultados de exames preliminares e testes laboratoriais. Dessa forma, uma acurácia de 94,88% é descrita [68]. Já no Brasil, Cascão desenvolveu modelos de redes neurais para a triagem e diagnóstico de TB pulmonar [69]. Foram utilizadas redes neurais não supervisionadas, no caso mapas auto-organizáveis, para analisar as variáveis relevantes e apoiar a triagem de pacientes em uma policlínica da cidade do Rio de Janeiro. Métodos para o diagnóstico, além da análise de grupos de risco, também foram realizados. Ao utilizar o discriminante de Fisher [70], o melhor modelo alcançou um resultado de 81,4% de sensibilidade e 61,3% de especificidade. Com o objetivo de propor um método não invasivo para o diagnóstico de TB pleural, Seixas, et al, desenvolveram uma rede neural baseada no histórico clínico e informação sobre a presença de HIV de 137 pacientes, relatando uma acurácia superior a 90% [71]. Ainda no Rio de Janeiro, além de CART, Aguiar, et al, também aplicaram RNA para a modelagem de TB pulmonar, gerando uma rede para classificação e outra para avaliação de grupo de risco, por meio de mapas auto-organizáveis. Ao utilizar dados clínicos e radiológicos de pacientes hospitalizados, a classificação mostrou uma sensibilidade de 96,0%

com especificidade de 89,0%. Foi ainda identificado que a variável com maior poder discriminativo foi relativa a classificação radiológica. [72].

Existem poucos estudos que avaliam essa abordagem em uma triagem que também considera os diagnósticos de TB-DR/MDR. Em Lima, Perú, Martinez, et al, construíram uma combinação entre análise bivariada e regressão logística para desenvolver um escore clínico de predição para estratificar o risco de TB-MDR em pacientes com TB pulmonar [73]. As variáveis são incluídas no modelo por meio da análise bivariada e a base de dados foi formada por achados clínicos (teve TB anteriormente, contato com TB-MDR, exames pulmonares anormais) e uma característica radiológica (cavidade pulmonar). Os resultados apresentaram uma sensibilidade de 72,6% e especificidade de 62,8%. Já Boonsarngsuk, et al, analisaram retrospectivamente pacientes com cultura positiva em Bangkok, Tailândia [74]. Com o objetivo de desenvolver um modelo para o diagnóstico de pacientes recém diagnosticados com TB, um escore foi criado para realizar a discriminação entre TB-DS e TB-DR. As variáveis foram incluídas de acordo com sua significância estatística em um modelo de regressão logística multivariado, incluindo: informação demográfica, exames radiológicos e para doenças sexualmente transmissíveis, contaminação por HIV, comorbidades, histórico de tratamento anterior para TB, resultado da baciloscopia para escarro. Com isso, o modelo atingiu uma sensibilidade de 57,7% e especificidade de 67,8%.

Apesar de existirem trabalhos desenvolvendo modelos computacionais para várias regiões do mundo, o que inclui não só para o diagnóstico de TB, mas também outras doenças, as variáveis e condições dos pacientes não são idênticas, dificultando aqui uma análise comparativa externa. Apesar disso, indicativos de que abordagens direcionadas a regiões podem apresentar melhores resultados em diagnóstico médico podem ser vistas. Ruffino, et al, avaliaram o nível primário de atenção a saúde em controle de TB no Brasil para mostrar que diferentes cidades apresentavam características locais na forma de lidar com a doença [75]. O estudo foi conduzido nas cidades de Itaboraí, Ribeirão Preto e São José do Rio Preto na região sudeste, e Campina Grande e Feira de Santana no nordeste. Os dados foram coletados ao

se entrevistar 514 pacientes de TB em tratamento no ano de 2007. Como resultados, o acesso ao tratamento foi considerado como satisfatório no sudeste e regular no nordeste, que apresentou os piores resultados em acompanhamento domiciliar e distância entre o local de tratamento e a casa do paciente. Os tipos de serviços oferecidos foram avaliados como regulares, com melhor desempenho nas cidades do sudeste. Educação em saúde, acompanhamento do tratamento e auxílio alimentação são menos oferecidos no nordeste. Já coordenação foi considerado um fator satisfatório em todas as cidades. Por outro lado, a iniciativa do profissional de saúde em questionar sobre a família do paciente com relação a outros problemas de saúde foi insatisfatório em todas as cidades, exceto Ribeirão Preto. Assim, obstáculos foram vistos no tratamento de TB nos diferentes locais. Dentre elas, a dificuldade estrutural e o acesso aos serviços de saúde, e a organização em conjunto com a forma em que a tecnologia em saúde e os serviços são distribuídos e integrados.

A busca tecnológica por novos métodos de auxílio à detecção de TB tem encontrado soluções em modelos computacionais. No caso, percebe-se que os modelos baseados em RNA vêm sendo amplamente utilizados no diagnóstico e na avaliação de condições médicas [76]. O objetivo é obter resultados rápidos por meio de testes não-invasivos. Nesse sentido, já são encontrados no mercado aplicativos para telefone móveis que utilizam tais soluções. Por exemplo, o CRyPTIC é uma aplicação que visa modelar as características de resistência às drogas por meio da apresentação de amostras de genomas os quais são atribuídos a TB resistente. O produto não só ajuda na identificação do diagnóstico, mas também sugere a linha de tratamento a ser adotada [77][78]. Outro aplicativo é o Babylon, que permite a pessoas, por meio respostas a uma série de perguntas, ter uma indicação de sua possível doença e da urgência de seu tratamento. Essa resposta é gerada por um modelo computacional que promete processar bilhões de combinações de sintomas de forma mais rápida e acurada que o cérebro humano [79].

Portanto, de forma geral, é visto que os modelos computacionais podem auxiliar no apoio ao diagnóstico, na triagem de pacientes, prognóstico e no direcionamento para o tratamento, facilitando a ação dos profissionais da área da saúde. Os re-

sultados da literatura indicam que, para o diagnóstico de TB, modelos de RNA apresentam um desempenho melhor em termos de sensibilidade, quando comparados a outros métodos computacionais. Resultados preliminares utilizando modelos de regressão linear corroboraram com os encontrados na literatura, reforçando o direcionamento para a utilização desse método. Além disso, os níveis de especificidade também se mantêm em taxas mais elevadas. Na triagem, esse tipo de resultado é altamente útil, pois os pacientes infectados serão identificados com uma boa taxa de detecção, assim como pacientes que não possuem a doença não serão encaminhados para tratamentos, ou mais exames, sem necessidade.



# Capítulo 3

## Método

Essa seção apresenta os métodos aplicados neste trabalho. São detalhadas as etapas de pré-processamento dos dados e para o desenvolvimento dos modelos computacionais, o que compreende a geração do escore clínico para cada diagnóstico, tanto por RNA, quanto por CART. A aplicação do método de *boosting* pelo algoritmo AdaBoost também é descrita. Em seguida, os critérios para a avaliação do desempenho dos modelos são definidos. As análises sobre a relevância das variáveis na discriminação dos diagnósticos são descritas, assim como as estratégias adotadas para a geração de modelos de menor complexidade. Dessa forma busca-se lidar com problemas de qualidade de dados, assim como permitir sua aplicação em regiões que apresentem uma maior escassez de recursos.

### 3.1 Pré-processamento dos dados

O objetivo final de um problema de classificação com modelos computacionais é encontrar os parâmetros do modelo em que os dados de entrada gerem saídas que sejam classificadas da forma mais correta possível. Para isso, existe a fase de treinamento, na qual ocorre o ajuste dos parâmetros, ou seja, o aprendizado em si, até chegarem a seu valor final. Entretanto, alguns cuidados devem ser tomados antes de realizar esse procedimento. Se a rede for treinada com o conjunto total de dados disponíveis, a configuração pode se tornar tendenciosa a apenas acertar os dados aos quais o modelo foi previamente apresentado. Com isso, a classificação

de novas amostras, as quais não foram utilizadas no treino, pode ter sua eficiência prejudicada.

Para evitar essa situação, os dados devem ser divididos entre conjuntos de treino, teste e validação. Os dados de treinamento são utilizados para determinar os parâmetros do modelo (normalização, pesos e viés no caso da RNA, e pontos de decisão em CART), a validação para garantir a generalização do modelo e os dados de teste para estimar o desempenho alcançado. No caso do estudo aqui descrito, e como também acontece na literatura revisada, o volume de dados disponíveis é limitado, chegando em alguns cenários a refletir poucas dezenas de pacientes como amostra. Dessa forma, objetivando aumentar a quantidade de dados dedicados a validação e ao teste, os pacientes foram subdivididos apenas entre treino e teste, em uma proporção de 75% e 25%, respectivamente. Assim, o subconjunto de teste também é utilizado para validação [80].

Além disso, os dados apresentam incertezas estatísticas que devem ser levadas em consideração. Caso contrário, ao realizar a separação dos dados entre treino e teste, é possível que alguma determinada característica de um grupo de eventos fique representada em apenas um desses conjuntos. Isso pode prejudicar o processo de aprendizado do modelo. Para amenizar esse problema, é utilizada a técnica de validação cruzada [81]. Essa técnica, quando aplicada a modelos que possuem vários parâmetros de configuração a serem escolhidos, pode ser utilizada para externalizar as possíveis flutuações estatísticas. Dentre suas opções de implementação, nesse trabalho foi utilizada a técnica de *k-fold* [82].

A técnica de validação cruzada *k-fold* consiste na divisão do conjunto total de dados entre subconjuntos de quantidade similar entre si. De forma a manter a proporção de 75% e 25% entre os conjuntos de treino e teste, foi utilizado um total de 12 subconjuntos, sendo 9 selecionados para treino e 3 para teste. Uma ilustração da técnica pode ser vista na Figura 3.1. Para uma maior variedade estatística, o procedimento é realizado até que todas as combinações de subconjuntos entre treino e teste sejam utilizadas, sem repetição. Por outro lado, considerando razões de praticidade na avaliação do resultado, assume-se como suficiente para uma boa

generalização um número limite definido em 50 combinações a serem consideradas para as análises.

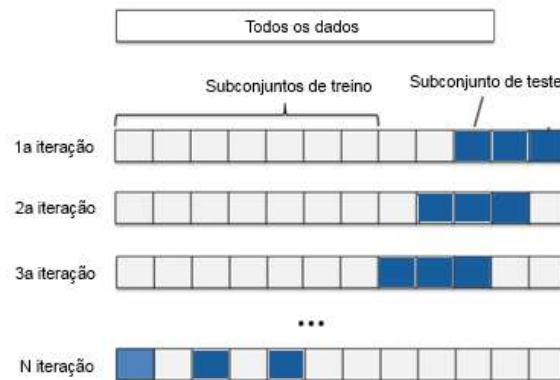


Figura 3.1: Ilustração da técnica de validação cruzada pelo método k-fold.

Os dados foram distribuídos entre os subconjuntos utilizados na validação cruzada após a aplicação da técnica de agrupamento de dados chamada de *k-means* [83]. O *k-means* é um método particional, amplamente utilizado pela comunidade científica, e que se baseia no conceito de centroides para realizar os agrupamentos. Dessa forma, para um conjunto de dados, definem-se N centroides (ou pontos de coordenada) espalhados aleatoriamente, e calcula-se a distância euclidiana quadrática de cada amostra a cada um desses pontos. As amostras mais próximas de cada centroide, em comparação com os outros, são considerados pertencentes ao mesmo agrupamento, e os centróides são recalculados para se tornarem o baricentro dos agrupamentos. O processo é repetido até não haver variação significativa da localização dos centróides. O número de agrupamentos a ser utilizado é indicado pela análise com o índice Davis-Bouldin, sendo que configurações utilizando de 1 a até 30 agrupamentos são testadas, e o menor valor obtido aponta o número de agrupamentos ótimo [84]. A utilização dessa técnica no processo de criação dos dados de treino e teste visa separar previamente o conjunto de dados por sua similaridade, e apoiar uma distribuição equilibrada garantindo a variedade estatística de cada subconjunto.

Conjuntos de treino, que possuam as classes com quantidades de pacientes desbalanceadas, também podem afetar a qualidade do treinamento. Para mitigar esse problema, amostras escolhidas de forma aleatória são replicadas dentro do conjunto

de menor representatividade, até as quantidades totais de cada classe serem iguais.

Ainda, uma das características de modelos computacionais gerados por CART é que eles conseguem lidar diretamente com dados numéricos e categóricos, assim como não necessitam de tratamentos para adequação das variáveis, como uma normalização [85]. Outro facilitador em termos de pré-processamento é que o modelo CART não possui limitadores em termos de dados faltantes, ou seja, um conjunto de dados pode ser aproveitado na geração do modelo mesmo que não exista um valor definido para alguma variável [86, 87]. Portanto, os modelos CART podem ser gerados pela aplicação direta dos dados disponíveis na base de pacientes original, ao contrário da RNA, que exige alguns cuidados adicionais para se adaptar a configurações do modelo e treinamento.

## 3.2 Desenvolvimento dos Modelos Computacionais

### 3.2.1 Redes Neurais Artificiais

As RNA são modelos computacionais capazes de realizar o aprendizado de máquina e o reconhecimento de padrões [82]. Elas são inspiradas pela arquitetura paralelizada do sistema nervoso central de um animal, no caso, o cérebro. No modelo, exemplificado na Figura 3.2, a informação é propagada nas redes pelas camadas de entrada, escondida e de saída, as quais são compostas de nós ou neurônios, por meio das sinapses ou conexões. Na RNA, a camada escondida atua como uma camada intermediária, tornando a rede capaz de extrair correlações não-lineares entre variáveis [88, 89]. Assim, os parâmetros do modelo são escolhidos pela apresentação de dados de entrada cujas saídas são conhecidos (treinamento supervisionado [82]),

Todos os modelos nesse trabalho utilizam redes neurais multicamadas (MLP) [71], sendo redes *feedforward*, ou seja, onde o fluxo de dados é sempre em uma única direção, além de conter uma única camada escondida, a qual é completamente conectada. Através das MLP é possível obter a probabilidade do diagnóstico de um paciente, em fase de triagem, a partir de seus dados, visando apoiar a decisão médica.

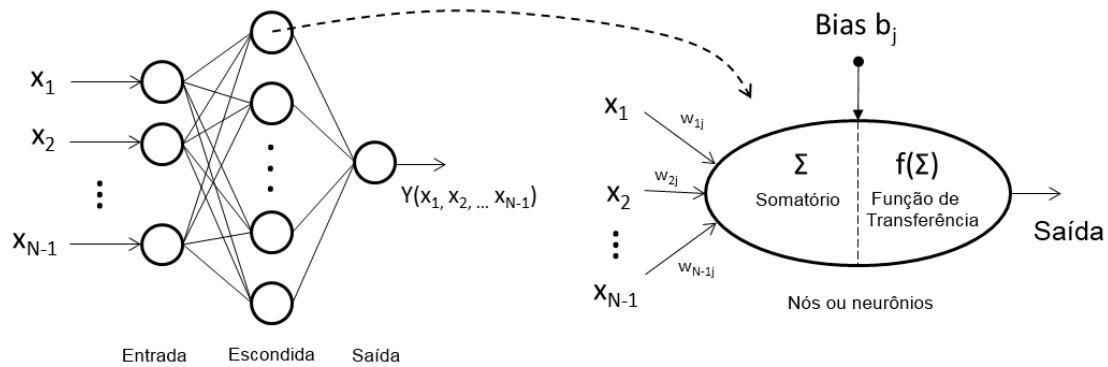


Figura 3.2: Ilustração de uma rede neural multicamadas (MLP), mostrando como a informação é processada, assim como a estrutura de um neurônio artificial.

### 3.2.1.1 Treinamento do modelo

Para escolher a melhor configuração de uma MLP, diversas redes devem ser treinadas e avaliadas, na busca por um resultado que alcance a melhor generalização possível. Uma RNA com um grande número de neurônios pode apresentar um bom ajuste para os dados de treinamento, porém, não garante um bom desempenho com a amostra de teste. Por outro lado, poucos neurônios podem não ser suficientes para realizar a modelagem e o aprendizado dos dados. Na literatura não é descrito um método ótimo para essa definição, sendo a escolha do número de neurônios da camada escondida realizada através de experimentos. Aqui, as topologias foram variadas para possuir dentre 1 a 10 neurônios na camada escondida, o que era avaliado como suficiente conforme a evolução do aprendizado em relação a essa configuração. Aumenta-se o total de neurônios na camada escondida apenas se o desempenho com o crescimento desse número, ao se avaliar dados de treino, indica uma possível melhora. Assim, existe uma preferência por redes com poucos neurônios, que tendem a possuir uma melhor generalização, possuindo menos graus de liberdade e reduzindo o problema de *overfitting* quando há restrição estatística no conjunto de dados. Assim, é considerado o conceito de parcimônia, onde busca-se o menor número possível de premissas para se conseguir explicar um fenômeno [86].

O algoritmo utilizado para treinamento foi o *Resilient Backpropagation*, o qual oferece uma bom tempo de convergência e necessita de pouco esforço computacional [90]. Entretanto, está sujeito a mínimos locais [82]. Para evitar esse cenário, foram

realizadas diversas inicializações para cada classificador proposto, selecionando-se apenas o melhor resultado. Para esse trabalho, foram feitas 100 inicializações para cada configuração de topologia. Dessa forma, levando em consideração a variação do número de neurônios, as 50 redes obtidas pela validação cruzada para cada configuração, e as inicializações realizadas para cada rede, ao todo 50 mil redes são geradas na etapa de treinamento para uma avaliação a posteriori.

Em relação à inicialização dos pesos sinápticos, o que pode influenciar na qualidade do treinamento, são utilizados valores aleatórios entre  $-0,2$  e  $0,2$ . Já a taxa de aprendizado escolhida depende da função a ser aproximada. A escolha de uma taxa muito pequena torna o treinamento lento, já valores muito grandes podem acarretar em sua divergência. Para os modelos gerados nesse trabalho, o valor inicial de  $0,01$  foi utilizado. Quando o sinal da derivada da função objetivo, em relação a um peso sináptico, alternava por 50 iterações consecutivas, a taxa era diminuída em  $30\%$ . Se o sinal permanece o mesmo, a taxa de aprendizado era aumentada em  $5\%$ .

Para alcançar a melhor generalização possível, o critério de parada do treinamento da rede deve considerar sua capacidade de generalização. Para isso, a cada iteração de treinamento, a rede é colocada em operação com o conjunto de dados de validação, sendo o seu erro de classificação monitorado. São utilizados dois critérios para a parada do treinamento, sendo eles a magnitude do gradiente sobre o desempenho e o número de validações realizadas. Assim, é considerado que se a magnitude do gradiente sobre o desempenho de treino se tornar menor que  $1^{-10}$  o treinamento deve ser encerrado, da mesma forma que se após 1000 validações sucessivas não houver decréscimo no seu valor de erro, o processo é finalizado.

Ao final desse processo, os pesos utilizados serão aqueles que obtiveram o menor valor de MSE de validação, seguindo o critério *Save the Best* [82]. A Figura 3.3 mostra um exemplo de evolução de treinamento e o critério de parada.

Já que todos os casos aqui estudados são voltados para a discriminação entre duas classes, a camada de saída das redes foi projetada para possuir apenas um nó. Portanto, os valores dos alvos são atribuídos como  $+1$  (o paciente possui a doença) e  $-1$  (o paciente não apresenta a doença). A função de transferência utilizada foi

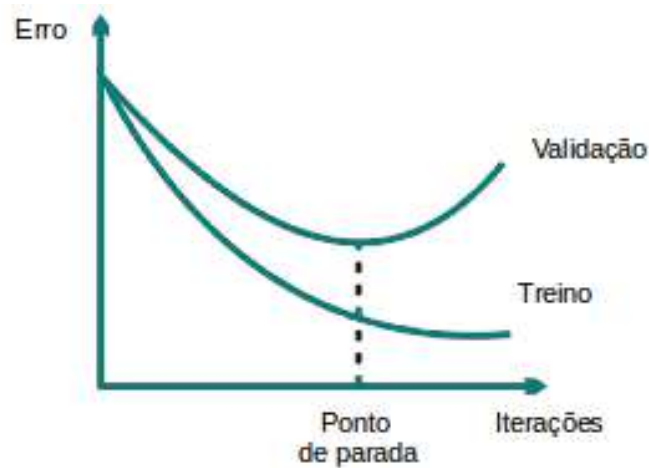


Figura 3.3: Representação da evolução do erro de classificação de um treinamento por iteração e escolha do ponto de parada com a melhor generalização.

a tangente hiperbólica para, dessa forma, o nó de saída produzir valores dentro do intervalo definidos como alvos.

Diferentes estratégias para a escolha do ponto de operação da rede ainda podem ser adotadas com o objetivo de favorecer algum critério de desempenho. Assim, os modelos podem ser direcionados para alcançar um dado valor de referência em termos de sensibilidade ou especificidade. Para isso, os modelos são colocados para operar no ponto de corte em que produz o desempenho desejado.

### 3.2.2 Árvores de Classificação

Em mineração de dados, uma árvore de decisão é um modelo preditivo que pode ser utilizado para representar tanto classificadores quanto modelos de regressão. Sua estrutura representa o modelo hierárquico de uma árvore invertida, o qual realiza a tarefa de previsão ou de classificação de forma progressiva. A cada nível da árvore, decisões são tomadas, as quais serão refletidas na estrutura seguinte, até atingir uma condição terminal. Quando uma árvore de decisão é utilizada para tarefas de classificação, o nome de referência do modelo é o de árvore de classificação, sendo esse o cenário de aplicação deste trabalho.

As árvores de classificação são formadas por nós, onde o primeiro nó é chamado de raiz e recebe o conjunto de dados de treino. A partir daí, os dados são particionados por nós intermediários, que representam um teste binário (verdadeiro, falso) sobre

o valor de uma variável específica. Para cada resultado possível de um nó, um novo ramo leva a estrutura para uma subárvore de decisão, e assim por diante, até alcançar uma folha, a qual representa uma determinada classe. A definição de uma folha se dá quando os dados são subdivididos entre os nós até alcançar um número mínimo pré-definido de amostras restantes.

Assim como a RNA, árvore de decisão é um método supervisionado, onde sua estrutura é gerada a partir de amostras de dados, e uma variável dependente é explicada através de  $n$  variáveis independentes. A Figura 3.4 mostra a estrutura de uma árvore de decisão, assim como a direção do fluxo dos dados conforme são processados.

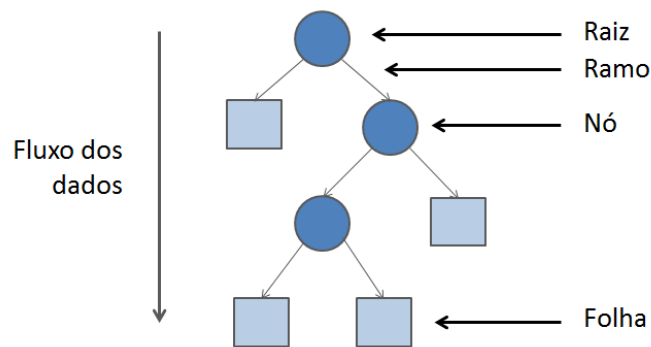


Figura 3.4: Ilustração de uma árvore de classificação, mostrando sua estrutura (raiz, nó, folha e ramos) e o fluxo da informação processada.

Uma das grandes vantagens de árvores de classificação é que a interpretação da sua estrutura e dos critérios de decisão é muito simples e clara, feita através de uma observação direta da árvore [49, 91]. Alia-se a isso a capacidade de absorver características complexas entre variáveis, e não apenas relações lineares, como acontece em boa parte dos procedimentos estatísticos.

### 3.2.2.1 Treinamento do modelo

As árvores de classificação foram desenvolvidas com base no algoritmo CART (*Classification And Regression Tree*), o qual visa realizar o particionamento do espaço disponível de forma recursiva, sendo representado pelos dados ao que se refere [19]. Dessa forma, a cada nível da árvore, um problema mais complexo de classificação é decomposto em subproblema mais simples, onde os dados se tornam cada



vez mais particularizados em termos de uma variável explicativa. No CART, esse processo ocorre até que um determinado critério seja atingido, como com a definição de um número mínimo de elementos em uma folha, que impeça sua subdivisão em mais ramos. Neste trabalho foi definido um número mínimo de 3 amostras devido a limitação no total de pacientes em cada cenário de triagem proposto.

O processo de geração da árvore é realizado utilizando o coeficiente Gini [92] como critério a ser maximizado em cada partição dos dados. As variáveis faltantes que correspondem a um nó de decisão são classificadas através da aplicação do método de substituição de variáveis [93]. Desse modo, para cada nó de decisão onde a variável não seja definida, todo o método de repartição dos dados é reaplicado com o objetivo de prever as condições do nó através das outras variáveis, e os erros de classificação são anotados. Assim, qualquer paciente que não possua o valor da variável definida é então classificado pela variável que obteve o menor erro nesse processo. Caso ela também não seja definida, a variável com o segundo menor erro é escolhida, e assim por diante.

Outra questão importante é que uma árvore pode crescer até o ponto onde sua estrutura não generalize mais a classificação, até pela definição do número mínimo de elementos em uma folha. Isso pode gerar um problema de *overfitting*, por a árvore ter mais estrutura do que o justificável. Além disso, sua interpretabilidade pode ser prejudicada por um excesso de nós de decisão [94]. Para mitigar essa situação, foi aplicado o processo de poda, o qual consiste em eliminar as partições que menos contribuem para a explicação da variável de resposta [95]. Com isso, cada nó é validado na tentativa de identificar árvores menos complexas e com taxas de erro menores, ou equiparáveis à original, ao classificar pacientes não utilizados no treinamento, ou seja, de teste. No caso de um nó ser considerado dispensável, ele é eliminado do modelo.

### **3.2.2.2 O *boosting* e o algoritmo AdaBoost**

No CART é possível, com o método de poda, reduzir consideravelmente a complexidade (e portanto ajudar na interpretabilidade) de uma árvore de classificação.

Porém, seu efeito na melhora de desempenho do modelo é limitado [96]. Essa melhoria pode ser alcançada ao conseguirmos agregar o resultado de vários modelos, desenvolvidos para resolver um mesmo problema, de forma que o seu conjunto consiga ser mais eficiente do que apenas um só. Essa técnica é chamada de *ensemble* [53]). Ainda, desejamos que os modelos possam ser capazes de melhorar as ineficiências um do outro. Para isso existem métodos de *boosting*, nos quais diferentes modelos são desenvolvidos sequencialmente, com o mais recente dando preferência a melhorar as deficiências encontradas no resultado do modelo anterior [97].

Dentro desse cenário, uma das aplicações do *boosting* é na melhora do desempenho utilizando modelos de fraca aprendizagem, ou seja, modelos que tenham estruturas muito simples e que são altamente enviesadas. Para isso, baseia-se na premissa de que cada um desses modelos terá um desempenho melhor do que 50%, sendo essa considerada a condição do fraco aprendizado. Essa premissa é intrínseca a definição matemática do *boosting*, que em conjunto com uma quantidade de dados suficiente, espera-se conseguir produzir um resultado com um erro de generalização menor [98]. Assim, o resultado final da classificação é obtido através de uma média ponderada sobre os resultados de todos os modelos gerados.

Um dos primeiros e mais utilizados algoritmos de *boosting* para classificação é o AdaBoost [20]. Nele, a principal ideia é dar foco, ao desenvolver cada modelo, na correção dos erros de classificação encontrados com o modelo anterior. Isso pode ser realizado ao se replicar os dados de treino que foram erroneamente classificados para um próximo treinamento, ou então, atribuindo-se pesos maiores aos mesmos. Dessa forma, o algoritmo força o modelo a tentar o aprendizado dessas instâncias que apresentaram uma maior dificuldade em seu aprendizado. Após essa etapa, os resultados de cada modelo são ponderados por pesos que refletem o desempenho alcançado no conjunto de treino por cada modelo, e uma votação por maioria simples decide qual a classificação final. O pseudo-algoritmo do AdaBoost pode ser visto em Algoritmo 1.

Além da busca pela melhora do desempenho pelos modelos, sua aplicabilidade em conjunto com CART, vista na literatura, foi também uma motivação para a

---

**Algoritmo 1:** Pseudo-código para o algoritmo AdaBoost

---

**início**Dado:  $(x_1, y_1), \dots, (x_m, y_m)$  onde  $x_i \in \mathcal{X}, y_i \in \{-1, +1\}$ .Inicializa:  $D_1(i) = 1/m$  para  $i = 1, \dots, m$ .**para**  $t = 1$  até  $T$  **faça**- Treina o modelo de fraco aprendizado utilizando a distribuição  $D_t$ - Obtém um resultado fraco  $h_t : \mathcal{X} \rightarrow \{-1, +1\}$ .- Objetivo: selecionar  $h_t$  com menor erro após ponderação:

$$\varepsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$$

- Escolher  $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ - Atualizar, para  $i = 1, \dots, m$ :

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

onde  $Z_t$  é um fator de normalização (escolhido para que  $D_{t+1}$  seja uma distribuição).**fim**

O resultado final será:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

**fim**

---

sua aplicação neste trabalho. No AdaBoost, modelos de fraca aprendizagem são comumente utilizados para se evitar o *overfitting*, o que no contexto de árvores de decisão implicou na utilização de modelos de baixa complexidade. Uma das estruturas mais utilizadas, e também aqui aplicadas, é a chamada de "*decision stumps*", onde o modelo possui apenas um nó de decisão. A geração desse tipo de modelo é realizada pela definição de critérios na parada do treinamento. Aqui, é definido um número máximo de 30 modelos para formar a arquitetura do *ensemble*. A Figura 3.5 mostra a arquitetura do algoritmo AdaBoost utilizando modelos de fraca aprendizagem baseados em CART. Como desvantagem de se utilizar esse método esta a perda obtida sobre a interpretabilidade do modelo, uma vez que o resultado final envolverá diversos modelos CART ao invés de uma única estrutura.

### 3.3 Avaliação do desempenho

Após a geração dos modelos computacionais, o desempenho é avaliado a partir da propagação dos dados de teste. Além da sensibilidade (capacidade de detectar pacientes que realmente possuem a doença - verdadeiros positivos - dentre os diagnosticados com a doença), que é prioritária quando se trata da triagem de pacientes,

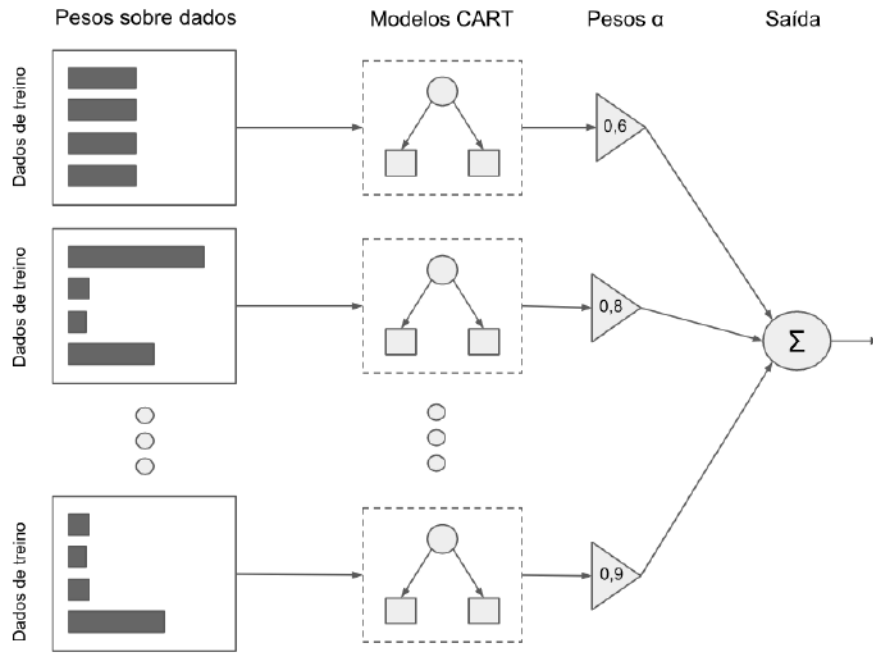


Figura 3.5: Arquitetura de *ensemble* e *boosting* adotado pelo algoritmo AdaBoost. Nele, as amostras que geram mais dificuldade de aprendizado recebem pesos maiores na geração do modelo CART, o qual contém apenas um nó de decisão (modelos de fraco aprendizado). Os resultados de cada modelo são então ponderados e somados para gerarem a classificação final.

outras estatísticas também são utilizadas para medir o desempenho, como a acurácia, ou seja, a taxa de classificação correta atingida pela aplicação do conjunto de teste, e a especificidade (capacidade de detectar pacientes que não possuem a doença - verdadeiros negativos - dentre os negativados pelo diagnóstico). Este trabalho também utiliza como critério o índice SP [99], cujo cálculo é definido por

$$SP = \sqrt{\frac{S + E}{2}} \times \sqrt{S \times E} \quad (3.1)$$

onde S representa a sensibilidade e E a especificidade. Esse índice visa ser uma medida que apoie o desenvolvimento de um classificador balanceado entre sensibilidade e especificidade [100], já que o seu valor tende a colapsar caso um dos dois valores que o compõe sofra uma redução significativa. Dessa forma, apesar de se buscar um modelo de alta sensibilidade, o índice SP é aqui utilizado para se avaliar a qualidade da triagem realizada. Todos esses índices são representados pelos seus valores médios e suas respectivas incertezas (valor RMS), obtidas através da validação cruzada, sendo esses os valores que refletem o desempenho esperado pelo modelo.

O índice SP é ainda utilizado na avaliação das RNA para apoiar a seleção da melhor topologia, ou seja, o número de neurônios na camada escondida a ser utilizado. Nesse caso, a topologia que apresenta a melhor média do índice SP, ao se propagar o conjunto de teste, é a escolhida como sendo a melhor configuração [71].

Já para facilitar a aplicação do modelo na unidade de saúde, é selecionado um único modelo de operação dentre todos os que são gerados pela validação cruzada. Esse procedimento é realizado por meio da propagação dos dados de treino, em conjunto com teste. Assim, o modelo de operação pode ser selecionado como o que alcançou o desempenho mais alto considerando algum dos critérios de avaliação. Aqui, o índice SP é utilizado, tanto com as MLP quanto CART. Nos dois casos, a seleção é feita considerando o modelo que apresentou o maior índice SP ao se propagar todos os dados disponíveis.

O valor do ponto de corte que definirá a sensibilidade e a especificidade é selecionado, a princípio, de forma a otimizar o valor do índice de SP. Valores de classificação acima do corte definido apontam pacientes classificados com o diagnóstico alvo (teste positivo, paciente com a doença), e valores abaixo desse limiar representam o resultado oposto (teste negativo, ausência de doença), como visto na Figura 3.6. Os resultados de cada índice e o ponto de corte escolhido devem ser avaliados de acordo com o objetivo final da rede.

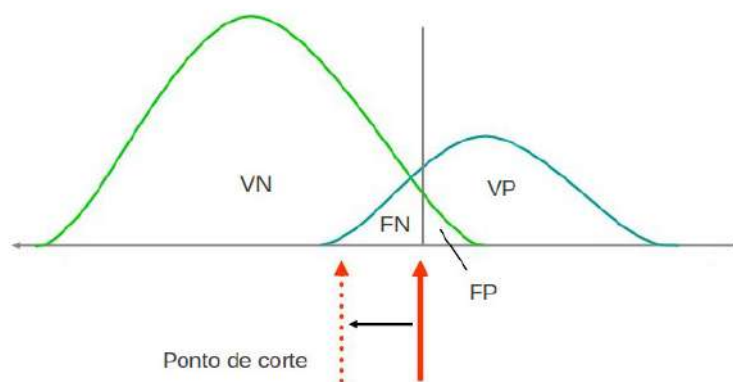


Figura 3.6: Distribuição de pacientes com diagnósticos positivos e negativos e exemplo de ponto de corte para definição da sensibilidade e especificidade. Regiões de verdadeiros negativos (VN), falsos negativos (FN), verdadeiros positivos (VP) e falsos positivos (FP) são indicadas.

Um modo de se analisar a relação entre sensibilidade e especificidade de um

modelo é através da curva ROC (*Receiver Operating Characteristic*) [101][102], a qual permite a visualização dos pontos de corte para os quais existe maior otimização entre esses dois critérios. Para cada ponto de corte um índice SP percentual é calculado, e o seu valor máximo indica o ponto onde a eficiência é alta. Na avaliação da curva ROC, ainda é calculada a área sob a curva (AUC), sendo o seu valor correspondente a sua integral medida em pontos discretos por meio da regra dos trapézios. Um exemplo de curva ROC pode ser visto na Figura 3.7, onde também é indicado o ponto relativo ao valor máximo de SP.

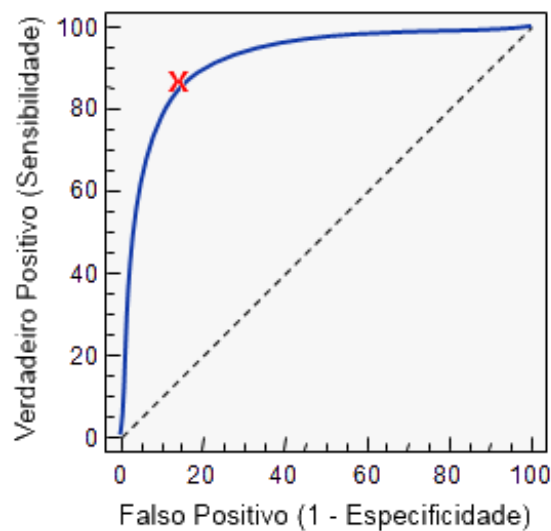


Figura 3.7: Exemplo de curva ROC para medir a relação entre sensibilidade e falsos positivos, indicando o ponto que representa o índice SP ótimo.

O Valor Preditivo Positivo (VPP), o qual é a probabilidade de um indivíduo ter uma determinada doença dado que ele é positivo a um teste diagnóstico, e o Valor Preditivo Negativo (VPN), que é definido como a probabilidade de um indivíduo não ser doente dado que tem teste diagnóstico negativo, também são avaliados. Dessa forma, quanto maior a sensibilidade, maior deve ser o VPN, isto é, maior será a probabilidade de, perante um resultado negativo, não haver doença. Por outro lado, quanto maior a especificidade, maior será o VPP, ou seja, maior será a probabilidade de, ao se obter um resultado positivo, haver doença. Considerando a prevalência da doença em uma população, quanto maior o seu valor, maior será o VPP e menor será o VPN. Portanto, quanto mais frequente é uma doença, mais provável é encontrar verdadeiros positivos (aumentando o VPP), mas também é mais

provável encontrar falsos negativos (VPN). De forma geral, tanto o VPP quanto o VPN conseguem fornecer um indicativo de quantos pacientes serão encaminhados para um determinado cenário em detrimento do outro. Os índices de VPP e VPN são definidos, através da definição do valor de prevalência ( $P$ ), como [103]:

$$VPP = \frac{S \times P}{S \times P + (1 - E) \times (1 - P)} \quad (3.2)$$

$$VPN = \frac{E \times (1 - P)}{(1 - S) \times P + E \times (1 - P)} \quad (3.3)$$

Para o cálculo desses índices, a prevalência da doença na população é analisada sobre o intervalo de 2% a 80%, sendo esses valores definidos com base em resultados observados na literatura. Micheletti, et al, aponta que no Rio Grande do Sul, considerando pacientes diagnosticados TB sem tratamento anterior, a prevalência de TB-DR e TB-MDR é de 8,5% e 2,2%, respectivamente [104]. No Rio de Janeiro, Brito, et al, descreve, também para pacientes sem tratamento anterior, valores de prevalência para TB-DR de 13,1% e para TB-MDR de 3,9%, e para pacientes que apresentam histórico de tratamento, números que chegam a 28,2% para TB-DR, e 17,3% para TB-MDR [105]. Nacionalmente, são estimados 1,5% de novos casos de TB-DR/MDR, além de uma taxa de 8% para casos que já passaram por tratamento [6]. Já para o estado do Ceará, não foram encontrados dados recentes sobre prevalência.

Os modelos computacionais podem ser aplicados nas unidades de saúde utilizando os critérios que melhor se adaptam a realidade do cenário em questão. Por exemplo, para aumentar o número de verdadeiros positivos, um modelo que possua uma taxa de sensibilidade elevada é o que se torna mais aderente a esse requisito. Por outro lado, se quisermos diminuir o número de diagnósticos falsos positivos e, com isso, evitar o encaminhamento de pacientes para tratamentos desnecessários (e seus efeitos colaterais devido a medicação), um modelo com foco em especificidade deve ser escolhido. Ainda, é possível utilizar ambos os modelos em um processo de triagem em etapas: pacientes diagnosticados pelo modelo de alta sensibilidade

são encaminhados para uma próxima etapa onde encontram o modelo de alta especificidade. Essa estratégia pode ser utilizada para maximizar a identificação de possíveis casos de TB-DR, e em seguida minimizar falsos casos de TB-MDR. Já uma outra opção é buscar um modelo que otimize o valor do índice SP, o que gera um resultado que visa balancear os resultados entre sensibilidade e especificidade, ambos índices de desempenhos importantes para o estudo. Portanto, essa avaliação utilização desse tipo de estratégia é realizado nesse trabalho. Para isso, é observado que na literatura é indicado que um nível mínimo de sensibilidade para triagem seria de 80%, quando comparado com a cultura [106].

### 3.4 Análise de relevância

Para um maior entendimento sobre o papel das variáveis em cada modelo, e qual o nível da sua influência para o diagnóstico, uma análise de relevância é proposta. Nesse caso, a relevância de uma variável é definida pelo nível de sua contribuição no modelo computacional. Quanto mais a variável for influente para o resultado do score, mais relevante é a variável. Como os modelos CART resultantes já são gerados por um processo de seleção de características inerente ao algoritmo [107], as variáveis utilizadas como nós de decisão são interpretadas como relevantes para o diagnóstico correspondente, aliando-se a isso o quanto é decisiva na discriminação dos pacientes. Já nos modelos RNA o valor de relevância de cada variável é obtido ao se propagar todos os dados através das 50 redes neurais geradas pela validação cruzada e o modelo que apresenta o maior valor de índice SP é selecionado. Assim, dois métodos são aplicados com o objetivo de se calcular o valor de relevância: *Weights* [108] e *Relevance mapping* [109]. Um índice de relevância consolidado também é proposto, obtido ao se combinar os resultados dos dois métodos de relevância. Com isso, objetiva-se simplificar a avaliação de um especialista na análise das variáveis mais relevantes em cada diagnóstico, já que esse processo resulta em uma listagem unificada.



### 3.4.1 O método *Weights*

Garson [110] propôs um método que particiona os pesos das conexões de uma rede neural para determinar a importância relativa das diferentes variáveis de entrada. Essa ideia foi simplificada por Goh [111] e aplicada em diferentes trabalhos [108, 112, 113]. Com essa proposta, para cada neurônio da camada escondida  $h$  o valor absoluto do peso de uma conexão de entrada é dividido pelo somatório de todos os pesos das conexões de entrada daquele neurônio:

$$Q_{ih} = \frac{|W_{ih}|}{\sum_{i=1}^{n_i} |W_{ih}|} \quad (3.4)$$

A seguir, o valor  $Q_{ih}$  referente a cada variável de entrada  $i$  é dividido pelo somatório de todos os valores de  $Q_{ih}$ . Ao multiplicar esse resultado por 100, obtém-se a importância relativa (IR) de todos os pesos de uma dada variável:

$$IR(\%)_i = \frac{\sum_{h=1}^{n_h} Q_{ih}}{\sum_{h=1}^{n_h} \sum_{i=1}^{n_i} Q_{ih}} * 100 \quad (3.5)$$

Com isso, se torna possível ordenar as variáveis por sua relevância IR, o que corresponde ao seu nível de contribuição para o diagnóstico.

### 3.4.2 O método *Relevance mapping*

Para se obter o valor de relevância, define-se o valor de uma variável  $x$  como a sua média ( $\bar{x}$ ) sobre seu valor para todos os pacientes. Em seguida, os dados dos pacientes, já com o valor de  $x$  alterado, são propagados pelo modelo neural e o impacto causado no desempenho de classificação é avaliado. Esse impacto é medido pelo valor da diferença entre o índice SP obtido com o resultado dessa alteração e com o valor original [109].

$$Relevance_x = SP|_{x=\bar{x}} - SP|_x \quad (3.6)$$

Com esses resultados é possível ordenar as variáveis pela sua importância ao modelo. Além disso, ao acompanhar o efeito de cada variável no desempenho fi-

nal do modelo, é possível identificar variáveis confundidoras, ou seja, que estejam diminuindo o índice SP. Esse método simplifica o método clássico *stepwise*, o qual consistiria em adicionar ou rejeitar uma variável antes da geração do modelo, realizar todo o processo de treinamento, para então analisar seu efeito na saída da rede [114].

### 3.4.3 Índice de relevância consolidado

Para unificar as relevâncias obtidas por cada variável nos métodos *Weights* e *Relevance mapping*, a geração de um índice consolidado é proposto. Esse processo consiste em normalizar os valores obtidos por cada método de relevância entre 0 e 1, para então somar os seus resultados, resultando no índice consolidado. Dessa forma, é possível apontar não só quais variáveis são mais relevantes em cada método, mas o quanto influenciam na triagem de forma conjunta, explorando topologias e o seu poder de discriminação. Assim, pequenas disparidades podem ser minimizadas ao se realizar o processo de consolidação, pois um resultado que tenha ganho mais relevância por conta de algum viés de um determinado método, pode ser minimizado pelo resultado do outro. Por outro lado, uma variável relevante nos dois métodos fica ainda mais proeminente quando o resultado é agregado. Ao final, por aplicar mais de um método de cálculo, os resultados podem ser comparados entre si em termos de coerência. A Figura 3.8 mostra um exemplo desse processo.

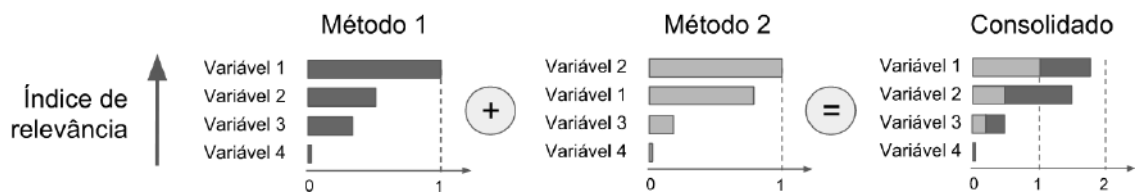


Figura 3.8: Exemplo de construção do índice de relevância consolidado a partir de dois métodos para o cálculo de relevância.

## 3.5 Geração de modelos de menor complexidade

Locais que apresentam uma menor disponibilidade de recursos podem não possuir toda a informação considerada necessária para o funcionamento dos modelos

computacionais desenvolvidos. Isso pode acontecer seja por falta de equipamento especializado para exame, profissionais de saúde qualificados, unidades de saúde com infraestrutura deficiente e/ou problemas de qualidade dos dados. Dessa forma, com o objetivo de investigar alternativas para possibilitar a aplicação dos modelos também nessas localidades, o impacto de uma redução das variáveis disponíveis no desempenho é analisado. Isso é realizado através da definição de estratégias que visam remover a necessidade de fornecer informações ao modelo, levando em consideração questões relacionadas a qualidade de dados. Dessa forma, é possível remover, de forma incremental, as informações de menor relevância em conjunto com definições de qualidade, analisando a viabilidade do modelo resultante.

Assim, a partir da ordenação das variáveis pelo valor de sua relevância nos modelos de RNA, diferentes subconjuntos de variáveis podem ser selecionados pela aplicação de estratégias que seguem critérios de corte incrementais, definidas como *loose*, *medium* e *tight*. A estratégia *tight* visa reduzir ao máximo o número de variáveis para apoiar a triagem em localidades onde poucos recursos estão disponíveis, e que a quantidade de informações necessárias pelo sistema de saúde possa ser significativamente reduzida. Quando poucas variáveis são necessárias, não só a complexidade do modelo diminui, mas também se torna mais resiliente à qualidade dos dados, incluindo a própria falta de dados. Por outro lado, ao se aplicar a estratégia *loose*, é possível remover as variáveis que não se mostram tão importantes e que podem até mesmo serem identificadas como prejudiciais ao processo de discriminação (variáveis confundidoras). Por fim, a estratégia *medium* age como um critério intermediário entre o *tight* e *loose*, tentando equilibrar a simplicidade do modelo com a eficiência de desempenho. Após as variáveis serem selecionadas seguindo uma das estratégias definidas, todo o procedimento de geração dos modelos neurais é repetido com esse subconjunto dos dados, e o desempenho resultante é comparado ao modelo original (38 variáveis).

# Capítulo 4

## Resultados e Discussão

Neste capítulo são apresentados os resultados obtidos e a discussão, além de uma análise sobre a base de dados disponível. O desempenho alcançado pelos modelos computacionais é descrito, detalhando os critérios de avaliação por diagnóstico, a técnica utilizada (RNA ou CART), assim como os resultados obtidos pelo algoritmo de AdaBoost. A qualidade do treinamento dos modelos é analisada, e as diferentes abordagens (nacional ou por estado) comparadas. A relevância das variáveis em cada estratégia proposta para o desenvolvimento de modelos de menor complexidade é discutida, assim como a influência da prevalência da doença na população por meio das proporções de VPP e VPN.

### 4.1 Base de Dados

Neste trabalho utilizou-se registros de pacientes suspeitos de TB-DR, avaliados num ensaio clínico pragmático (Registro N. RBR-4rprbd), atendidos em 5 unidades de saúde de referência em diferentes estados do Brasil: pacientes internados nos Instituto Estadual Ary Parreiras (Rio de Janeiro/RJ), Hospital Messejana (Fortaleza/CE), Hospital Partenon (Porto Alegre/RS); e pacientes ambulatoriais do Centro de Referência Hélio Fraga (Rio de Janeiro/RJ) e do Instituto Clemente Ferreira (São Paulo/SP). Esses estados foram escolhidos por decisão de projeto em conjunto com o Ministério da Saúde, de forma a priorizar recursos em locais com maior prevalência de TB-DR. Outro critério de escolha utilizado foi o acesso a avaliação pelo meio de

cultura líquido *Mycobacteria growth indicator tube* (MGIT). Dessa forma, estados como Amazonas e Bahia, que também possuem alta prevalência de TB, não foram incluídos.

Os dados dos pacientes foram coletados de fevereiro de 2011 a maio de 2013 e foram incluídos pacientes maiores de 18 anos apresentando tosse por 3 semanas ou mais. Além disso, de acordo com diretrizes nacionais [21], ter a presença de, ao menos, uma das seguintes condições clínico-sociais que os definem como suspeitos de TB-DR: a) Tratamento anterior de TB: histórico de falha em tratamento ou mais de dois tratamentos padrão; b) sem tratamento anterior de TB: portadores do vírus da AIDS, pacientes que tiveram contato próximo com casos positivos de TB-MDR, moradores de rua ou hospitalizados em centros de referência em TB (ambulatório ou enfermarias). Participantes foram excluídos por: a) não possuírem confirmação bacteriológica apontando ou não a presença da doença, b) presença de micobactérias atípicas, c) sem resultados clínicos ou laboratoriais disponíveis nos registros médicos. Um total de 28 pacientes com tratamento inicializado há mais de 30 dias e diagnóstico negativo foram incluídos como participantes, mas não foram considerados no desenvolvimento do escore proposto, pois seus sintomas clínicos já poderiam ter sido descaracterizados pela medicação. As ocorrências desses casos estavam distribuídos dentre os estados da seguintes forma: 4 oriundos do RJ, 10 do RS, 5 no CE, 8 em SP, além de um 1 de Minas Gerais. No estado de SP esse volume de pacientes acabou representando dois terços dos dados totais coletados relativos a diagnósticos negativos para TB.

Destaca-se que no processo de coleta dos dados, os pacientes foram avaliados prospectivamente por meio de um formulário padrão por profissionais de saúde treinados, que não possuíam conhecimento do diagnóstico final. Algumas limitações nas análises também podem ser ressaltadas. Não é possível avaliar a utilização do escore clínico nos primeiro e segundo níveis de atenção à saúde, que são onde as decisões de encaminhamento desses pacientes são feitas. Isso acontece pelas características dos dados, que já são de pacientes que foram encaminhados para um centro de referência. Da mesma forma, poucos casos de TB com HIV positivo foram incluídos

na análise, devido a baixa prevalência dessa co-infecção na base de dados.

Os dados consolidados representam os 502 pacientes que tiveram diagnósticos confirmados de TB-DR/MDR (206), TB-DS (76) e não TB (220). Apesar dos dados serem relativos a pacientes encaminhados com provável TB-DR, comparativamente existe uma grande quantidade de casos onde a TB não foi confirmada. Essa situação pode indicar uma falha no processo de triagem nas Unidades Básicas de Saúde (primeiro e segundo nível de atenção básica) e no posterior encaminhamento para as unidades de referência. Observa-se ainda que dos 220 pacientes diagnosticados como não TB, 14 (RJ-7, CE-2, RS-3, SP-2) haviam iniciado tratamento há menos de 30 dias. Essa situação corrobora com uma possível suspeita de falha do processo, pois caso houvesse uma forte suspeita de TB, o tratamento empírico (por probabilidade) deveria ter sido iniciado antes da sua confirmação pelo resultado laboratorial. Dessa forma, essa característica da base de dados é uma das limitações do presente estudo.

O desenvolvimento de escores clínicos foi direcionado para o suporte de diagnósticos de resistência a medicamentos, sendo a classificação deles realizada entre: TB-DR/MDR e não TB-DR/MDR (TB-DS + não TB), TB-DR/MDR e TB-DS, e TB-MDR de TB-DR. A Figura 4.1 exibe um esquemático dos diagnósticos de TB pela complexidade de tratamento. Como o banco de dados não apresenta casos de TB-XDR, nenhum estudo foi realizado nesse sentido.

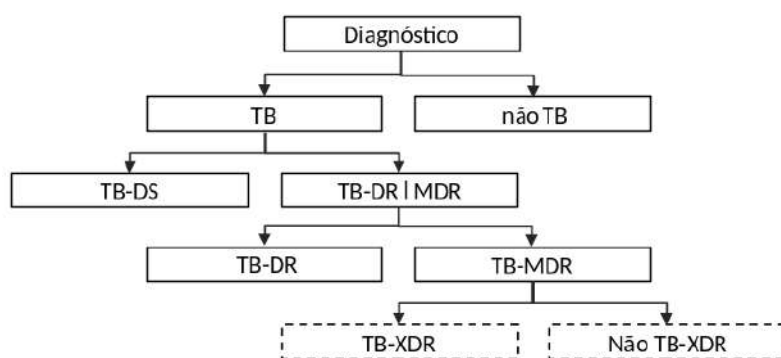


Figura 4.1: Diagnósticos de TB considerando a complexidade de tratamento. O estudo desenvolve escores clínicos para a classificação dos diagnósticos de TB-DR/MDR dentre todos os outros pacientes e de TB-DS, além de TB-MDR e TB-DR

Os desfechos da base de dados seguem as definições listadas a seguir:

- Não TB: Paciente apresentou resultado de exame de cultura negativo.

- TB-DS: Paciente apresentou resultado de exame de cultura positivo, porém sem resistência as drogas.
- TB-DR, não TB-MDR: Paciente apresentou resultado de cultura positivo, e teste de sensibilidade apontou resistência para drogas anti-TB, exceto a combinação de Rifampicina e Isoniazida.
- TB-MDR: Paciente apresentou resultado de cultura positivo, e teste de sensibilidade apontou resistência para drogas anti-TB, incluindo a combinação de Rifampicina e Isoniazida.

Dentre os estados, o que teve mais participantes incluídos foi o Rio de Janeiro (RJ - 276), seguido do Ceará (CE - 112) e do Rio Grande do Sul (RS - 82). Já pacientes de outros lugares somaram um total de 32 registros. É observado que, mesmo quando se reúnem dados de todos os estados, excluindo o RJ, o total de 226 registros obtidos se mostra inferior ao volume de dados advindos do RJ. Pacientes que foram triados nos centros de referência, mas que eram de outros estados, também tiveram dados incluídos no estudo. São identificados registros de Minas Gerais, Paraíba e Paraná, com apenas um paciente de cada lugar. A distribuição de pacientes, por regiões e diagnósticos, pode ser vista na Tabela 4.1.

Diagnóstico	CE	RJ	RS	SP	Outros estados	Não RJ	Nacional
Não TB-DR/MDR	54	190	42	7	3	106	296
Não TB	33	158	22	4	3	62	220
TB-DS	21	32	20	3	0	44	76
TB-DR/MDR	58	86	40	22	0	120	206
TB-DR	27	38	19	18	0	64	102
TB-MDR	31	48	21	4	0	56	104
Total	112	276	82	29	3	226	502

Tabela 4.1: Número de pacientes incluídos no estudo separados por regiões (RJ, CE, RS, outras localizações e todos as regiões exceto RJ) e diagnósticos de TB (Não TB, TB-DS, além de TB-DR e TB-MDR).

Nas Tabelas 4.2 e 4.3 são apresentadas todas as variáveis que caracterizam os pacientes, com estatísticas sobre os dados para os diagnósticos de TB-DR/MDR

e não TB-DR/MDR, considerando a abordagem nacional e para o estado do RJ. Para a geração dos modelos computacionais, as variáveis binárias são codificadas como +1 para representar a presença de uma característica, e -1 para sua ausência. Se a informação foi ignorada na triagem, ou seja, não possui um valor definido, o valor 0 é atribuído. Esses valores correspondem aos alvos dos modelos, os quais são definidos para a fase de treinamento dos mesmos. Como o valor nulo não ativa os neurônios das RNA devido a função de ativação escolhida (tangente hiperbólica), dados faltantes acabam não influenciando o treino do modelo para realizar a discriminação do diagnóstico. Ao atribuir o valor de +1 à variável, a classificação tende a um diagnóstico positivo, enquanto quando utilizado o valor -1 o efeito esperado é o contrário. Portanto, a escolha de atribuir esses valores aos dados foi feita com o objetivo de apoiar os modelos de RNA com conhecimento especialista. Já as variáveis nominais, como estado civil e desfecho do último tratamento, são expandidas em diferentes características para então serem codificadas como variáveis binárias. Por exemplo, a variável que trata do desfecho do último tratamento foi dividida como o desfecho sendo de cura, abandono, entre outras, onde para todas foram atribuídos os valores de sim ou não. Variáveis numéricas, como valores referentes a quantidade de tempo, são normalizadas para terem média zero e variância unitária, de forma a lidar com o intervalo dinâmico da função de ativação dos neurônios das RNA (tangente hiperbólica).

Após as codificações descritas acima, um total de 38 variáveis passam a constituir a base de dados. Ressalta-se que a seleção das variáveis utilizadas neste trabalho foi realizada de forma a contemplar ao máximo as informações descritas na literatura médica como importantes para a caracterização de TB, assim como o relatado em trabalhos anteriores que desenvolvem escores clínicos correspondentes através de modelos computacionais. Essas variáveis podem ser subdivididas em informações sobre características gerais do paciente (idade, delta IMC - variação do peso sobre a altura elevada ao quadrado do paciente, gênero, estado civil, cor da pele, fumante/ex-fumante, alcoolismo), sintomas (emagrecimento, tosse, expectoração, escarro com sangue, suor noturno, febre, perda de apetite, falta de ar), histórico (há quanto



tempo tem tosse/possui escarro, quantas vezes tratou TB, tratamento anterior para TB, desfecho do tratamento anterior, estadia em prisão ou delegacia, contato anterior com TB-DR/DS) e resultados de exames (testagem anti-HIV/AIDS e cavitação pulmonar).

Em relação a qualidade de dados, as Tabelas 4.2 e 4.3 também indicam as variáveis que podem apresentar problemas relacionados a falta de infraestrutura e acessibilidade, assim como a validade da resposta obtida. Algumas das variáveis vistas na literatura como marcadoras de TB-DR exigem equipamento de raio X (para detecção de cavitação pulmonar) ou para medir peso e altura (obter o delta IMC), além de exames laboratoriais para doenças sexualmente transmissíveis, como o HIV/AIDS. Ainda, a unidade de saúde pode não ter tido acesso aos resultados do desfecho de tratamentos anteriores. Ressalta-se que no período da coleta dos dados ainda não existia o Sistema Gerenciador de Ambiente Laboratorial (GAL), desenvolvido posteriormente para informatizar o Sistema Nacional de Laboratórios de Saúde Pública, facilitando a distribuição de informação para as Unidades de Saúde. Entretanto, mesmo hoje em dia, obter a informação do desfecho de tratamento anterior ainda implica em infraestrutura, pois é necessário um computador conectado à internet para se ter o acesso necessário. Existe também o caso onde o paciente precisa ser encaminhado para outra unidade para exames adicionais, e não o faz se não tiver uma supervisão de um profissional de saúde. Dessa forma, essas situações acabam gerando dados faltantes nas amostras coletadas.

Por outro lado, mesmo havendo dados, estes podem não ser válidas por serem fornecidos pelo paciente sem confirmação por outra fonte. Assim, as respostas obtidas podem acabar sendo enganosas, seja por viés de memória, ou influência de aspectos culturais e sociais. Dentre as informações que se encaixam nesse perfil estão o tempo em que a pessoa apresenta um determinado sintoma, se já teve contato com alguém que possuía TB, condição de fumante e estadia em delegacia e instituição prisional. Portanto, quando aplicável, as variáveis são rotuladas com essas características de qualidade de dados, as quais serão consideradas na análise de relevância e no desenvolvimento de modelos menos complexos.

Tabela 4.2: Lista de todas variáveis, e suas respectivas estatísticas em termos de resposta, para o diagnóstico de TB-DR/MDR, considerando a abordagem nacional. Os dados são expressos como o total de pacientes (% do total) ou a média  $\pm$  RMS (um valor de raiz média quadrática da variância). Possíveis problemas em qualidade de dados são indicados, como: (1) falta de infraestrutura e/ou problemas relacionados a acessibilidade da informação e (2) validade da resposta do paciente.

Variáveis	TB-DR/MDR (N=206)			Não TB-DR/MDR (N=296)		
	Média $\pm$ DP			Média $\pm$ DP		
Idade	39.174 $\pm$ 11.995			42.969 $\pm$ 13.869		
Delta IMC <sup>(1)</sup>	2.816 $\pm$ 3.053			2.507 $\pm$ 3.170		
Tempo tem tosse <sup>(2)</sup>	3.737 $\pm$ 4.910			4.722 $\pm$ 9.923		
Tempo possui escarro <sup>(2)</sup>	3.082 $\pm$ 3.515			4.070 $\pm$ 9.323		
Quantas vezes tratou TB <sup>(2)</sup>	1.558 $\pm$ 1.270			0.836 $\pm$ 1.167		
Gênero	Masc.	Fem.	Ign	Masc.	Fem.	Ign
	140(68)	66(32)	-	174(59)	122(41)	-
Estado civil	Sim	Não	Ign	Sim	Não	Ign
	-	-	1(1)	-	-	0(0)
Solteiro	113(55)	92(44)	-	190(64)	106(34)	-
Casado	36(18)	169(82)	-	42(14)	254(86)	-
Amigado	23(11)	182(88)	-	27(9)	269(91)	-
Divorciado	9(4)	196(95)	-	9(3)	287(97)	-
Separado	17(8)	188(91)	-	17(6)	279(94)	-
Viúvo	7(3)	198(96)	-	11(4)	285(96)	-
Cor da pele	-	-	0(0)	-	-	0(0)
Branco	73(35)	133(65)	-	86(29)	209(71)	-
Negro	47(23)	159(77)	-	89(30)	206(70)	-
Amarelo	2(1)	204(99)	-	3(1)	292(99)	-
Pardo	82(40)	124(60)	-	113(38)	182(62)	-
Indígena	2(1)	204(99)	-	4(1)	291(98)	-
Emagrecimento	140(68)	64(31)	2(1)	143(48)	152(51)	1(1)
Tosse	191(93)	12(6)	3(1)	216(73)	78(26)	2(1)
Expectoração	175(85)	30(14)	1(1)	192(65)	101(34)	3(1)
Catarro com sangue	61(29)	144(70)	1(1)	47(16)	244(82)	5(2)
Suor noturno	129(63)	70(34)	7(3)	133(45)	152(51)	11(4)
Febre	135(66)	67(32)	4(2)	125(42)	165(56)	6(2)
Perda de apetite	138(67)	66(32)	2(1)	142(48)	150(51)	4(1)
Falta de ar	110(53)	91(44)	5(3)	163(55)	126(43)	7(2)
Tratamento anterior TB <sup>(2)</sup>	157(76)	44(21)	5(3)	141(47)	153(52)	2(1)
Desfecho tt. anterior <sup>(1)</sup>	-	-	49(24)	-	-	156(53)
Cura	36(17)	121(59)	-	54(18)	86(29)	-
Completo	5(2)	152(74)	-	13(4)	127(43)	-
Abandono	43(21)	114(55)	-	57(19)	83(28)	-
Falência	71(34)	86(42)	-	14(5)	126(43)	-
Prisão ou delegacia <sup>(2)</sup>	31(15)	174(84)	1(1)	57(19)	234(79)	5(2)
Contato TB-DS <sup>(2)</sup>	83(40)	71(34)	52(26)	162(55)	90(30)	44(15)
Contato TB-DR <sup>(2)</sup>	39(19)	59(29)	108(52)	19(6)	113(38)	132(56)
Fumante <sup>(2)</sup>	76(37)	128(62)	2(1)	156(53)	138(46)	2(1)
Ex-fumante <sup>(2)</sup>	65(32)	139(67)	2(1)	67(23)	227(76)	2(1)
Cavitação pulmonar <sup>(1)</sup>	172(83)	27(13)	7(3)	107(36)	56(19)	133(45)
Alcoolismo	40(19)	166(81)	0	38(13)	258(87)	0(0)
HIV/AIDS <sup>(1)</sup>	25(12)	181(88)	0	58(20)	238(80)	0(0)

Tabela 4.3: Lista de todas variáveis, e suas respectivas estatísticas em termos de resposta, para o diagnóstico de TB-DR/MDR, considerando pacientes do RJ. Os dados são expressos como o total de pacientes (% do total) ou a média  $\pm$  RMS (um valor de raiz média quadrática da variância). Possíveis problemas em qualidade de dados são indicados, como: (1) falta de infraestrutura e/ou problemas relacionados a acessibilidade da informação e (2) validade da resposta do paciente.

Variáveis	TB-DR/MDR (N=86)			Não TB-DR/MDR (N=190)		
	Média $\pm$ DP			Média $\pm$ DP		
Idade	40.325 $\pm$ 12.528			42.716 $\pm$ 14.378		
Delta IMC <sup>(1)</sup>	3.017 $\pm$ 3.582			2.756 $\pm$ 3.681		
Tempo tem tosse <sup>(2)</sup>	2.673 $\pm$ 2.998			6.770 $\pm$ 13.970		
Tempo possui escarro <sup>(2)</sup>	2.380 $\pm$ 2.615			5.829 $\pm$ 13.737		
Quantas vezes tratou TB <sup>(2)</sup>	1.525 $\pm$ 1.262			0.508 $\pm$ 0.882		
Gênero	Masc. 62 (72)	Fem. 24(28)	Ign 0(0)	Masc. 113(59)	Fem. 77(41)	Ign 0(0)
Estado civil	Sim	Não	Ign 0(0)	Sim	Não	Ign 0(0)
Solteiro	45(52)	41(48)	-	125(66)	65(34)	-
Casado	19(22)	67(78)	-	16(8)	174(92)	-
Amigado	10(12)	76(88)	-	22(12)	168(88)	-
Divorciado	2(2)	84(98)	-	9(5)	181(95)	-
Separado	8(9)	78(91)	-	14(7)	176(93)	-
Viúvo	2(2)	84(98)	-	4(2)	186(96)	-
Cor da pele			0(0)			1(1)
Branco	19(22)	67(78)	-	50(26)	139(73)	-
Negro	28(33)	58(67)	-	67(35)	122(64)	-
Amarelo	0(0)	86(100)	-	1(1)	188(98)	-
Pardo	39(45)	47(55)	-	70(36)	119(63)	-
Indígena	0(0)	86(100)	-	1(1)	188(98)	-
Emagrecimento	63(73)	22(26)	1(1)	83(44)	106(55)	1(1)
Tosse	79(92)	6(7)	1(1)	128(67)	61(32)	1(1)
Expectoração	70(81)	16(19)	0(0)	104(55)	84(44)	2(1)
Escarro com sangue	27(31)	59(69)	0(0)	17(9)	170(89)	3(1)
Suor noturno	51(59)	33(38)	2(1)	71(37)	114(60)	5(3)
Febre	55(64)	30(35)	1(1)	68(36)	119(63)	3(1)
Perda de apetite	55(64)	31(36)	0(0)	62(33)	125(66)	3(1)
Falta de ar	54(63)	31(36)	1(1)	85(45)	102(54)	3(1)
Tratamento anterior TB <sup>(2)</sup>	64(74)	20(23)	2(3)	58(31)	130(68)	2(1)
Desfecho tt. anterior <sup>(1)</sup>			21(24)			133(70)
Cura	13(15)	52(61)	-	17(9)	40(21)	-
Completo	4(5)	61(71)	-	8(4)	49(26)	-
Abandono	20(23)	45(52)	-	27(14)	30(16)	-
Falência	27(31)	38(44)	-	3(2)	54(28)	-
Prisão ou delegacia <sup>(2)</sup>	8(9)	77(90)	1(1)	44(23)	142(75)	4(2)
Contato TB-DS <sup>(2)</sup>	28(33)	27(31)	31(36)	123(65)	42(22)	25(13)
Contato TB-DR <sup>(2)</sup>	9(10)	21(24)	56(66)	9(5)	62(33)	119(52)
Fumante <sup>(2)</sup>	32(37)	53(62)	1(1)	115(60)	74(39)	1(1)
Ex-fumante <sup>(2)</sup>	19(22)	66(77)	1(1)	31(16)	158(83)	1(1)
Cavitação <sup>(1)</sup>	75(87)	8(9)	3(4)	41(22)	34(18)	115(60)
Alcoolismo	10(12)	76(88)	0(0)	9(5)	181(95)	0(0)
HIV/AIDS <sup>(1)</sup>	7(8)	79(92)	0(0)	34(18)	156(82)	0(0)

Dessa forma, são apontadas quatro dimensões possíveis em termos de qualidade de dados: infraestrutura, acessibilidade, validade e dados faltantes. Entretanto, vê-se como limitação a falta da coleta de informações, por meio de questionário, por exemplo, para avaliar o nível de acesso de cada variável, seja por estado ou centro de referência. O mesmo se aplica a validade e infraestrutura. Dessa forma, a única dimensão de qualidade possível de ser quantificada é o total de dados faltantes, ou completude dos dados.

Define-se como um nível ótimo de completude quando se tem a informação para mais de 90% dos dados, razoável entre 70% e 90% e ruim abaixo de 70% [115, 116]. Ao analisar a proporção de dados faltantes no cenário nacional, é possível observar que os sintomas possuem porcentagens que variam de 1 a até 4% do total de respostas cuja a informação não está disponível. Essa situação é melhor quando se analisa apenas o RJ, onde no máximo 1% das respostas não estão respondidas, podendo indicar um melhor processo na coleta de informações nesse estado. Já a falta de informação sobre a existência de tratamento anterior do paciente para TB alcança uma representatividade de até 3% do total, tanto nacionalmente, quanto no RJ. Nesses casos, é possível que o paciente não soubesse responder a pergunta, mas também pode apontar uma falha no processo de preenchimento dos dados. Já detalhes sobre o desfecho do tratamento anterior não são conhecidos em mais da metade dos pacientes ao se analisar os diagnósticos negativos para TB-DR, tanto para a abordagem nacional (53%), quanto para o RJ (70%). Dados sobre contato prévio com TB-DS e TB-DR também apresentam altas números de dados faltantes, até 66% para TB-DR no RJ e 52% nacionalmente. Já para o diagnóstico de não TB-DR/MDR esses números são de 56% na abordagem nacional, e 52% no RJ. Ainda, é possível observar que a variável que depende de uma maior infraestrutura, a cavitação pulmonar, apresenta um total representativo de respostas ignoradas quando o diagnóstico é negativo para TB, 60% no RJ e 45% em todos os dados. Essa situação pode indicar que, nas unidades de saúde, existe uma tendência de se realizar esse exame em pacientes que acabam confirmados como TB-DR. Dessa forma, avalia-se que a maioria das variáveis se encontra em uma situação de completude entre

ótima e razoável. Apenas 4 estão em um nível apontado como ruim, o que será levado em consideração no desenvolvimento de modelos de menor complexidade.

## 4.2 Análise dos modelos computacionais

Os modelos computacionais são aqui desenvolvidos utilizando-se tanto uma abordagem direcionada a estados, como pelo agrupamento de todos os pacientes sem nenhuma distinção geográfica, sendo esta definida como uma abordagem nacional. Modelos específicos para estados são desenvolvidos para o RJ (região sudeste), CE (região nordeste) e RS (região sul), os quais são importantes centros para essas regiões do Brasil. Sabe-se que esses estados possuem diferentes aspectos culturais e podem apresentar níveis de desenvolvimento e infraestrutura distintos. Assim, é analisado se a eficiência na triagem de pacientes com suspeita de TB-DR pode ser melhorada ao considerarmos modelos computacionais para cada estado ao invés de uma única solução para o país. Como o estado de SP incluiu apenas 3 pacientes com o diagnóstico de TB-MDR e 4 com TB-DS, e também pelo fato de já termos um estado representante da região sudeste (RJ) com mais dados disponíveis, SP não foi considerado para análise no desenvolvimento de modelos por estados. Ainda, como RJ possui muito mais pacientes na base de dados que qualquer outro estado, é realizado um estudo da sua influência sobre a abordagem nacional ao se gerar um modelo que exclui os seus dados.

Dentre os objetivos deste estudo está a necessidade de se encontrar novas abordagens para a triagem de pacientes suspeitos de TB-DR que apresente uma alta taxa de sensibilidade, mas que ainda possuam um bom nível de especificidade, e que seja de simples implementação. Dessa forma, analisando a literatura e as potenciais alternativas encontradas, foi escolhido desenvolver modelos computacionais baseados em RNA e CART, os quais já haviam mostrado potencial não só em desempenho, mas em conseguir se adaptar e refletir as características de dados médicos. Esse mesmo direcionamento foi aplicado na avaliação comparativa entre se realizar uma abordagem nacional de diagnóstico ou localizada por estados.

Os resultados alcançados tanto para cada estado, quanto pela abordagem naci-

onal, podem ser vistos na Tabela 4.4. De forma a buscar um bom desempenho em sensibilidade, mas também não ter grandes perdas em especificidade, os modelos escolhidos buscaram o equilíbrio ao alcançar o maior índice SP possível. Assim como indicado em estudos anteriores, os modelos baseados em RNA também mostraram aqui um desempenho superior quando comparados aos desenvolvidos com CART, apresentando uma estrutura compacta (entre 3 e 4 neurônios, com exceção para o diagnóstico de TB-DR/MDR e TB-DS, com 6 em dois casos). No diagnóstico entre TB-DR/MDR e não TB-DR/MDR com RNA, o modelo para pacientes do RJ obteve uma sensibilidade de  $95,1(\pm 5,0)\%$ , sendo esta apropriada para triagem, enquanto CE e RS alcançaram valores de  $84,9(\pm 7,5)\%$  e  $84,8(\pm 8,8)\%$ , respectivamente. Observa-se que o desempenho do RJ é melhor do que o valor alcançado pelo modelo nacional, o qual exibe uma sensibilidade de  $82,7(\pm 6,0)\%$ . Ainda, o RJ também apresenta um desempenho superior quando considerado o índice SP, com um valor de  $89,9(\pm 3,2)\%$  contra  $80,9(\pm 2,5)\%$  do nacional. Apesar desse cenário também ser visto com o modelo do CE, no RS o desempenho acaba sendo comparável à abordagem nacional. Quando pacientes do RJ são removidos da base de dados, os resultados indicam um desempenho similar em sensibilidade e especificidade em relação ao modelo nacional. Entretanto, se com esse resultado é sugerido que possa não existir um viés por se ter um alto número de pacientes do RJ, a alta sensibilidade alcançada pelo modelo somente com pacientes do RJ não torna possível fazer essa afirmação. Outro ponto importante de observação é o total de pacientes incluídos na base com o diagnóstico de não TB. Ao obter, por estado, valores de sensibilidade maiores (em média) que aproximadamente 85%, os modelos, ao serem implantados nas unidades de saúde, podem vir a corrigir os erros da sistemática de encaminhamento que gerou esse volume de pacientes. Dessa forma, diminui-se a probabilidade da realização de tratamentos desnecessários e seus efeitos colaterais. Em geral, os modelos RNA mostraram um desempenho melhor que CART e a abordagem por estado apresenta melhores resultados que a nacional. Entretanto, as incertezas aos quais estão associados ainda são consideradas elevadas.

No diagnóstico entre TB-DR/MDR e TB-DS com modelos RNA, é visto que os

estados do CE e RS apresentam um índice SP mais elevado do que o apresentado pelo modelo nacional, com destaque para o CE, que chega ao um valor de  $94,7(\pm 3,5)\%$  contra  $77,1(\pm 3,6)\%$  ao incluir todos os estados. Assim, é observado que a abordagem estadual consegue alcançar uma sensibilidade, em média, maior que  $83\%$ , também considerada boa para triagem. Já o desempenho do RJ em índice SP, que é de  $83,3\pm 3,5\%$ , se mostra similar ao do modelo que não inclui os dados desse estado, com  $83,7(\pm 3,7)\%$ , o que também não indica viés nesse sentido.

Já para o diagnóstico entre TB-MDR e TB-DR, não só o CE, com  $87,3(\pm 5,5)\%$  de sensibilidade, mas também o RS apresentam valores maiores que o da abordagem nacional, chegando nesse caso a  $93,0(\pm 10,0)\%$  contra  $73,4(\pm 8,4)\%$  ao se reunir todos os estados. Como no diagnóstico de TB-DR, o RJ apresenta um valor de índice SP e sensibilidade similar ao que exclui os pacientes desse estado. Entretanto, o RJ mostra uma sensibilidade em média de  $76,7\%$  com uma incerteza associada de  $13,6\%$ , o que levanta dúvidas sobre sua aplicação em triagem. Em média, para TB-MDR, a sensibilidade alcançada pelos modelos estaduais é acima de  $77\%$ , novamente sendo superior a abordagem nacional.

De forma similar, os modelos CART também favorecem uma abordagem por estado. Ao discriminar TB-DR/MDR de não TB-DR/MDR, apesar dos modelos nacional e do RJ apresentarem sensibilidade similares de  $73,9(\pm 8,6)\%$  e  $73,4(\pm 12,1)\%$ , respectivamente, o índice SP de  $78,4(\pm 6,0)\%$  do RJ é maior do que o alcançado pelo modelo nacional, de  $69,7(\pm 3,3)\%$ . Já o pior desempenho, tanto em sensibilidade quanto índice SP, é obtido pelo RS e quando se desenvolve o modelo sem os pacientes do RJ. Considerando o suporte ao diagnóstico de TB-DS, é visto que o RJ e CE possuem sensibilidade maior que o modelo nacional, da mesma forma que quando o RJ é excluído. Já os resultados obtidos pelo modelo do RS, assim como os alcançados no diagnóstico de TB-MDR contra TB-DR, são mais difíceis de analisar por apresentarem altos valores em sua incerteza.

Tabela 4.4: Desempenho dos modelos de RNA e CART por região, obtidos através da validação cruzada, com informações sobre o número (#N) de neurônios na camada escondida ou profundidade da árvore de classificação, e ponto de corte (RNA) considerando o suporte ao diagnóstico entre TB-DR/MDR e não TB-DR/MDR; TB-DR/MDR e TB-DS; e TB-MDR e TB-DR.

Diag.	Região	Método	#N	Corte	Índice SP (%)	Sensib. (%)	Especif. (%)	Acurácia (%)
TB-DR/MDR x não TB-DR/MDR	Nacional	RNA	4	0.20	80,9(±2,5)	82,7(±6,0)	79,4(±6,2)	80,1(±2,9)
		CART	6	-	69,7(±3,3)	73,9(±8,6)	65,8(±5,9)	68,9(±2,9)
		AdaBoost	-	-	75,4(±2,9)	69,5(±5,6)	81,5(±3,6)	76,6(±2,6)
	Não RJ	RNA	4	-0.10	80,0(±2,6)	76,0(±8,7)	84,3(±7,0)	80,6(±3,1)
		CART	8	-	57,1(±4,5)	51,7(±10,3)	63,6(±10,7)	57,9(±4,5)
		AdaBoost	-	-	66,7(±6,4)	72,7(±6,8)	61,1(±9,7)	67,3(±6,1)
	RJ	RNA	4	0.51	89,9(±3,2)	95,1(±5,0)	85,0(±4,9)	88,8(±4,4)
		CART	6	-	78,4(±6,0)	73,4(±12,1)	83,9(±5,8)	80,6(±4,4)
		AdaBoost	-	-	80,6(±3,9)	71,2(±7,5)	90,8(±4,8)	84,7(±3,3)
	CE	RNA	3	0.22	89,7(±3,7)	84,9(±7,5)	94,7(±4,7)	90,5(±4,1)
		CART	5	-	67,3(±8,0)	65,4(±12,1)	69,9(±12,5)	67,7(±7,7)
		AdaBoost	-	-	74,0(±6,5)	76,5(±10,0)	72,1(±11,1)	74,2(±6,5)
	RS	RNA	3	0.00	83,6(±4,8)	84,8(±8,8)	82,8(±8,7)	82,9(±4,9)
		CART	4	-	53,3(±10,5)	54,5(±17,6)	54,4(±15,1)	54,6(±9,2)
		AdaBoost	-	-	51,1(±13,3)	56,3(±13,6)	48,1(±16,2)	52,2(±11,3)
TB-DR/MDR x TB-DS	Nacional	RNA	6	0.29	77,1(±3,6)	73,7(±7,6)	80,9(±8,9)	77,9(±4,7)
		CART	8	-	57,7(±6,6)	62,0(±8,0)	54,3(±13,2)	59,9(±5,6)
		AdaBoost	-	-	56,2(±7,0)	87,2(±5,1)	32,7(±9,9)	73,1(±4,9)
	Não RJ	RNA	3	-0.24	83,7(±3,7)	79,5(±10,4)	88,6(±9,6)	84,5(±4,5)
		CART	6	-	64,2(±8,2)	68,1(±9,8)	61,6(±16,6)	66,4(±6,7)
		AdaBoost	-	-	55,5(±11,3)	86,1(±7,6)	33,5(±12,1)	72,5(±6,5)
	RJ	RNA	4	0.11	83,3(±3,5)	82,6(±8,1)	84,3(±8,9)	83,8(±5,2)
		CART	4	-	52,8(±14,4)	72,6(±11,6)	39,2(±18,6)	63,5(±9,5)
		AdaBoost	-	-	58,2(±12,4)	83,5(±10,7)	40,0(±16,3)	71,8(±7,8)
	CE	RNA	3	-0.17	94,7(±3,5)	91,3(±5,3)	98,2(±5,3)	96,2(±3,8)
		CART	3	-	68,2(±13,8)	74,7(±13,9)	64,2(±23,5)	71,6(±11,0)
		AdaBoost	-	-	64,1(±21,9)	85,4(±9,9)	52,0(±29,9)	76,8(±9,9)
	RS	RNA	6	-0.06	88,2(±4,0)	84,2(±9,0)	92,7(±9,6)	89,1(±4,9)
		CART	4	-	45,6(±21,4)	58,8(±17,5)	42,8(±28,1)	53,1(±13,5)
		AdaBoost	-	-	36,6(±20,1)	72,9(±13,6)	21,1(±14,6)	55,8(±10,4)
TB-MDR x TB-DR	Nacional	RNA	4	0.09	73,8(±2,9)	73,4(±8,4)	74,5(±8,1)	74,3(±3,8)
		CART	7	-	56,5(±5,9)	60,6(±12,6)	53,8(±11,5)	57,2(±5,6)
		AdaBoost	-	-	56,4(±5,7)	55,9(±9,8)	57,5(±8,6)	56,7(±5,7)
	Não RJ	RNA	3	-0.30	79,3(±3,8)	85,2(±7,4)	74,0(±9,0)	80,7(±5,3)
		CART	2	-	59,6(±8,8)	64,9(±15,0)	55,4(±11,2)	59,9(±8,3)
		AdaBoost	-	-	58,4(±7,4)	57,7(±13,6)	60,1(±12,2)	58,7(±7,2)
	RJ	RNA	4	0.48	79,1(±4,5)	76,7(±13,5)	82,5(±11,8)	78,9(±4,9)
		CART	4	-	46,2(±13,1)	59,0(±17,6)	39,3(±18,4)	50,5(±9,7)
		AdaBoost	-	-	52,1(±12,8)	65,7(±14,7)	41,8(±18,3)	55,1(±11,6)
	CE	RNA	3	-0.69	88,9(±5,5)	87,3(±5,5)	90,8(±10,1)	88,6(±5,4)
		CART	3	-	51,1(±16,8)	60,0(±21,3)	49,5(±23,4)	55,1(±11,9)
		AdaBoost	-	-	53,6(±10,9)	66,0(±17,5)	44,0(±16,0)	55,6(±11,1)
	RS	RNA	3	-0.25	94,3(±6,1)	93,0(±10,0)	95,8(±8,4)	94,1(±5,7)
		CART	2	-	50,8(±22,3)	51,4(±25,2)	59,2(±25,7)	54,4(±16,6)
		AdaBoost	-	-	57,1(±17,9)	59,4(±19,3)	59,4(±20,0)	59,5(±13,3)



Dessa forma, tanto os resultados obtidos pelas RNA, quanto por CART, indicam que uma abordagem por estado, ao gerar escores clínicos para a triagem de pacientes com suspeita de TB-DR, pode ser o mais indicado. Ao se comparar com os resultados de AdaBoost, nota-se que, em geral, ocorre uma diminuição dos valores de incerteza em relação ao CART. Essa é uma vantagem encontrada para esse tipo de modelo computacional. Para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR, as incertezas do AdaBoost são menores em sensibilidade e especificidade, onde alcançam os valores de  $69,5(\pm 5,6)\%$  e  $81,5(\pm 3,6)\%$ , respectivamente, tanto em comparação com o modelo RNA,  $82,7(\pm 6,0)\%$  e  $79,4(\pm 6,2)\%$ , quanto com o CART,  $73,9(\pm 8,6)\%$  e  $65,8(\pm 5,9)\%$ . Em especificidade, ainda consegue ser melhor, em média, que a RNA. Diferenças nas análises também podem ser encontradas. Se comparamos o modelo do RJ com o que exclui esses pacientes, ao contrário do CART, os resultados se encontram similares em termos de sensibilidade, sendo  $71,2(\pm 7,5)\%$  para a abordagem estadual e  $72,7(\pm 6,8)\%$  ao agrupar os dados restantes. O modelo CART sem o RJ havia alcançado  $51,7(\pm 10,3)\%$  em sensibilidade. Em termos de especificidade, o resultado com AdaBoost para o RJ ainda consegue ser superior, em média, ao modelo RNA. Já para os estados do CE e RS, apesar de diminuir as incertezas obtidas com CART, seus resultados são similares. Assim como com CART, as incertezas aumentam com a menor disponibilidade de dados. Na discriminação entre TB-DR/MDR e TB-DS, apesar de diminuir incertezas em sensibilidade e especificidade, e atingir um valor de sensibilidade maior que RNA e CART,  $87,2(\pm 5,1)\%$  contra  $73,7(\pm 7,6)\%$ , a especificidade é menor que a metade, em média, que RNA, alcançando  $32,7(\pm 9,9)\%$  contra  $80,9(\pm 8,9)\%$ . Esse resultado pode apontar uma existência de um custo-benefício para se atingir valores melhores que outros métodos. Para o modelo que exclui o RJ, a sensibilidade com o AdaBoost é comparável ao modelo RNA,  $86,1(\pm 7,6)\%$  e  $79,5(\pm 10,4)\%$ , e similar ao alcançado pelo modelo do RJ,  $83,5(\pm 10,7)\%$ , assim como já tinha acontecido com CART e RNA. Para o estado do CE, é visto um aumento na incerteza com AdaBoost em comparação com CART, tanto em sensibilidade,  $52,0(\pm 29,0)\%$  e  $64,2(\pm 23,5)\%$ , respectivamente, quanto índice SP,  $61,1(\pm 21,9)\%$  e  $68,2(\pm 13,8)\%$ . Já no modelo do RS,

apesar de ter menores incertezas que CART, novamente é visto que a especificidade diminui para a metade, em média, passando de  $42,8(\pm 28,1)\%$  para  $21,1(\pm 14,6)\%$ . Os resultados para o diagnóstico entre TB-MDR e TB-DR apresentam resultados similares a CART, que apesar de terem menores incertezas, ainda continuam com valores altos. Ao analisarmos os desempenhos de todos os modelos desenvolvidos, é possível observar que as incertezas tendem a aumentar conforme a quantidade de pacientes na amostra diminui. Essa situação pode ser mitigada se mais dados sobre os diagnósticos de pacientes forem coletados.

Com o objetivo de ilustrar o processo de desenvolvimento dos modelos RNA, a Figura 4.2 exibe um exemplo de escolha da melhor topologia. A quantidade de neurônios na camada escondida da rede é variada e o índice SP calculado ao se aplicar os dados de teste, considerando, para esse caso, a discriminação de pacientes entre TB-DR/MDR e não TB-DR/MDR para o RJ. Desse modo, a configuração das redes que possuem um total de quatro neurônios é escolhida, observando, para isso, o pico do valor de SP. Em geral, os modelos de RNA desenvolvidos apresentaram uma boa generalização, o que pode ser visto na evolução da curva de treinamento da rede de operação mostrada na Figura 4.3. É possível ver a situação de *overfitting*, assim como a aplicação do critério de parada da rede para evitá-lo. A condição de generalização também pode ser analisada por meio da distribuição dos valores de saída da rede, como observado na Figura 4.4. Ambas as classes, apesar da estatística limitada, possuem seus picos nas extremidades do histograma, seja aplicando os dados de treino ou teste. Os picos são próximos aos valores de +1 e -1, e sendo esses os alvos da rede neural, é portanto, um bom resultado.

Ainda analisando as redes de operação, quando comparamos as curvas ROC (Figura 4.5) dos modelos do RJ e nacional, também para o diagnóstico de TB-DR/MDR, é observado que, ao se propagar todos os dados, a abordagem estadual mostra resultados melhores que a nacional. Isso acontece tanto em termos de sensibilidade quanto em especificidade, ao se variar o ponto de operação da RNA. Para o RJ, é possível escolhermos um ponto de operação que forneça altos valores de sensibilidade (acima de 90%), mantendo a especificidade acima de 80%, o que não

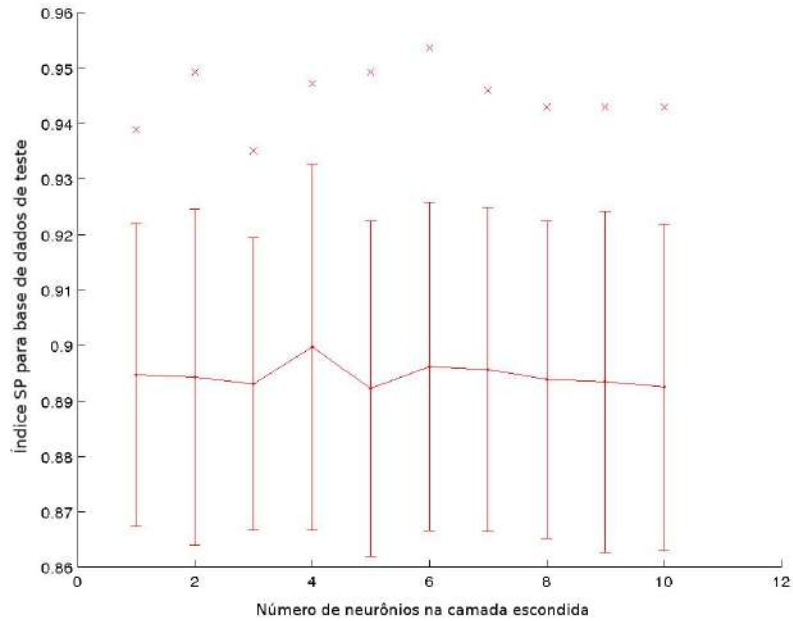


Figura 4.2: Índice SP obtido pela RNA, a partir da validação cruzada utilizando até 10 neurônios na camada escondida da rede, para o suporte a discriminação entre TB-DR/MDR e não TB-DR/MDR com o modelo RJ.

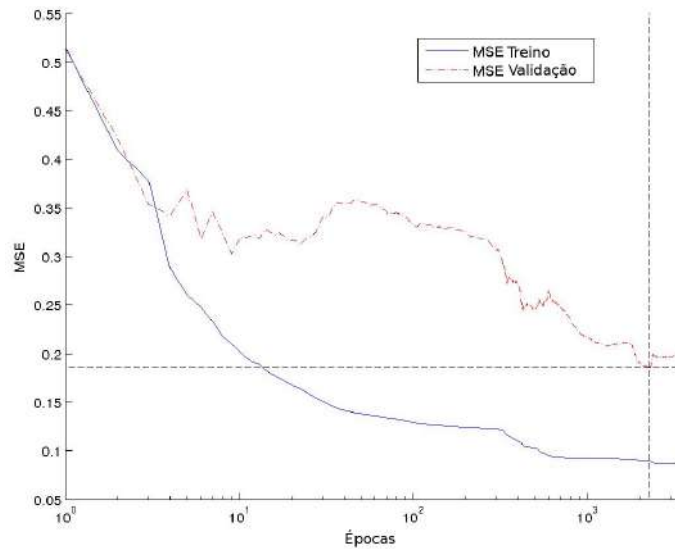


Figura 4.3: Curva de aprendizado da RNA, em termos de erro quadrático médio, assim como a indicação da época referente a parada do treinamento da rede, considerando o suporte ao diagnóstico entre TB-DR/MDR e não TB-DR/MDR no modelo RJ.

é viável quando se agrupa os pacientes nacionalmente.

Por meio do critério de seleção por maior índice SP, é possível visualizar, também, a estrutura da árvore de classificação de operação obtida pelo CART. Como visto na Figura 4.6, para o diagnóstico de TB-DR/MDR considerando pacientes do RJ, a

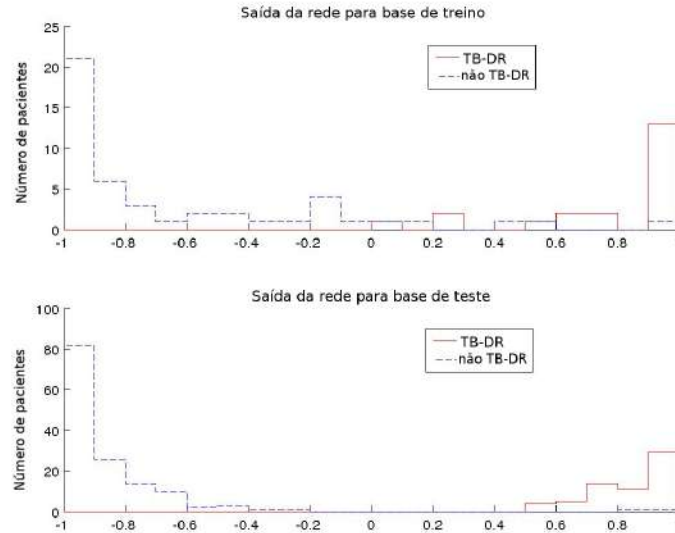


Figura 4.4: Histograma com os valores de saída da rede neural que apresentou maior índice SP, na discriminação dos diagnósticos de TB-DR/MDR e não TB-DR/MDR, para o modelo RJ.

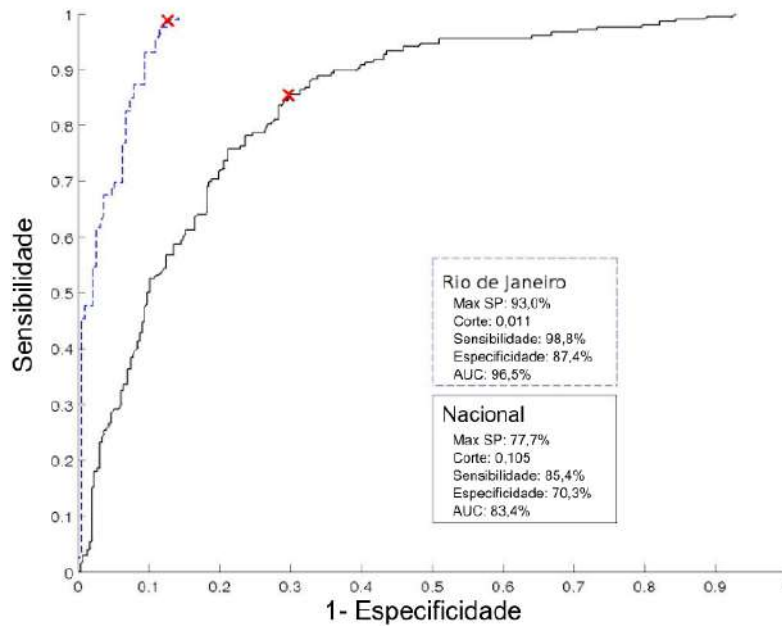


Figura 4.5: A curva ROC para as RNA desenvolvidas para o estado do RJ e pela abordagem nacional, quando propagando todos os dados. Aqui é considerado o diagnóstico para discriminar TB-DR/MDR de não TB-DR/MDR. Os pontos de corte que representam o maior valor do índice SP são indicados, assim como os valores de sensibilidade, especificidade, índice SP e AUC.

variável localizada na raiz da árvore corresponde a cavitação pulmonar. Em termos de qualidade de dados é algo a ser ressaltado, pois essa é uma informação que depende da existência de infraestrutura na unidade de saúde, tendo apresentado um

alto número de respostas ignoradas na base de dados. Em seguida, a quantidade de vezes anteriores que o paciente tratou TB é também incluída como ponto de decisão, onde o diagnóstico de TB-DR/MDR é apontado quando esse valor é acima de 2,5. Entende-se que uma recorrência de tratamentos é um dos indícios esperados pela literatura médica como marcador de resistência. Da mesma forma, um tempo maior que 1,25 meses em que o paciente apresenta escarro também é visto indicando o diagnóstico de TB-DR/MDR, assim como uma variação de IMC maior que 5,013 (reflete a variação de peso do paciente). Ainda aparece na árvore o desfecho do paciente, que também possui questões de qualidade relacionadas a acessibilidade e dados faltantes. Sendo a falência do tratamento anterior um dos indicadores de TB-DR, esse se mostra um resultado razoável da perspectiva médica. Em níveis de menor poder de discriminação estão a perda de apetite e o contato anterior com TB-DR, em que a resposta positiva para ambos aponta um diagnóstico negativo para TB-DR/MDR. Esse resultados pode refletir uma característica da base de dados. Assim, a árvore resultante apresenta um total de 8 nós e altura 4, contendo 9 folhas, as quais são responsáveis por apontar os diagnósticos.

Já a evolução do desempenho que é obtido ao se construir os modelos que compõem o AdaBoost é visto na Figura 4.7. Nela, é também considerado o diagnóstico de TB-DR/MDR contra não TB-DR/MDR para o modelo do RJ. Tanto com os dados de treino, quando com os de teste, é possível observar que um total de 30 modelos, operando em conjunto, são suficientes para não haver indícios de melhora nos resultados que justifique uma necessidade da inclusão de mais estruturas. Ao analisar os resultados com dados de teste, é visto que a especificidade aumenta com a inclusão de mais modelos até não apresentar grande variação a partir de um total de 10 modelos. Em contrapartida, a sensibilidade sofre uma queda até a inclusão do terceiro modelo, recuperando aos poucos conforme o método evolui no treinamento. Entretanto, os pesos dados a cada modelo indicam que os dois primeiros possuem uma relevância bem maior que os demais, com uma queda mais suave, comparativamente, a partir do terceiro. Esse fator ajuda a explicar o resultado alcançado pelo método quando consolida todas as decisões sobre o diagnóstico, além de reforçar a

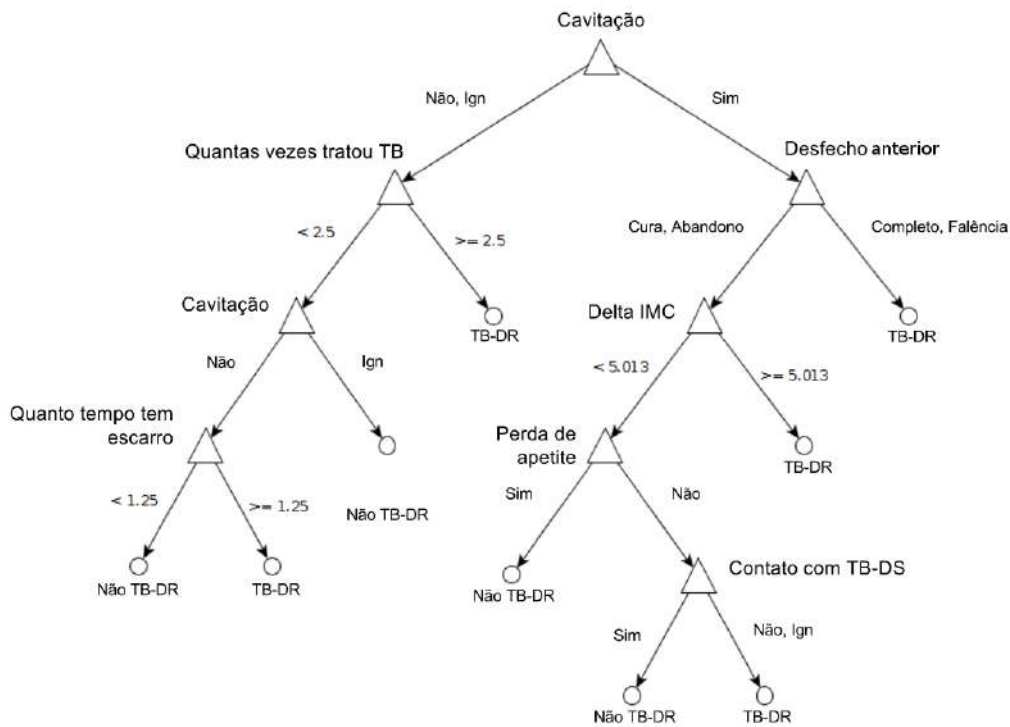


Figura 4.6: Estrutura da árvore de classificação, gerada utilizando CART, para a discriminação entre TB-DR/MDR e não TB-DR/MDR, considerando o desenvolvimento do modelo do estado do RJ.

suficiência do total de modelos utilizados.

Após essa análise, pode-se observar os principais modelos utilizados pelo algoritmo do AdaBoost na sua classificação. Assim, na Figura 4.8 observa-se a estrutura dos 10 modelos aos quais foram atribuídos os maiores pesos. O modelo de maior peso utiliza a variável de cavitação pulmonar, seguido do desfecho do tratamento anterior que indica a falência do tratamento. Contato com TB pulmonar positiva aparece tanto no modelo 3, quanto no modelo 9, mostrando que o algoritmo pode, após um determinado número de iterações, retornar com um modelo de fraca aprendizagem que utilize novamente uma variável, mas em outro ponto de corte. Esse é um comportamento que pode também ser visto em uma estrutura CART, com a mesma variável presente em níveis diferentes da hierarquia. Entretanto, como analisado anteriormente, um dos resultados alcançados pelo AdaBoost foram as incertezas menores que o CART. Outras variáveis que aparecem são a quantidade de vezes que o paciente tratou TB, contato anterior com TB-DR, tempo de expectoração e tosse, além da perda de apetite e cor da pele preta.

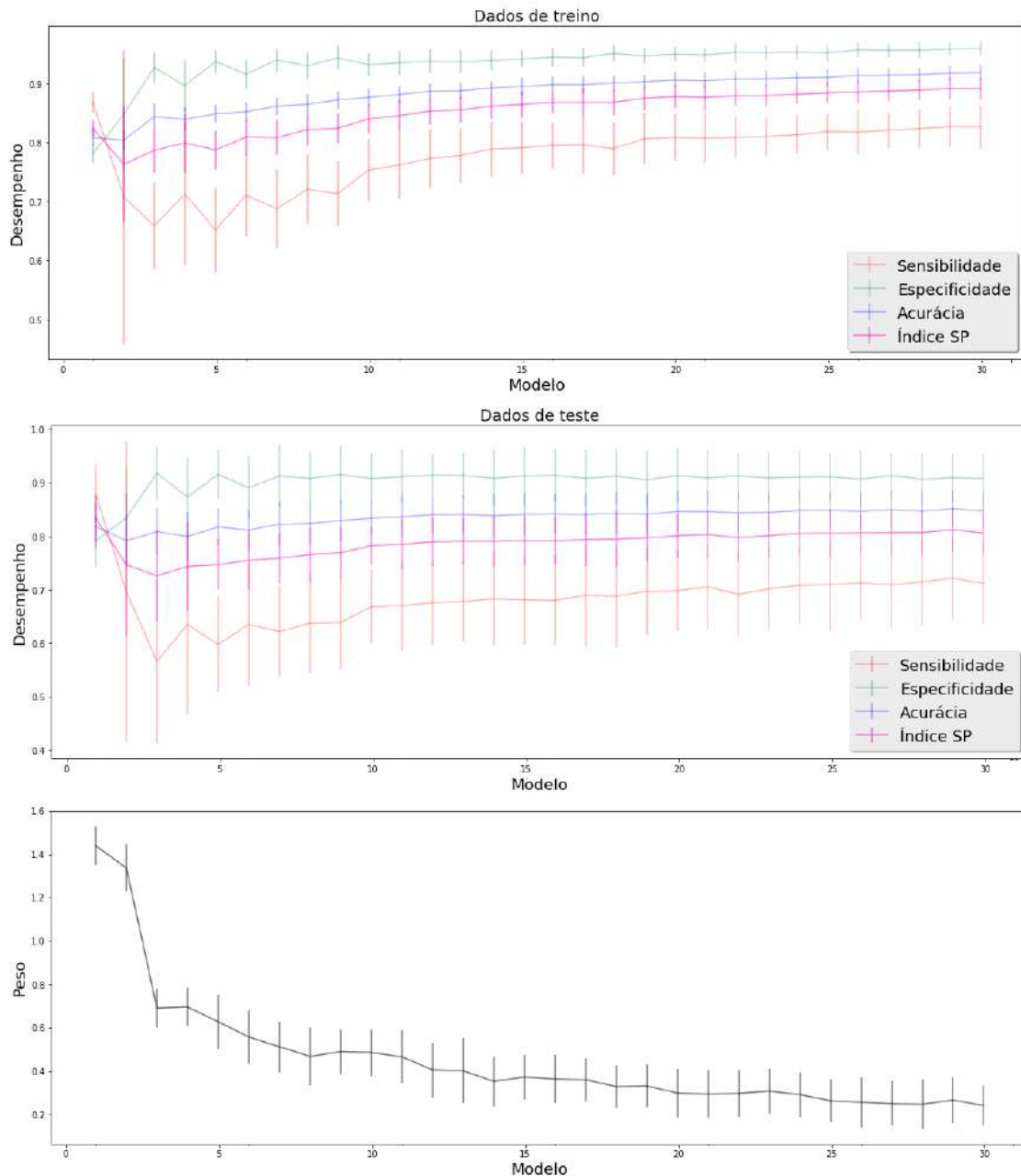


Figura 4.7: Evolução do desempenho por modelo CART de fraco aprendizado, para dados de treino e teste, utilizando o algoritmo AdaBoost. É considerado o diagnóstico entre TB-DR/MDR e não TB-DR/MDR para o estado do RJ. O peso de cada modelo para o resultados final também é mostrado.

De forma a conseguir avaliar também o estado de SP, apesar de seus pacientes não terem sido considerados para o desenvolvimento de um modelo de abordagem estadual, o desempenho ao propagá-los pelo modelo do RJ é avaliado. Dessa forma, é estudada a possível portabilidade dos modelos para outros estados da mesma região, no caso, a sudeste. Para também investigar compatibilidades com outros locais, o

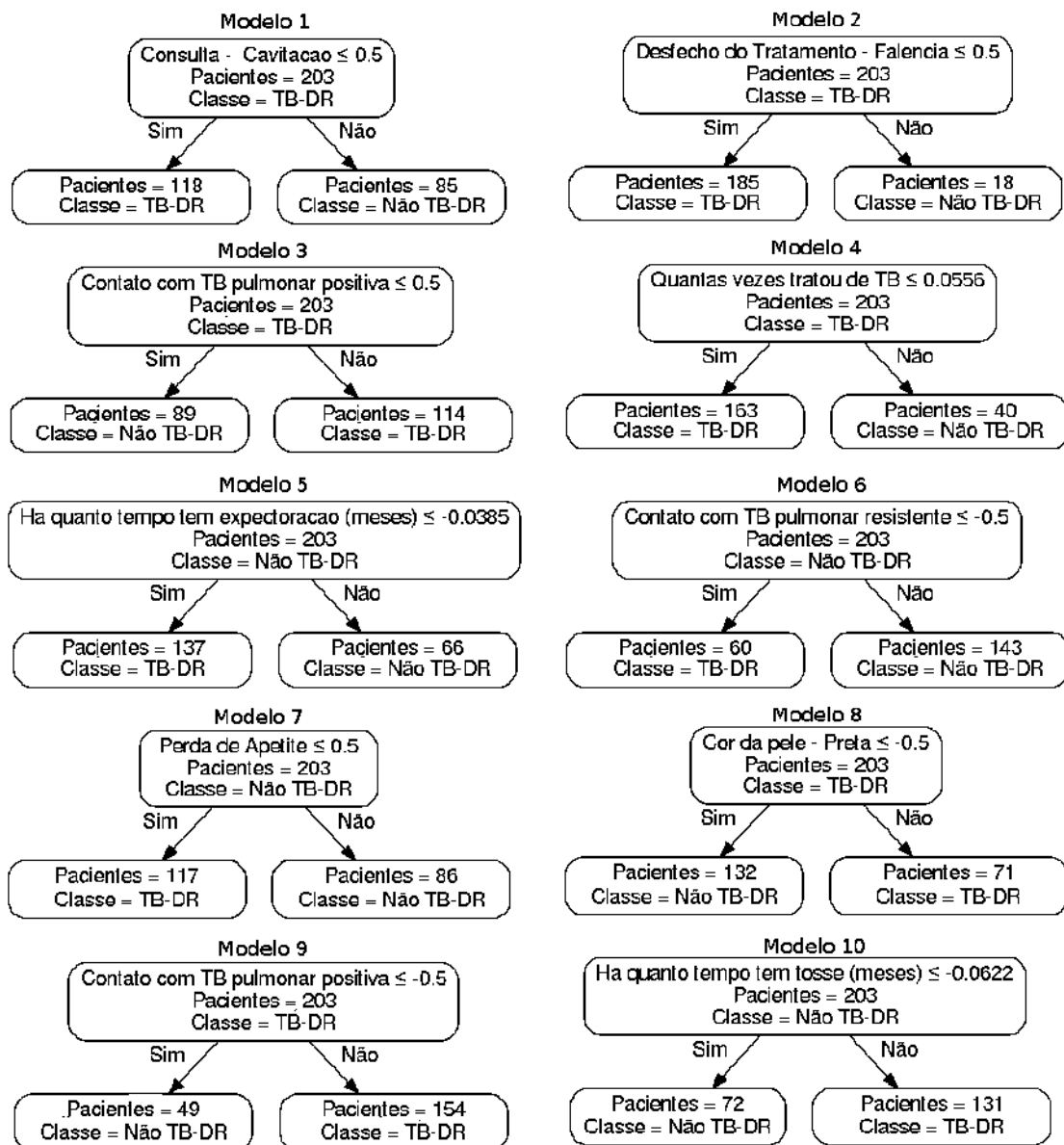


Figura 4.8: Estrutura dos 10 modelos CART de fraco aprendizado de maior peso na decisão final obtidos pelo algoritmo AdaBoost. Os modelos são para o diagnóstico de TB-DR contra não TB-DR para o estado do RJ.

desempenho dos pacientes de SP com os modelos do CE e RS é observado. A Tabela 4.5 mostra os resultados obtidos, utilizando para isso os modelos RNA. Assim, nota-se que o desempenho com os pacientes de SP não é alto em nenhum dos modelos estaduais, incluindo o RJ. Há uma queda nos índices, tanto em sensibilidade, quanto especificidade, em todos os casos, quando comparados com o alcançado ao se diagnosticar os pacientes dos locais de especialização dos modelos. Portanto, levanta-se a hipótese de que a geração de um modelo específico para SP seja necessário, o que neste estudo é uma limitação dado a quantidade de dados disponíveis desse estado.



Essa situação pode ser corroborada quando se leva em consideração as diferenças existentes entre os programas de TB em cada unidade da federação, os quais podem ser particulares a cada local.

Tabela 4.5: Desempenho alcançado, por diagnóstico, ao se propagar os pacientes do estado de SP pelos modelos RNA desenvolvidos para os estados do RJ, CE e RS.

Diagnóstico	Estado	Índice SP (%)	Sensibilidade (%)	Especificidade (%)	Acurácia (%)
TB-DR/MDR x Não TB-DR/MDR	RJ	37,8(±7,6)	50,5(±8,6)	28,5(±15,5)	45,2(±5,8)
	CE	57,6(±6,3)	46,2(±12,0)	71,1(±8,4)	52,2(±8,5)
	RS	43,1(±20,1)	47,7(±19,2)	52,0(±30,9)	48,7(±11,4)
TB-DR/MDR x TB-DS	RJ	27,9(±27,3)	65,5(±17,0)	26,0(±28,8)	60,8(±13,5)
	CE	37,6(±30,0)	65,5(±11,7)	36,6(±35,1)	62,0(±9,1)
	RS	36,0(±31,2)	63,0(±22,2)	42,0(±40,2)	60,5(±17,1)
TB-MDR x TB-DR	RJ	50,5(±14,0)	58,5(±26,0)	47,3(±17,1)	49,3(±12,8)
	CE	48,8(±9,9)	58,0(±24,9)	43,2(±12,5)	45,9(±8,5)
	RS	45,3(±8,8)	62,5(±24,8)	35,1(±18,2)	40,0(±12,2)

Outro estudo realizado é sobre a influência dos pacientes do estado de SP nos resultados alcançados. É visto que o desempenho dos modelos estaduais são, geralmente, melhores do que os nacionais. Com isso, apesar da baixa representatividade, considera-se que a presença de dados de SP poderia estar influenciando para diminuir os índices de desempenho. Dessa forma, com o objetivo de avaliar esse possível impacto, são desenvolvidos modelos RNA para uma abordagem nacional onde não se inclui pacientes de SP. Os resultados são vistos na Tabela 4.6. Como pode ser observado, os resultados são muito similares aos vistos anteriormente, ao se incluir pacientes de SP. Como um exemplo, a sensibilidade na triagem entre TB-DR/MDR e não TB-DR/MDR sem SP é de 83,3(±5,5)%, enquanto com todos os dados incluídos havia se alcançado 82,7(±6,0)%.

Com os modelos desenvolvidos, é também possível analisar como seria o comportamento, em uma situação de triagem, dos diagnósticos de não TB em que os pacientes apresentavam tratamento há mais de 30 dias. Esses pacientes foram excluídos da base de dados, e portanto, não utilizados para o desenvolvimento dos modelos. Porém, em uma situação de triagem, é investigado como seria o diagnóstico desses casos. Caso suas características sejam similares aos dos pacientes não TB sem tratamento (ou com menos de 30 dias de tratamento), espera-se que o mo-

Tabela 4.6: Desempenho alcançado, por diagnóstico, ao se desenvolver modelos RNA utilizando os dados dos pacientes do RJ, CE e RS, e excluindo dados de pacientes do estado de SP.

Diagnóstico	Índice SP (%)	Sensibilidade (%)	Especificidade (%)	Acurácia (%)
TB-DR/MDR x Não TB-DR/MDR	81,2(±2,1)	83,3(±5,5)	79,3(±5,7)	80,9(±2,4)
TB-DR/MDR x TB-DS	77,7(±2,9)	75,1(±7,9)	80,7(±8,1)	76,6(±4,3)
TB-MDR x TB-DR	77,1(±3,8)	78,0(±9,8)	76,7(±9,1)	77,9(±4,1)

delo treinado para não TB-DR indique uma tendência a classificá-los no diagnóstico menos complexo. Na Figura 4.9 observam-se os resultados, para cada abordagem, da classificação desses casos. Para essa análise, utiliza-se o modelo RNA para a discriminação entre TB-DR/MDR e não TB-DR/MDR. Assim, é observado que na maioria dos casos o modelo não indica um diagnóstico claro para a maioria desses pacientes. Pelo contrário, tanto para o modelo nacional, quanto para o RJ e RS, os resultados apresentam valores quase que classificando 50% dos casos em TB-DR/MDR, e os outros 50% em não TB-DR/MDR. A exceção é vista com o estado do CE, onde dos 5 pacientes, 76,0(±8,0)% são classificados como TB-DR/MDR. Esses resultados podem confirmar a hipótese utilizada para exclusão desses pacientes do estudo aqui realizado, na qual o diagnóstico confirmado por exames como não TB poderia ter sido influenciado pelo tratamento anti-TB ao qual esses pacientes já estavam utilizando.

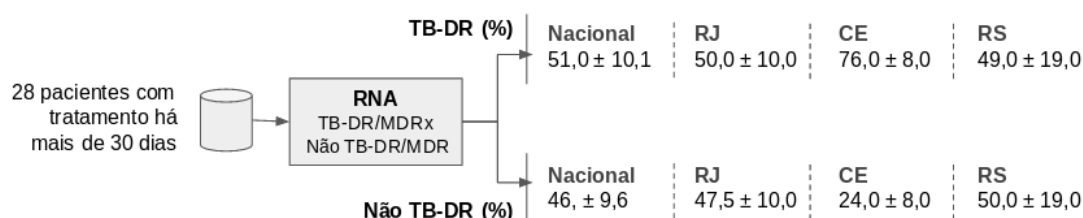


Figura 4.9: Classificação dos pacientes que foram excluídos do estudo por apresentarem tratamento há mais de 30 dias e diagnóstico de não TB. Para isso, é utilizado o modelo RNA para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR, e o resultado para cada abordagem é indicado.

Além disso, é possível avaliar o que acontece na discriminação dos pacientes

incluídos e que possuem o diagnóstico de não TB. Considera-se os pacientes que estavam em tratamento antes da triagem (menos de 30 dias) e os que não estavam. Assim, é possível avaliar se os modelos conseguem ser robustos a uma falha no processo de encaminhamento. A Tabela 4.7 indica a classificação dos pacientes diagnosticados como não TB pelos modelos de abordagem nacional e por estado. Observa-se que, independentemente da abordagem e do estado, os pacientes são de fato, em sua maioria, encaminhados para o diagnóstico de não TB-DR/MDR. Dessa forma, entende-se que os modelos conseguem lidar com possíveis problemas de processo nas unidades de saúde e gerar resultados que refletem o comportamento desejado.

Tabela 4.7: Análise da discriminação dos pacientes com o diagnóstico de não TB para avaliar as possíveis diferenças de comportamento do modelo com os que tinham menos de 30 dias de tratamento e os sem tratamento. São utilizados os modelos RNA de classificação entre TB-DR/MDR e não TB-DR/MDR para os estados do RJ, CE e RS.

Abordagem	Não TB com tratamento (menos 30 dias)		Não TB sem tratamento	
	TB-DR/MDR	Não TB-DR/MDR	TB-DR/MDR	Não TB-DR/MDR
Nacional	5,28( $\pm 1,47$ )	8,72( $\pm 1,47$ )	27,66( $\pm 11,07$ )	178,34( $\pm 11,07$ )
RJ	2,42( $\pm 1,23$ )	4,58( $\pm 1,23$ )	11,36( $\pm 7,38$ )	139,64( $\pm 7,38$ )
CE	0,10( $\pm 0,36$ )	1,90( $\pm 0,36$ )	1,68( $\pm 1,91$ )	29,3( $\pm 1,91$ )
RS	0,36( $\pm 0,63$ )	2,64( $\pm 0,63$ )	3,96( $\pm 3,92$ )	15,04( $\pm 3,92$ )

### 4.3 Análise de relevância das variáveis

Após a geração dos modelos computacionais, a relevância das variáveis é calculada de forma a apontar qual delas possui mais influência no diagnóstico realizado pelo modelo. Um exemplo de como o índice de relevância consolidado é composto, utilizando para isso os resultados dos métodos *Weights* e *Relevance mapping*, é mostrado na Figura 4.10. Aqui é considerado o modelo do RJ para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR. Ao observar as variáveis selecionadas para cada estratégia (Loose, Medium e Tight), é possível notar que diferentes fatores podem influenciar essa escolha. Variáveis podem aumentar sua relevância quando o resultado alcançado pelos dois métodos são somados, ou então manter sua posição, sendo

esse o caso da cavitação pulmonar (destaque em ambos). Por outro lado, um valor muito maior em um determinado método pode compensar um valor menor atingido pelo outro e, com isso, influenciar na inclusão da variável em uma das estratégias que visam modelos menos complexos. Essa situação é vista com a variável que indica quantas vezes o paciente tratou de TB, assim como com a cor de pele indígena. Ambas alcançam valores baixos por meio do *Relevance mapping*, mas altos pela avaliação do *Weights*. Outro fator que gera impacto, e que pode também ser observado, refere-se aos aspectos relacionados a qualidade de dados levados em consideração na composição das estratégias. Por exemplo, contato anterior com TB-DR não é incluída na seleção Tight. Nesse caso, a decisão de excluí-la é pela potencialidade de se encontrar problemas relacionados a validade, os quais são inerentes desse tipo de informação. Esse mesmo critério pode ser visto na não inclusão da condição de fumante em conjunto com as variáveis Loose.

Um resumo do desempenho alcançado, para cada região e diagnóstico, considerando o índice consolidado, é mostrado na Tabela 4.8. Comparando os resultados das diferentes estratégias, os modelos menos complexos mostram ser competitivos quando comparados em termos de índice SP. Entretanto, é possível observar que, para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR, o número de neurônios na camada escondida é geralmente maior do que o visto no modelo original (4 neurônios). Observa-se ainda que o número de variáveis pode ser menor, mas mesmo assim não implicar em uma diminuição no número de neurônios necessários para a modelagem dos dados, como acontece na abordagem nacional. Da mesma forma, a estratégia Medium, na abordagem nacional, chega a alcançar um melhor desempenho do que a Loose. Esse é um resultado que pode indicar a presença de variáveis confundidoras no modelo mais complexo, já que o acréscimo de informação está prejudicando o desempenho, e não auxiliando. Ao observar as incertezas, seus valores, em todos os casos, tendem a aumentar conforme o volume de dados se torna mais escasso. Uma exceção pode ser vista ao se diagnosticar pacientes do RJ, quando apresentam valores menores tanto para sensibilidade, quanto para especificidade. Já no diagnóstico entre TB-DR/MDR e TB-DS, para o estado do CE na estratégia

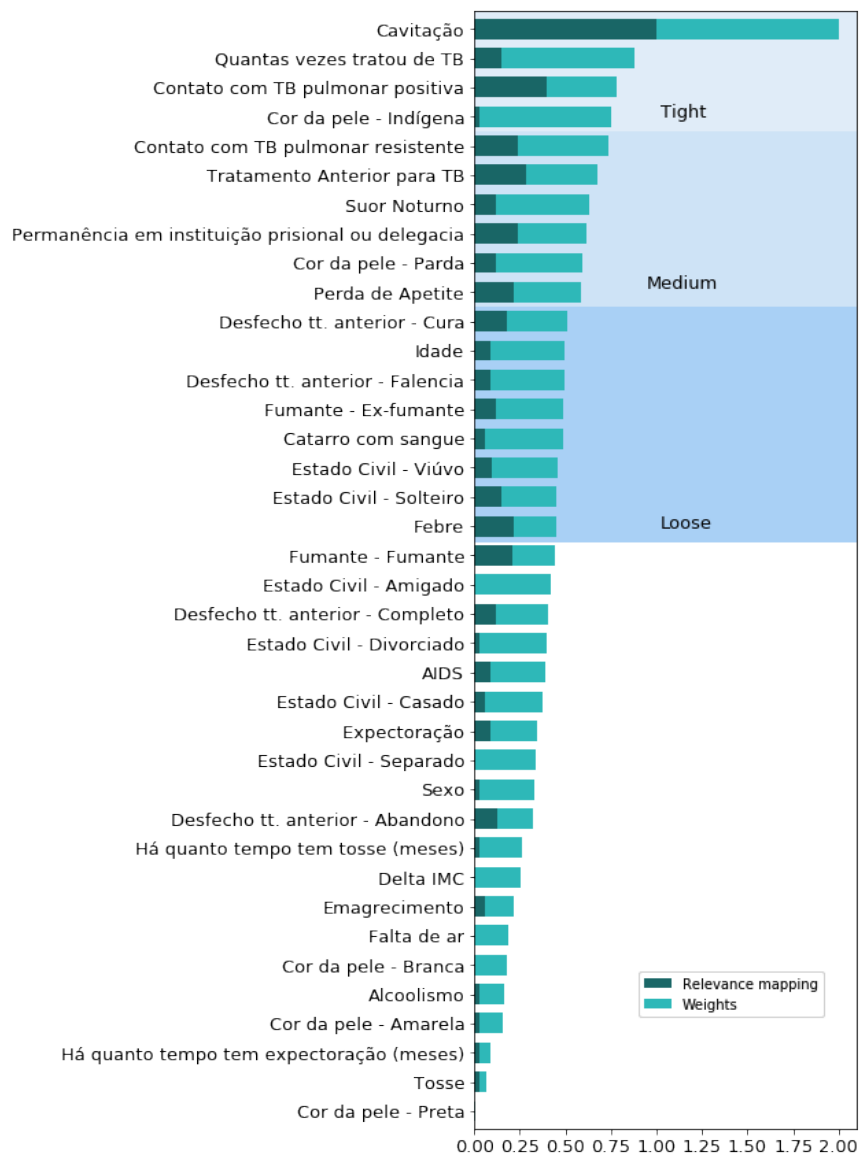


Figura 4.10: Lista com todas as variáveis ordenadas pelo índice de relevância consolidado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR. Os resultados obtidos pelos métodos *Relevance mapping* e *Weights* são somados para se chegar ao resultado mostrado, sendo que cada respectiva contribuição é ainda indicada como referência. As variáveis selecionadas em cada abordagem (Tight, Medium e Loose) são ressaltadas, refletindo também os critérios de qualidade de dados utilizados.

Tight, o modelo consegue alcançar uma especificidade similar ao gerado com todas as variáveis disponíveis. Apesar da estratégia Tight atingir valores mais baixos de índice SP, a maioria dos modelos estaduais ainda conseguem atender a um cenário de triagem mesmo em ambientes com baixa disponibilidade de recursos. Em geral, uma pequena perda em desempenho ainda pode ser aceitável em troca de menos informações necessárias para um diagnóstico. Portanto, um desempenho adequado

para triagem pode ser visto quando realiza-se o diagnóstico entre: TB-DR/MDR e não TB-DR/MDR, considerando sensibilidade, para pacientes do RJ e CE em todas as estratégias, além do RS com exceção do cenário Tight; TB-DR/MDR e TB-DS, também considerando sensibilidade, para o RJ (Medium e Loose), CE e a estratégia Loose no RS, e ainda, em termos de especificidade, as estratégias Medium e Loose no RS; e TB-MDR e TB-DR nas abordagens estaduais do CE e RS em sensibilidade e especificidade, assim como no RJ apenas em especificidade para os critérios Medium e Loose.

É possível identificar características importantes sobre os critérios para o diagnóstico ao se analisar as variáveis mais relevantes em cada modelo (Tabela 4.9). Analisando o exemplo do diagnóstico entre TB-DR/MDR e não TB-DR/MDR, é visto que o histórico de tratamento anterior (desfecho) possui, em geral, um papel de importância no resultado. Nesse escopo, a variável que indica a falência do tratamento é incluída na estratégia Tight para as abordagens nacional e quando se exclui os pacientes do RJ, assim como nas seleções Loose e Medium para os modelos do RJ e CE, respectivamente. Um cenário similar é visto com um desfecho anterior de cura, o qual é incluído nacionalmente e no CE na estratégia Tight, e no Loose para o RJ e quando os pacientes desse estado são excluídos. Já o contato anterior com uma pessoa contaminada com TB-DR, o qual é um fator apontado na literatura como importante para a contaminação de TB-MDR, é incluído em todos os modelos seja nas estratégias Tight ou Medium. Nessa direção, o contato com TB-DS e o número de vezes que o pacientes tratou TB são incluídos na estratégia Tight no modelo do RJ, sendo a última também selecionada para o estado do CE. Entretanto, a existência de tratamento anterior para TB é incluído na estratégia Tight apenas na abordagem nacional. Cavitação pulmonar é outro exemplo de característica que demonstra relevância, presente na estratégia Tight para os modelos nacional, RJ e sem o RJ, e também na estratégia Loose no CE e RS. Porém, nesses dois últimos locais, é importante ressaltar que com estratégias Tight e Medium, não se faz necessária a existência de um aparelho de raio-X na unidade de saúde, o que é um resultado importante para aplicação em locais com pouca infraestrutura. Já a

Tabela 4.8: Desempenho por estratégia de relevância e abordagens (estados e nacional) considerando o índice consolidado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR; TB-DR/MDR e TB-DS; e TB-MDR e TB-DR, utilizando os modelos RNA. O total de neurônios na camada escondida é representado pela coluna #N e o número de variáveis utilizado na coluna #V.

Diagnóstico	Região	Corte	#N	#V	Índice SP (%)	Sensibilidade (%)	Especificidade (%)	Acurácia (%)
TB-DR/MDR x não TB-DR/MDR	Nacional	Tight	10	5	78,3(±2,3)	83,8(±5,9)	73,3(±5,9)	77,6(±2,5)
		Medium	10	10	80,3(±3,3)	85,2(±6,1)	75,9(±7,9)	79,7(±3,7)
		Loose	6	18	78,9(±1,8)	80,3(±6,3)	77,7(±6,4)	78,8(±2,1)
	Não RJ	Tight	2	5	75,9(±2,7)	70,7(±6,8)	81,6(±8,7)	78,8(±2,6)
		Medium	6	9	78,9(±3,7)	76,7(±8,1)	81,6(±9,7)	75,7(±3,6)
		Loose	10	17	82,3(±4,7)	79,6(±8,0)	85,4(±8,5)	82,3(±4,6)
	RJ	Tight	9	4	84,1(±3,7)	88,9(±5,4)	79,5(±4,7)	82,4(±3,6)
		Medium	7	10	87,7(±2,6)	92,3(±5,7)	83,4(±4,7)	86,1(±2,8)
		Loose	9	18	88,7(±3,1)	94,3(±4,4)	83,4(±5,3)	86,7(±3,6)
	CE	Tight	10	4	84,1(±4,8)	83,4(±9,3)	85,3(±9,2)	84,2(±4,7)
		Medium	10	7	90,6(±5,4)	88,8(±6,5)	92,5(±7,2)	90,5(±5,4)
		Loose	9	12	90,7(±3,5)	86,7(±7,5)	95,0(±5,8)	90,7(±3,5)
	RS	Tight	9	4	76,3(±6,3)	78,6(±11,9)	74,5(±10,8)	76,5(±6,3)
		Medium	6	9	85,1(±5,2)	86,6(±9,6)	83,9(±8,2)	85,2(±5,3)
		Loose	5	16	87,7(±5,3)	86,7(±9,4)	89,0(±9,5)	87,7(±5,1)
TB-DR/MDR x TB-DS	Nacional	Tight	2	4	66,6(±4,3)	63,6(±9,7)	70,4(±11,9)	65,3(±5,4)
		Medium	2	8	71,5(±3,3)	61,1(±9,4)	83,5(±10,4)	67,1(±4,9)
		Loose	7	15	76,7(±3,2)	72,3(±9,7)	81,9(±10,9)	74,9(±4,9)
	Não RJ	Tight	2	4	68,2(±5,7)	75,6(±10,4)	62,0(±11,7)	71,9(±6,5)
		Medium	10	11	84,2(±4,2)	82,6(±8,9)	86,2(±8,5)	83,6(±5,5)
		Loose	9	17	85,3(±3,7)	80,7(±7,7)	90,3(±7,7)	83,3(±4,7)
	RJ	Tight	2	4	73,7(±6,6)	73,3(±11,8)	74,9(±13,1)	73,6(±7,7)
		Medium	8	10	85,9(±4,3)	83,5(±9,8)	88,8(±8,6)	84,8(±6,1)
		Loose	5	16	84,7(±5,0)	82,6(±8,3)	87,2(±11,6)	83,7(±5,0)
	CE	Tight	2	4	91,1(±5,6)	84,4(±9,4)	98,3(±4,4)	88,2(±6,9)
		Medium	2	9	96,4(±3,7)	94,7(±6,5)	98,2(±5,0)	95,6(±4,5)
		Loose	2	15	96,9(±3,7)	94,7(±6,4)	99,2(±3,9)	95,8(±4,7)
	RS	Tight	2	4	70,5(±6,9)	71,3(±18,8)	72,0(±17,9)	71,3(±9,0)
		Medium	2	8	83,0(±8,5)	77,0(±10,5)	89,8(±13,3)	81,2(±8,2)
		Loose	2	15	83,7(±7,0)	81,9(±12,8)	86,3(±12,5)	83,3(±7,6)
TB-MDR x TB-DR	Nacional	Tight	10	6	66,5(±3,4)	70,3(±8,9)	63,4(±9,2)	66,9(±3,4)
		Medium	9	11	71,8(±4,1)	70,6(±10,9)	73,9(±12,2)	72,2(±4,0)
		Loose	9	16	74,7(±3,9)	75,5(±10,4)	74,4(±10,0)	75,0(±3,8)
	Não RJ	Tight	2	4	68,0(±6,8)	64,1(±14,8)	73,4(±14,2)	69,1(±6,5)
		Medium	8	9	76,5(±5,4)	83,7(±10,7)	70,3(±11,8)	76,6(±5,3)
		Loose	9	15	80,0(±4,9)	81,7(±9,8)	78,8(±11,0)	80,1(±4,9)
	RJ	Tight	9	4	72,4(±6,9)	68,4(±16,2)	77,9(±14,4)	72,8(±6,9)
		Medium	10	8	82,3(±5,6)	79,9(±12,8)	85,5(±11,3)	82,4(±5,8)
		Loose	7	14	81,9(±6,5)	76,7(±13,0)	87,9(±11,0)	81,6(±6,7)
	CE	Tight	2	4	81,0(±7,1)	85,4(±12,1)	77,1(±10,4)	81,4(±7,0)
		Medium	2	9	89,9(±7,4)	88,0(±11,1)	92,1(±10,8)	89,9(±7,3)
		Loose	5	17	93,0(±6,4)	91,8(±10,1)	94,5(±8,6)	93,0(±6,3)
	RS	Tight	2	4	87,4(±10,9)	86,7(±17,0)	88,6(±11,7)	87,8(±10,7)
		Medium	4	9	89,8(±7,6)	90,2(±12,8)	90,0(±13,1)	89,8(±7,9)
		Loose	5	15	94,5(±6,5)	91,9(±10,4)	97,4(±6,7)	94,6(±6,5)

condição de fumante do paciente só não é incluída com alguma relevância no CE. Ao analisarmos os sintomas, somente febre, perda de apetite e falta de ar são incluídos em estratégia Tight, o que acontece no modelo onde se exclui o RJ. Uma exceção é falta de ar, também vista no modelo do CE. Ao se analisar os resultados para o RS, é possível ver que as cores de pele preta, onde 45,8% possuem TB-DR/MDR, e parda, com 42,1% nesse mesmo diagnóstico, são incluídas na seleção Tight e Medium. Apesar da cor de pele indígena ser incluída na estratégia Tight no modelo do RJ, existe apenas um paciente com essa característica na base de dados. Dessa forma, mais estatísticas sobre essa variável seriam necessárias para se poder fazer maiores considerações. Já o estado civil exibe relevância no CE e RS quando o paciente é casado. Ainda, para pacientes com estadia em prisão ou delegacia, a qual é uma situação que favorece a contaminação, pois as pessoas ficam confinadas em espaços pequenos, é incluída na estratégia Tight e Medium nos modelos do RS e CE, respectivamente. Entretanto, o HIV/AIDS, uma importante comorbidade quando se estuda TB e também uma variável frequentemente utilizada no desenvolvimento de escores, não é incluída em nenhuma estratégia de relevância. Esse fato pode ser explicado pela sua baixa prevalência na base de dados disponível.

Pode-se também comparar os resultados de relevância obtidos por RNA com as variáveis presentes na árvore de classificação gerada pelo CART, as quais são incluídas em sua estrutura por serem decisoras no diagnóstico. Assim, com exceção do delta IMC e quanto tempo o paciente apresenta escarro, todas as variáveis presentes na árvore de classificação do RJ estão incluídas em estratégias Medium ou Tight do modelo RNA. Assim, os modelos parecem concordar sobre as variáveis que são mais discriminantes em um determinado diagnóstico, o que pode validar os resultados obtidos. Entretanto, dificuldades práticas, em termos de operação em uma unidade de saúde, aparecem quando se obtém um modelo CART que se ampara em variáveis com um volume considerável de dados faltantes, como cavitação pulmonar.

As variáveis utilizadas com o algoritmo AdaBoost foram similares as encontradas no modelo CART. Nessa abordagem, a cavitação também foi a variável mais relevante, sendo o modelo ao qual foi atribuído o maior peso. Em seguida, o desfecho



Tabela 4.9: Lista de variáveis incluídas em cada estratégia de desenvolvimento considerando relevância (*Tight*, *Medium* e *Loose*), para as abordagens estaduais e nacional, incluindo quando se exclui os pacientes do RJ. A relevância utilizada é a resultante do índice agregado para o diagnóstico entre TB-DR/MDR e não TB-DR/MDR.

Variáveis	Nacional	Não RJ	RJ	CE	RS
Idade	Medium	-	Loose	-	-
Delta IMC	-	-	-	-	-
Tempo - tosse	-	-	-	-	-
Tempo - escarro	-	-	-	-	-
Quantas vezes trat TB	Medium	-	Tight	Tight	-
Gênero	Medium	-	-	Loose	Loose
Estado civil	-	-	-	-	-
Solteiro	Loose	-	Loose	-	Loose
Casado	-	Medium	-	Tight	Tight
Amigado	-	-	-	-	Loose
Divorciado	-	Loose	-	-	Loose
Separado	-	-	-	Loose	-
Viúvo	-	-	Loose	-	-
Cor da pele	-	-	-	-	-
Branco	Loose	-	-	Loose	Loose
Negro	-	Loose	-	-	Tight
Amarelo	-	-	-	-	-
Pardo	-	Loose	Medium	-	Medium
Indígena	Loose	-	Tight	-	-
Emagrecimento	Loose	Medium	-	-	-
Tosse	Loose	-	-	-	Medium
Expectoração	Loose	-	-	-	-
Catarro com sangue	-	-	Loose	Loose	-
Suor noturno	-	Loose	Medium	-	Medium
Febre	-	Tight	Loose	Medium	Medium
Perda de apetite	-	Tight	Medium	-	-
Falta de ar	-	Tight	-	Tight	-
Trat. anterior TB	Tight	-	Medium	-	-
Desfecho anterior	-	-	-	-	-
Cura	Tight	Loose	Loose	Tight	-
Completo	-	-	-	-	-
Abandono	Medium	Loose	-	-	-
Falecimento	Tight	Tight	Loose	Medium	-
Prisão ou delegacia	Loose	-	Medium	-	Tight
Contato TB-DS	Medium	-	Tight	-	-
Contato TB-DR	Tight	Medium	Medium	Medium	Tight
Fumante	Loose	Medium	-	-	Medium
Ex-fumante	-	Loose	Loose	-	Loose
Cavitação pulmonar	Tight	Tight	Tight	Loose	Loose
Alcoolismo	-	Loose	-	-	-
HIV/AIDS	-	-	-	-	-

do tratamento como falência, outra coincidência entre métodos. Da mesma forma aparecem a quantidade de vezes que tratou TB, perda de apetite e contato anterior com TB-DS. Apesar de ser positivo que os métodos apontem para uma mesma direção em termos de resultado, é visto que tentar diferentes técnicas para geração de

modelos computacionais pode não ajudar a evitar uma variável que possa apresentar problemas em qualidade de dados.

Em geral, problemas relacionados com qualidade de dados e falta de infraestrutura podem ser mitigados com a geração de modelos menos complexos. Por exemplo, ao analisar as variáveis de acordo com sua relevância nas RNA, cavitação pulmonar é incluída apenas na estratégia Loose quando diagnosticando pacientes dos estados do CE e RS. Portanto, mesmo se não houver disponibilidade de equipamento de raio X em unidades de saúde nesses locais, ainda assim os modelos vão conseguir ter um bom desempenho em termos de sensibilidade. Uma situação similar pode ser vista com relação ao contato anterior com TB-DS. Sendo essa uma informação que pode sofrer de validade sobre a resposta dada, é observado que se torna desnecessária por meio das estratégias de seleção nos estados do CE e RS. Já a relevância da variável que implica na condição de fumante, que também depende da validade da resposta do paciente, não a torna dispensável, com exceção de no estado do CE. Nota-se também que o CE possui menos variáveis selecionadas na estratégia Medium do que na Tight. Seguindo o método de seleção de variáveis, a próxima a ser incluída por ordem de relevância seria a cavitação pulmonar. Então, para mitigar os problemas relacionados a falta de estrutura, essa informação foi deixada para ser incluída apenas na estratégia Loose, tornando o modelo mais resiliente a falta de informação.

Características específicas do RJ podem ser comparadas com as do resto do Brasil quando as bases de dados são avaliadas de forma separada. Em ambas as abordagens o desfecho continua relevante, assim como contato anterior com TB-DR. O gênero do paciente também não é relevante, assim como os sintomas de tosse e expectoração. Entretanto, a falta de ar não é relevante no RJ, mas é selecionado na estratégia Tight no modelo que exclui o RJ. A mesma situação ocorre com o emagrecimento, incluído na estratégia Medium. Já tratamento anterior para TB é visto como importante para o estado do RJ, mas não para o restante dos pacientes. Dessa forma, esses são alguns exemplos que mostram que quando se considera uma abordagem nacional, é possível que o desempenho do escore seja afetado por não estar considerando as diferentes características das populações envolvidas. Esse fato

pode ser observado por meio dos diferentes impactos que as variáveis causam nos resultados de cada modelo estudado.

Ao observar variáveis que apresentam baixa relevância de uma forma geral, é possível destacar o sintoma de tosse. Essa variável é normalmente utilizada para ser referência de TB, porém acaba sendo apenas incluída na estratégia Medium no estado do RS e Loose na abordagem nacional. Porém, é observado que 81% dos pacientes disseram que possuíam esse sintoma, algo razoável já que a base de dados é constituída por pacientes que encaminhados com suspeita de TB-DR. Ainda, a variável que representa o tempo em que o paciente possui escarro é excluída de todas as estratégias, assim como expectoração possui baixa relevância nos estados. A mesma situação é vista com o delta IMC. Já o alcoolismo é incluído como relevante quando se excluiu o RJ, mas somente na estratégia Loose. A maioria dos estados civis também mostra baixa relevância, como divorciado, separado, viúvo e amigado. Apesar da alta relevância de alguns desfechos de tratamentos anteriores, a informação que indica que o tratamento apenas foi completado não é incluída em nenhuma estratégia. Isso indica que, para ter um bom indicador de diagnóstico, é preciso possuir mais informações sobre o resultado dos acontecimentos anteriores relacionados a TB.

Em geral, pode-se avaliar que os modelos RNA apresentam um bom desempenho em triagem mesmo em cenários com disponibilidade menor de recursos (estratégias Medium e Loose), sugerindo que é possível discriminar os diagnósticos de TB-DR com menos informação. Ao avaliar a classificação de pacientes entre TB-DR/MDR e não TB-DR/MDR, a sensibilidade dos modelos direcionados aos estados variam de 76,0% até 100%, sendo esse um resultado superior, por exemplo, ao descrito por Boonsarngsuk, et al [74], e Martinez, et al [73], que reportaram sensibilidades para TB-DR de 57,7% e 72,6%, assim como especificidades par TB-MDR de 67,8% e 62,8%, respectivamente.

## 4.4 Estudo de prevalência

Com o objetivo de estudar o impacto da prevalência de TB na população, e dessa forma, entender como o tipo de encaminhamento dado aos pacientes na triagem pode influenciar no cenário de TB-DR, são calculados índices de VPP e VPN para diferentes exemplos. A perspectiva clínica considera um bom resultado quando, dentre 100 indivíduos sem a doença, apenas cinco são encaminhados de forma incorreta para uma avaliação diagnóstica adicional. Isso significa 95% de VPN. Quando se avalia os modelos que discriminam TB-DR/MDR de não TB-DR/MDR para cada estado, cujos resultados são vistos na Figura 4.11, esse cenário é uma realidade em quase todos os casos até uma prevalência de 20%, incluindo quando diferentes estratégias de seleção de variáveis são adotados. A exceção é observada quando se aplica a estratégia Tight para pacientes do RS, mas mesmo assim, os resultados alcançam 93,3% de VPN. Ainda, o modelo do RJ atinge 95% de VPN até uma prevalência de 40% mesmo utilizando a estratégia Medium. Por outro lado, por causa do seu resultado em especificidade, o modelo do CE apresenta um VPP em torno de 80% para uma prevalência de 20%. Portanto, é possível observar que modelos de menor complexidade conseguem atender estados com menos recursos dentro do cenário desejado de VPN, mitigando problemas de qualidade de dados por meio de estratégias Medium e Loose. Ainda, é possível observar as indicações das prevalências da base de dados, para cada estado, em relação a TB-DR e TB-MDR, sendo elas: RJ, TB-DR=13,7% e TB-MDR=17,3%; CE, TB-DR=24,1% e TB-MDR=27,6%; RS, TB-DR=23,1% e TB-MDR=25,6%. Deve-se ressaltar que as prevalências do estudo são maiores do que as apontadas para cada local, pelo fato de que a base de dados aqui utilizada é composta por pacientes já encaminhados para unidade de saúde com suspeita de TB-DR. Apesar disso, é visto que um índice de VPN superior a 95% também é alcançado na maior parte dos casos. De forma complementar, a Tabela 4.10 descreve os valores de VPP e VPN, para cada abordagem e estratégia de relevância, considerando prevalências de 1, 5, 10, 20 e 30% e resultados para modelos baseados em RNA.

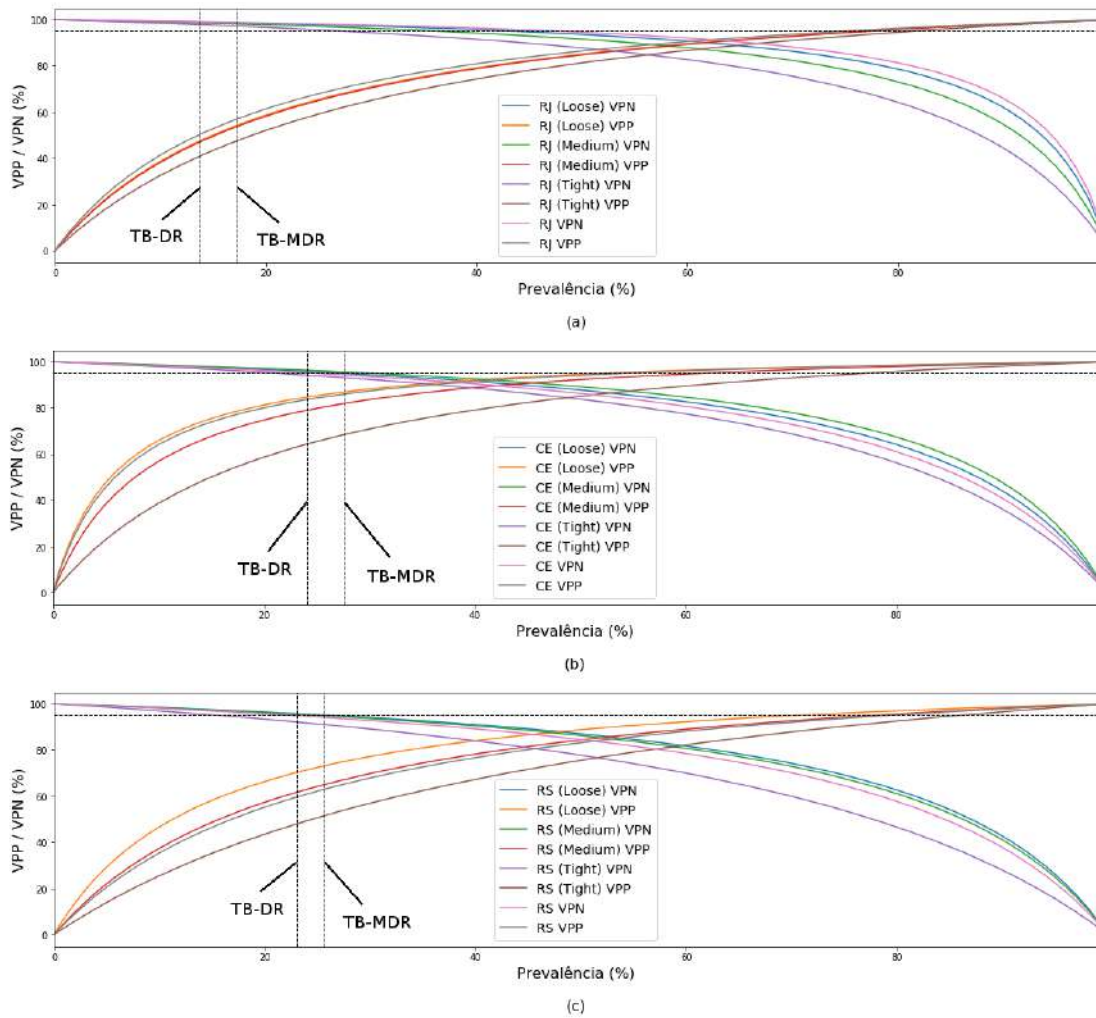


Figura 4.11: Valores de VPP e VPN para triagem de TB-DR relativos a abordagem por estados, de acordo com respectiva prevalência e diferentes estratégias de desenvolvimento de modelos, sendo os resultados para: (a) RJ, (b) CE e (c) RS. A probabilidade de 95% é indicada, assim como as respectivas prevalências de TB-DR e TB-MDR, em relação a base de dados, para cada estado.

Tabela 4.10: Tabela com os valores de VPP e VPN conforme o aumento da prevalência, considerando a abordagem por estados e cada estratégia de relevância adotada. Aqui os resultados são relativos ao diagnóstico de TB-DR e não TB-DR.

	P (%)	Estratégia adotada							
		Todas as variáveis		Loose		Medium		Tight	
		VPP (%)	VPN (%)	VPP (%)	VPN (%)	VPP (%)	VPN (%)	VPP (%)	VPN (%)
RJ	1	6,0	99,9	5,4	99,9	5,3	99,9	4,2	99,9
	5	25,0	99,7	23,0	99,6	22,6	99,5	18,6	99,3
	10	41,3	99,4	38,7	99,2	38,2	99,0	32,5	98,5
	20	61,3	98,6	58,7	98,3	58,2	97,7	52,0	96,6
	30	73,1	97,6	70,9	97,2	70,4	96,2	65,0	94,4
CE	1	13,9	99,8	14,9	99,9	10,7	99,9	5,4	99,8
	5	45,7	99,2	47,7	99,3	38,4	99,4	23,0	99,0
	10	64,0	98,3	65,8	98,5	56,8	98,7	38,7	97,9
	20	80,0	96,2	81,3	96,6	74,7	97,1	58,6	95,4
	30	87,3	93,6	88,1	94,3	83,5	95,1	70,9	92,3
RS	1	4,7	99,8	7,4	99,8	5,2	99,8	3,0	99,7
	5	20,6	99,0	29,3	99,2	22,1	99,2	14,0	98,5
	10	35,4	98,0	46,7	98,4	37,4	98,3	25,5	96,9
	20	55,2	95,6	66,3	96,4	57,4	96,2	43,5	93,3
	30	67,9	92,7	77,2	94,0	69,7	93,6	56,9	89,0

# Capítulo 5

## Conclusões

Neste trabalho, os escores clínicos desenvolvidos por meio de modelos computacionais para triagem alcançaram resultados promissores para o diagnóstico de pacientes suspeitos de TB-DR. As altas probabilidades de sensibilidade atingidas pelos modelos, especialmente quando aplicadas técnicas de RNA, indicam que os escores devem ser estudados prospectivamente de forma complementar a novos testes moleculares para diagnóstico, e/ou radiografia de tórax, avaliando, com isso, o seu impacto clínico e o seu custo-benefício. Os diagnósticos que apresentaram maior aplicabilidade para triagem foram os que discriminaram TB-DR/MDR de não TB-DR/MDR e TB-DR/MDR de TB-DS. O resultado da classificação de TB-MDR entre pacientes com TB-DR é considerado incerto, dado a incerteza associada.

Em geral, os modelos direcionados a uma localidade específica, no caso, a um estado do Brasil, apresentaram resultados superiores ao de uma abordagem nacional. Entretanto, ainda se levanta a possibilidade dos modelos estarem capturando as características operacionais de cada local, e não de sua população. Apesar do RJ ser o estado mais representativo em termos de dados disponíveis, os outros estados também indicaram o mesmo resultado quando comparados aos modelos desenvolvidos ao se agrupar todos os pacientes sem distinção. Diferentes estratégias podem ser adotadas no desenvolvimento de escores, levando a construção de modelos com menor complexidade em relação ao número de variáveis necessárias em uma triagem. Para isso, as estratégias adotadas se apoiam na relevância das variáveis e na

qualidade da mesmas, a qual é modelada em 4 dimensões (dados faltantes, validade, acessibilidade e necessidade de infraestrutura). Mesmo com menos informações, esses modelos conseguem alcançar um desempenho similar aos que utilizam toda a base de dados disponível. Ainda, ao diminuir o número de variáveis, problemas relacionados a qualidade de dados podem ser mitigados.

Ainda, a partir do desenvolvimento de modelos de menor complexidade, é possível, por meio da análise de relevância das variáveis, obter informações sobre qual característica é a mais importante em um diagnóstico de TB-DR/MDR em diferentes regiões/estados. Assim, variáveis relevantes podem ser ressaltadas e o impacto das características de cada local levados em consideração ao se avaliar pacientes em triagem. Entretanto, resultados que apontam variáveis como cor da pele como importantes para a triagem ainda devem ser mais aprofundados, dado limitações da base de dados utilizada neste trabalho. As diferenças encontradas indicam que modelos desenvolvidos para um estado podem não ser capazes de serem utilizados na triagem de um estado vizinho, mesmo sendo de uma mesma região. De forma geral, tais resultados podem apoiar a decisão clínica e complementar o conhecimento obtido sobre variáveis marcadoras de resistência por meio da literatura médica. Modelos computacionais são baratos em termos de estrutura operacional (não é necessário adquirir uma licença de software), e podem ser implantado em dispositivos móveis. Os resultados podem ser obtidos no momento em que os pacientes estão em fase de triagem, teste *point of care*, com baixa complexidade em relação a entrada de dados. Além disso, permite que ambientes com poucos recursos possam ser apoiados pelo escore desenvolvido.

Apesar dos novos diagnósticos moleculares rápidos (o que inclui o Xpert MTB RIF) mostrarem resultados promissores e simples de usar, eles ainda são caros, e múltiplos espécimes podem ser necessários para atingirem uma sensibilidade adequada em casos que apresentem exame de baciloscopia negativo. A utilização da abordagem aqui proposta para realizar a triagem, mesmo em regiões onde o Xpert MTB tenha sido implementado, pode diminuir o tempo necessário para o resultado diagnóstico e o tratamento empírico, favorecendo a gestão apropriada dos pacientes



no sistema de saúde. Nota-se ainda, que no Rio de Janeiro, uma alta taxa de pacientes suspeitos de TB-DR, e que não possuem TB ativa, acabam sendo enviados para centros de referência. Portanto, a abordagem aqui proposta para a triagem de pacientes suspeitos de TB-DR mostra vantagem em relação aos procedimentos hoje existentes, possibilitando corrigir esse erro no processo. Assim, os modelos consideram esse possível fluxo de pacientes e a qualidade das amostras a serem avaliadas nas unidades médicas locais. Apesar de pacientes com diagnóstico de não TB poderem ser encaminhados para as unidades de referência já com algum tratamento iniciado, o que pode influenciar sua situação no momento de triagem, os modelos se mostram robustos a essa situação.

Apesar das diferentes abordagens adotadas no desenvolvimento dos modelos computacionais, observa-se que as incertezas associadas aos resultados ainda são consideradas altas, em sua maioria devido ao baixo número de pacientes. No futuro, a coleta de mais dados em cada região estudada é vista como uma ação com potencial de apoiar a melhora dos valores associados à incertezas. Isso pode ajudar, especialmente, nos estados com menos amostras de dados, como RS e SP, e no diagnóstico entre TB-MDR e TB-DR. Essa iniciativa é ainda mais incentivada a partir dos resultados promissores já alcançados. A inclusão de outros estados também pode ser feita em projetos futuros, como Amazonas e Bahia, locais de alta prevalência. Outra abordagem possível é a ampliação do estudo para o nível de macrorregiões do Brasil, ou então municípios com características similares. Além disso, incluir a informação, na triagem, de dados operacionais dos estados, como indicadores de programas de saúde da família, ou que possam apontar uma possível falência do sistema de saúde local (alta taxa de mortalidade), pode ajudar no desenvolvimento de um modelo nacional. Dessa forma, os modelos poderiam se adaptar a locais com características similares em termos de sistema de saúde. Métodos para apoiar a classificação de pacientes em grupos de risco são considerados úteis e podem ser desenvolvidos para serem incluídos como uma avaliação seguinte na etapa de triagem. Treinamentos dos modelos com o objetivo de alcançar altas taxas de sensibilidade e especificidade, também podem ser estudados, assim como avaliar a fusão dos resultados obtidos

por diferentes métodos computacionais. Em termos da aplicação, em unidades de saúde, o desenvolvimento de um aplicativo para dispositivo móvel pode ser realizado visando a experimentação do modelo em cenários de operação em triagem.

# Referências Bibliográficas

- [1] THEODORIDIS, S., KOUTROUMBAS, K. “Pattern Recognition”, *Academic Press*, v. 1999.
- [2] DUDA, R. O., HART, P. E., STORK, D. G. “Pattern Classification”, *Wiley-Interscience*, v. 2000.
- [3] WANG, R. Y., STRONG, D. M. “Journal of Management Information Systems”, *Volume 12 - Issue 4*, v. 1996.
- [4] DENSEN, P. “Challenges and Opportunities Facing Medical Education”, *Transactions of the American Clinical and Climatological Association*, v. 122, 48–58. 2011.
- [5] BERNER, E. S. “Clinical decision support systems”, *New York: Springer Science+ Business Media, LLC*, v. 2007.
- [6] WHO. *Global Tuberculosis Report 2017*. World Health Organization, WHO, 2017.
- [7] WHO. *Supplement Global Tuberculosis Report-Drug-Resistant TB Surveillance and Response*. WHO, 2014.
- [8] FERREIRA, A.A., QUEIROZ, K.C., TORRES, K.P., FERREIRA, M.A., ACIOLY, H., ALVES, M.S. *Os fatores associados à tuberculose pulmonar e a baciloscopia: uma contribuição ao diagnóstico nos serviços de saúde pública*. Rev Bras Epidemiol 2005; 8(2): 142-9., 2005.
- [9] WHO. *The global plan to stop TB 2011-2015*. WHO, 2006.
- [10] TRÉBUCQ, A., ENARSON, D.A., CHIANG, C.Y., VAN DEUN, A., HARRIES, A.D., BOILLOT, F., DETJEN, A., FUJIWARA, P.I., GRAHAM, S.M., MONEDERO, I., RUSEN, I.D., RIEDER, H.L. “Xpert® MTB/RIF for national tuberculosis programmes in low-income countries: when, where and how?” *Int J Tuberc Lung Dis* 15, v. doi: 10.5588/ijtld.11.0392. Epub 2011 Oct 13., n. 12, pp. 1567–72, Dec 2011.

- [11] SECRETARIA DE VIGILÂNCIA EM SAÚDE. “Boletim Epidemiológico”, *Ministério da Saúde*, v. volume 47, N 13 -2016, .
- [12] SECRETARIA DE VIGILÂNCIA EM SAÚDE. “Boletim Epidemiológico”, *Ministério da Saúde*, v. volume 48, N 8 -2017, .
- [13] LANGLEY, I., SQUIRE, S.B., DACOMBE, R., MADAN, J., LAPA E SILVA, J.R., GALIEZ, R., OLIVEIRA, M.M., KRITSKI, A. “Developments in impact assessment of new diagnostic algorithms for TB control”, *Clinical Infectious Disease* 15, v. doi: 10.1093/cid/civ580, 2015, n. 61, pp. Suppl 3:S126–34, Oct 2015.
- [14] VAN’T HOOG, A.H., COBELENS, F., VASSALL, A., VAN KAMPEN, S., DORMAN, S.E., ALLAND, D., ELLNER, J. “Optimal triage test characteristics to improve the cost-effectiveness of the Xpert MTB/RIF assay for TB diagnosis: a decision analysis”, *PLoS One*, v. doi: 10.1371/journal.pone.0082786. eCollection 2013, n. 18;8(12):e82786, Dec 2013.
- [15] WHO. *The end TB strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015*. WHO, 2015.
- [16] DIRLIKOV, E., RAVIGLIONE, M., SCANO, F. “Global Tuberculosis Control: Toward the 2015 Targets and Beyond”, *Ann Intern Med* 2015; 163(1):52-58.
- [17] UPLEKAR, M., WEIL, D., LONNROTH, K., JARAMILLO, E., LIENHARDT, C., DIAS, H.M., FALZON, D., FLOYD, K., GARGIONI, G., GETAHUN, H., GILPIN, C., GLAZIOU, P., GRZEMSKA, M., MIRZAYEV, F., NAKATANI, H., RAVIGLIONE, M. “WHO’s Global TB Programme. WHO’s new end TB strategy”, *Lancet*, v. 385(9979):1799-801, 2015.
- [18] HORNIK, K. “Approximation capabilities of multilayer feedforward networks”, *Neural networks* 4.2, v. (1991): 251-257.
- [19] BREIMAN, L., FRIEDMAN, J., STONE, C.J., OLSHEN, R.A. “Classification and regression trees”, *CRC press*, v. 1984.
- [20] FREUND, Y., SCHAPIRE, R., ABE, N. “A short introduction to boosting”, *Journal-Japanese Society For Artificial Intelligence*, v. 14(771-780), 1612, 1999.

- [21] MINISTÉRIO DA SAÚDE, SECRETARIA DE VIGILÂNCIA EM SAÚDE - BRASILIA. *Manual de recomendações para o controle da Tuberculose no Brasil*. Ministério da Saúde, 2011.
- [22] LOPES, A. J., CAPONE, D., MOGAMI, R., et al. *Tuberculose extrapulmonar: aspectos clínicos e de imagem*. Pulmão RJ 2006;15(4):253-261, 2006.
- [23] WORLD HEALTH ORGANIZATION. *Guidelines for the Prevention of Tuberculosis in Health Care Facilities in Resource-Limited Settings*. WHO, 1999.
- [24] CENTERS FOR DISEASE CONTROL AND PREVENTION - CDC. *Guidelines for preventing the transmission of Mycobacterium tuberculosis in health care facilities*. MMWR 1994;43(RR-13):1-131, 1994.
- [25] “Global Health Observatory, Map Gallery”.  
Acessado em 23 Fev. 2017.  
<http://gamapserv.who.int/mapLibrary/app/searchResults.aspx>,  
.
- [26] MINISTÉRIO DA SAÚDE. *Tratamento diretamente observado da tuberculose na atenção básica*. 2011.
- [27] SCHIRM, J., OOSTENDORP, L. A., MULDER, J. G. *Comparasion of amplicolor, in house PCR and conventional culture for detection of mycobacterium in clinical samples*. Journal Clinical Microbiology, pp. 3321–3324, 1995.
- [28] A, V. R., PAGE-SHIPPI, L. S., et al. *Xpert(®) MTB/RIF for point-of-care diagnosis of TB in high-HIV burden, resource-limited countries: hype or hope?* Expert Rev Mol Diagn. 2010; 10(7):937-46., 2010.
- [29] GLOBAL LABORATORY INITIATIVE/WHO. *Mycobacteriology Laboratory Manual*. 2014.
- [30] CONDE, M.B., MELLO, F., LIMA, M.A., GUERRA, R.L., MIRANDA, S.S., GALVÃO, T.S., PINHEIRO, V.G., CARVALHO, N.B. *Tuberculose Pulmonar: Diagnóstico - Técnicas Convencionais*. Sociedade Brasileira de Pneumologia e Tisiologia, 2011.
- [31] COULTER, J.B. *Diagnosis of pulmonary tuberculosis in young children*. Ann Trop Paediatr 2008;28:3-12., 2008.
- [32] WHO. *Implementing the end TB strategy: the essentials*. WHO, 2015.

- [33] WHO. *Companion handbook to the WHO guidelines for the programmatic management of drug-resistant tuberculosis*. WHO, 2014.
- [34] BARROSO, E. C., MOTA, R. S., SANTOS, R. O., et al. *Fatores de risco para tuberculose multirresistente adquirida*. J Pneumol 2003;29(2):89-97, 2003.
- [35] DE SOUZA, M. B., DE FIGUEIREDO ANTUNES, C. M., GARCIA, G. F. *Perfil de sensibilidade e fatores de risco associados à resistência do Mycobacterium tuberculosis, em centro de referência de doenças infecto-contagiosas de Minas Gerais*. J. bras. pneumol. vol.32 no.5 São Paulo Set./Out, 2006.
- [36] ARBEX, M.A. “O desafio do tratamento da tuberculose extensivamente resistente em um hospital de referência no estado de São Paulo: um relato de três casos”, *Jornal Brasileiro de Pneumologia*, v. 41, n. 6, p. 554-559, v. 2015.
- [37] SECRETARIA DE VIGILÂNCIA EM SAÚDE, MINISTÉRIO DA SAÚDE. *Programa Nacional de Controle da Tuberculose*. Ministério da Saúde, 2017.
- [38] SECRETARIA DE VIGILÂNCIA EM SAÚDE. “Rede de Teste Rápido para Tuberculose no Brasil - Primeiro ano da implantação”, *Ministério da Saúde*, v. Brasília, DF, 2015, .
- [39] “Programa Nacional de Controle da Tuberculose. CGPNCT/DEVIT. Secretaria de Vigilância em Saúde, Ministério da Saúde”.  
Acessado em 14 Fev. 2018.  
<http://portalarquivos.saude.gov.br/images/pdf/2017/fevereiro/21/Apresentacao-sobre-os-principais-indicadores-da-tuberculose.pdf>, .
- [40] SECRETARIA DE VIGILÂNCIA EM SAÚDE. “Panorama da Tuberculose no Brasil: A mortalidade em números”, *Ministério da Saúde*, v. Brasília, DF, 2016, .
- [41] BAXT, W.G., SHOFER, F.S., SITES, F.D., HOLLANDER, J.E. “A neural network aid for the early diagnosis of cardiac ischemia in patients presenting to the emergency department with chest pain”, *Annals of emergency medicine*, v. 40(6), 575-583, 2002.
- [42] BASSAN, R., PIMENTA, L., SCOFANO, M., SOARES, J.F. “Accuracy of a neural diagnostic tree for the identification of acute coronary syndrome in

patients with chest pain and no ST-segment elevation”, *Critical Pathways in Cardiology*, v. 3(2), 72-78, 2004.

- [43] DAS, A. BEN-MENACHEM, T., FAROOQ, F.T., COOPER, G.S., CHAK, A., SIVAK, M.V., WONG, R.C. “Artificial neural network as a predictive instrument in patients with acute nonvariceal upper gastrointestinal hemorrhage”, *Gastroenterology*, v. 134(1), 65-74, 2008.
- [44] EFTEKHAR, B., MOHAMMAD, K., ARDEBILI, H., GHODSI, M., KETABCHI, E. “Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data”, *BMC Medical Informatics and Decision Making*, v. 5(1), 1, 2005.
- [45] GUILLERMO, J.E., ET AL. “Detection of heart murmurs based on radial wavelet neural network with Kalman learning”, *Neurocomputing*, v. 164 (2015): 307-317.
- [46] NANNI, L. “A reliable method for the diagnosis of gastric carcinoma”, *Neurocomputing*, v. 69(7), 862-865, 2006.
- [47] PEARL, A., BAR-OR, R., BAR-OR, D. “An artificial neural network derived trauma outcome prediction score as an aid to triage for non-clinicians”, *Studies in health technology and informatics*, v. 136, 253, 2008.
- [48] FORTHOFFER, M.S., BRYANT, C.A. “Using audience-segmentation techniques to tailor health behavior change strategies”, *American Journal of Health Behavior*, v. 24(1), 36-43.
- [49] LEMON, S.C., ROY, J., CLARK, M.A., FRIEDMANN, P.D., RAKOWSKI, W. “Classification and regression tree analysis in public health: methodological review and comparison with logistic regression”, *Ann Behav Med* 2003;26:172-81, v. 2003.
- [50] KORB, K.B., NICHOLSO, A.E. *Bayesian Artificial Intelligence*. Chapman Hall /CRC, 2003.
- [51] MITCHELL, T. *Machine Learning*. McGraw Hill, 1997.
- [52] HAYKIN, S. *Neural Networks and Learning Machines*. Prentice-Hall, Inc., 2008.
- [53] WEST, D., MANGIAMELI, P., RAMPAL, R., WEST, V. “Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application”, *European Journal of Operational Research*, v. 162(2), 532-551.

- [54] RIDGEWAY, G. “The state of boosting”, *Computing Science and Statistics*, v. p. 172-181, 1999, .
- [55] DRUCKER, H., CORTES, C. “Advances in Neural Information Processing Systems”, 1996, v. pp. 479–485.
- [56] BAUER, E., KOHAVI, R. “Machine Learning”, 1999, v. 36, 105–139.
- [57] ZHOU, J. “Comparing regularized B-spline neural network, multilayer perceptron and boosted-CART on two problems of heart arrhythmia diagnosis”, *In: Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, v. IEEE, 2004. p. 3601-3604.
- [58] TAN, C., CHEN, H., XIA, C. “The prediction of cardiovascular disease based on trace element contents in hair and a classifier of boosting decision stumps”, *Biological trace element research*, v. v. 129, n. 1-3, p. 9-19, 2009, .
- [59] TAN, C., CHEN, H., XIA, C. “Early prediction of lung cancer based on the combination of trace element analysis in urine and an Adaboost algorithm”, *Journal of pharmaceutical and biomedical analysis*, v. v. 49, n. 3, p. 746-752, 2009, .
- [60] HRDLICKA, J., KLEMA, J. “Schizophrenia prediction with the adaboost algorithm”, *Studies in health technology and informatics*, v. v. 169, p. 574-578, 2011.
- [61] SOLARI, L., ACUNA-VILLAORDUNA, D., SOTO, A., AGAPITO, J., PERES, F., SAMALVIDES, F., ET AL. “A clinical prediction rule for pulmonary tuberculosis in emergency departments”, *Int J Tuberc Lung Dis*, v. 12(6):619–624, 2008.
- [62] MELLO, F.C., BASTOS, L.G., SOARES, S.L., REZENDE, V.M., CONDE, M.B., CHAISSON, R.E., ET AL. “Predicting smear negative pulmonary tuberculosis with classification trees and logistic regression: a cross-sectional study”, *BMC Public Health*, v. Feb 23;6:43, 2006.
- [63] AGUIAR, F.S., ALMEIDA, L.L., RUFFINO-NETTO, A., KRITSKI, A.L., MELLO, F.C.Q., WERNECK, G.L. “Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients”, *BMC Pulmonary Medicine*, v. 12:40, doi:10.1186/1471-2466-12-40, 2012.



- [64] WISNIVESKY, J.P., SEREBRISKY, D., MOORE, D., SACKS, H.S., IANNUZZI, M.C., MCGINN, T. “Validity of Clinical Prediction Rules for Isolating Inpatients with Suspected Tuberculosis. A Systematic Review”, *J Gen Intern Med*, v. 20:947–952, 2005.
- [65] DE O. SOUZA FILHO, J.B., DE SEIXAS, J.M., GALLIEZ, R., PEREIRA, B. DE B., DE Q MELLO, F.C., DOS SANTOS, A.M., KRITSKI, A.L. “A Screening System for Smear-Negative Pulmonary Tuberculosis using Artificial Neural Networks”, *International Journal of Infectious Diseases*, v. v. 49, p. 33-39, 2016.
- [66] EL-SOLH, A.A., HSIAO, C.B., GOODNOUGH, S., SERGHANI, J., GRANT, B.J. “Predicting active pulmonary tuberculosis using an artificial neural network”, *Chest*, v. Oct;116(4):968-73, 1999.
- [67] ER, O., TEMURTAS, F., TANRIKULU, A.C. “Tuberculosis disease diagnosis using artificial neural networks”, *J Med Syst*, v. Jun;34(3):299-302, 2010.
- [68] ELVEREN, E., YUMUSAK, N. “Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm”, *J Med Syst*, v. Jun;35(3):329-32, 2011.
- [69] CASÇÃO, L. V. C. *Modelos de inteligência computacional para apoio da triagem de pacientes e diagnóstico clínico de tuberculose pulmonar*. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2011.
- [70] WELLING, M. “Fisher linear discriminant analysis”, *Department of Computer Science*, v. University of Toronto, v. 3, n. 1, 2005.
- [71] SEIXAS, J.M., FARIA, J., SOUZA FILHO, J.B.O., VIEIRA, A.F.M., KRITSKI, A., TRAJMAN, A. “Artificial Neural Network Models for Diagnosis Support of Pleural Tuberculosis in Adult Patients”, *I J Tuberc Lung Dis*, v. 17 (5): 682-686, 2013.
- [72] AGUIAR, F.S., TORRES, R.C., PINTO, J.V., KRITSKI, A.L., SEIXAS, J.M., MELLO, F.C. “Development of two artificial neural network models to support the diagnosis of pulmonary tuberculosis in hospitalized patients in Rio de Janeiro, Brazil”, *Medical and biological engineering and computing*, v. 54(11), 1751-1759, 2016.
- [73] MARTINEZ, D., HEUDEBERT, G., SEAS, C., ET AL. “Clinical Prediction Rule for Stratifying Risk of Pulmonary Multidrug-Resistant Tuberculosis”, *PLoS One*, v. Aug 11;5(8):e12082. doi: 10.1371/journal.pone.0012082, 2010.

- [74] BOONSARNGSUK, V., TANSIRICHAIIYA, K., KIATBOONSRI, S. “Thai drug-resistant tuberculosis predictive scores”, *Singapore Med J*, v. 50(4): 378-384, 2009.
- [75] VILLA, T. C., RUFFINO NETTO, A., SCATENA L. M., ANDRADE, R. L., ET AL. “Health services performance for TB treatment in Brazil: a cross-sectional study”, *BMC Health Serv Res*, v. 2011 Sep 28;11(1):241.
- [76] DANDE, P., SAMANT, P. “Acquaintance to Artificial Neural Networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review”, *Tuberculosis*, v. v. 108, p. 1-9, 2018.
- [77] “AI known as CRyPTIC can diagnose drug-resistant tuberculosis in minutes”, *[Citado Fev. 2018]*, v. Disponível em: <http://today.mims.com/topic/ai-known-as-cryptic-can-diagnose-drug-resistant-tuberculosis-in-minutes/>, .
- [78] “Gene-reading software to cut TB diagnosis from months to minutes”, *[Citado Fev. 2018]*, v. Disponível em: <https://www.newscientist.com/article/mg23130923-000-genereading-software-to-cut-tb-diagnosis-from-months-to-minutes/>, .
- [79] BURGESS, M. “The NHS is trialling an AI chatbot to answer your medical questions.” *[Citado Fev. 2018]*, v. Disponível em: [www.wired.co.uk/article/babylon-nhs-chatbot-app](http://www.wired.co.uk/article/babylon-nhs-chatbot-app).
- [80] ÉVORA. L. H. R. A., SEIXAS, J. M., KRITSKI, A. L. “Neural network models for supporting drug and multidrug resistant tuberculosis screening diagnosis”, *Neurocomputing*, v. v. 265, p. 116-126, 2017.
- [81] STONE, M. “Cross-validators: choice and assessment of statistical predictions”, *Journal of the Royal Statistical Society*, v. B26 pp 111-147, 1974.
- [82] HAYKIN, S. *Neural Networks and Learning Machines*. Prentice-Hall, Inc., 2011.
- [83] HERTIGAN, A., WONG, M. A. “A K-means Clustering Algorithm: Algorithm AS 136”, *Journal of the Royal Statistical Society*, v. Series C (Applied Statistics), Vol. 28, No. 1 (1979), pp. 100-108 , DOI: 10.2307/2346830, 1979.
- [84] DAVIES, D.L., BOULDIN, D.W. “A Cluster Separation Measure, Pattern Analysis and Machine Intelligence”, *IEEE Transactions on Volume:PAMI-1*, v. Issue: 2, , pp 224 - 227, DOI: 10.1109/TPAMI.1979.4766909, 1979.

- [85] COOK, E.F., GOLDMAN, L. “Asymmetric stratification an outline for an efficient method for controlling confounding in cohort studies”, *American journal of epidemiology*, v. 127(3), 626-639, 1988.
- [86] SANTOS, A.M., PEREIRA, B.B., SEIXAS, J.M., MELLO, F.C.Q., KRITSKI, A.L. “Árvore de Classificação e Redes Neurais Artificiais: Uma Aplicação à Predição de Tuberculose Pulmonar”, *Proceedings of the VI Brazilian Conference on Neural Networks - VI Congresso Brasileiro de Redes Neurais pp.427-432*, v. 2003.
- [87] MELLO, F.C., BASTOS, L.G., SOARES, S.L., REZENDE, V.M., CONDE, M.B., CHAISSON, R.E., KRITSKI, A.L., RUFFINO-NETTO, A., WERNECK, G.L. “Predicting smear negative pulmonary tuberculosis with classification trees and logistic regression: a cross-sectional study”, *BMC Public Health. 2006 Feb 23;6:43. PubMed PMID: 16504086; PubMed Central PMCID:PMC1402281*, v. 2006.
- [88] HORNIK, K. *Approximation capabilities of multilayer feedforward networks*. Neural networks 4.2, (1991): 251-257, 1991.
- [89] LEK, S., GIRAUDEL, J. L., GUÉGAN, J. F., . *Neuronal networks: algorithms and architectures for ecologists and evolutionary ecologists*. Artificial Neuronal Networks, (pp. 3-27). Springer Berlin Heidelberg, 2000.
- [90] RIEDMILLER, M., BRAUN, H. “A Direct Adaptive Method for Faster Back-propagation Learning: the RPROP Algorithm”, *Proceedings of the IEEE Conference on Neural Networks*, v. pp.586-591, ISBN 0-7803-0999-5, San Francisco, California, USA, 28 Mar ch 28 - April 01, 1993.
- [91] GROBMAN, W.A., STAMILIO, D.M. “Methods of clinical prediction”, *Am J Obstet Gynecol. 2006 Mar;194(3):888-94*, v. 2006.
- [92] ROKACH, L., MAIMON, O. “Data mining with decision trees: theory and applications”, *World scientific*, v. 2014.
- [93] ATKINSON, E.J., THERNEAU, T.M. “An introduction to recursive partitioning using the RPART routines”, *Rochester: Mayo Foundation*, v. 2000.
- [94] RODRIGUES, M.A.S. “Árvores de Classificação”, *Monografias da SEIO. Depto. Matemática da Univ. dos Açores: Ponta Delgada, www.uac.pt/ amendes (ID 54.218)*, v. 2006.

- [95] BRADFORD, J.P., KUNZ, C., KOHAVI, R., BRUNK, C., BRODLEY, C.E. “Pruning decision trees with misclassification costs”, *Machine Learning: ECML-98*, v. (pp. 131-136). Springer Berlin Heidelberg, 1998.
- [96] MINGERS, J. “Machine Learning”, 1989, v. 4, 227-243.
- [97] RIDGEWAY, G. “The state of boosting”, *Computing Science and Statistics*, v. p. 172-181, 1999, .
- [98] SCHAPIRE, R. E. “Explaining adaboost”, *Empirical inference. Springer Berlin Heidelberg*, v. 2013. p. 37-52.
- [99] SIMAS FILHO, E. *Análise Não-Linear de Componentes Independentes para uma Filtragem Online Baseada em Calorimetria de Alta Energia e com Fina Segmentação*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, 2010.
- [100] CIODARO, T., DEVA, D., DE SEIXAS, J., ET AL. *Online particle detection with Neural Networks based on topological calorimetry information*. In: *Journal of Physics: Conference Series*, v. 368, p. 012030. IOP Publishing, 2012.
- [101] ALTMAN, D. G., BLAND, J. M. *Diagnostic tests 3: receiver operating characteristic plots*. *BMJ: British Medical Journal*, 1994. Jul 16;309(6948):188. PubMed PMID: 8044101, PubMed Central PMCID: PMC2540706.
- [102] GROBMAN W. A., STAMILIO, D. M. *Methods of clinical prediction*. *Am J Obstet Gynecol*, 2006 Mar; 194(3):888-94.
- [103] KAWAMURA, T. “Interpretação de um teste sob a visão epidemiológica: eficiência de um teste.” *Arquivos Brasileiros de Cardiologia*, v. v.79, n.4, p.437-441, 2002.
- [104] MICHELETTI, V.C.D., MOREIRA, J.D.S., RIBEIRO, M.O., KRITSKI, A.L., BRAGA, J.U. “Drug-resistant tuberculosis in subjects included in the Second National Survey on Antituberculosis Drug Resistance in Porto Alegre, Brazil”, *Jornal Brasileiro de Pneumologia*, v. 40(2), 2014, 155-163.
- [105] BRITO, C.R., MELLO, F.C.Q., ANDRADE, M.K., OLIVEIRA, H., COSTA, W., ET AL. “Drug-resistant tuberculosis in six hospitals in Rio de Janeiro, Brazil”, *The international journal of tuberculosis and lung disease*, 14(1), 24-33, v. 2010.

- [106] DENKINGER, C. M., ET AL. “Defining the needs for next generation assays for tuberculosis”, *The Journal of infectious diseases*, v. v. 211, n. suppl2, p. S29-S38, 2015.
- [107] EURTS, P., IRRTHUM, A., WEHENKEL, L. “Supervised learning with decision tree-based methods in computational and systems biology”, *Molecular Biosystems*, v. v. 5, n. 12, p. 1593-1605, 2009.
- [108] GEVREY, M., DIMOPOULOS, I., LEK, S. “Review and comparison of methods to study the contribution of variables in artificial neural network models”, *Ecological modelling* 160.3, v. 249-264, 2003.
- [109] XAVIER, T.C., DOS ANJOS RABELLO, A., DE SEIXAS, J.M. “Discriminadores neurais de partículas para um detector submetido a uma alta taxa de eventos”, *VIII Congresso Brasileiro de Redes Neurais*, v. 2007.
- [110] GARSON, G.D. “Interpreting neural network connection weights”, *Artificial Intelligence Expert* 6, v. 47-51, 1991.
- [111] GOH, A.T.C. “Back-propagation neural networks for modeling complex systems”, *Artificial Intelligence in Engineering*, v. 9, 143-151, 1995.
- [112] ZHOU, B., VOGT, R. D., LU, X., XU, C., ZHU, L., SHAO, X., ET AL. “Relative Importance Analysis of a Refined Multi-parameter Phosphorus Index Employed in a Strongly Agriculturally Influenced Watershed”, *Water, Air, and Soil Pollution*, v. 226(3), 1-13, 2015.
- [113] BECK, M.W., WILSON, B.N., VONDRACEK, B., HATCH, L.K. “Application of neural networks to quantify the utility of indices of biotic integrity for biological monitoring”, *Ecological Indicators*, v. 45, 195-208, 2014.
- [114] SUNG, A.H. “Ranking importance of input parameters of neural networks”, *Expert Systems with Applications* 15, v. 405-411, 1998.
- [115] BARTHOLOMAY OLIVEIRA, P., PINTO DE OLIVEIRA, G., CODENOTTI, S. B., SARACENI, V.; DA NÓBREGA, A. A.; SOBEL, J. “Avaliação do sistema de vigilância da tuberculose no município do Rio de Janeiro”, *Caderno de Saúde Coletiva*, v. Rio de Janeiro, 18 (3): 337-46, 2010.
- [116] SILVA, G. D. M., ET AL. “Avaliação da qualidade dos dados, oportunidade e aceitabilidade da vigilância da tuberculose nas microrregiões do Brasil”, *Ciência e Saúde Coletiva*, v. v. 22, n. 10, p. 3307-3319, 2017.

# Apêndice A

## Publicações Relativas à Tese

Abaixo são listadas as publicações produzidas durante a realização deste trabalho, incluindo o resumo correspondente a cada uma das produções citadas.

ÉVORA, L.H.R.A, SEIXAS, J.M., KRITSKI, A.L., "Neural Network Models for Supporting Drug and Multidrug Resistant Tuberculosis Screening Diagnosis". *Neurocomputing*, v. 265, p. 116-126, 2017.

**Abstract:** Tuberculosis (TB) is the leading cause of global mortality among communicable diseases. The diagnosis of Drug-Resistant Tuberculosis (DR-TB) demands even more attention, leading to longer treatments and higher deceased rates. All diagnostic methods available have deficiencies in their detection rates, time release results, or have a higher cost and need a complex infrastructure to setup. New molecular diagnostics, such as the Xpert MTB/RIF assay, have great potential for revolutionizing the diagnosis of Rifampicin Resistance (RR). But, a positive RR result with this test should be carefully interpreted and take into consideration the risk of Multidrug-Resistant TB (MDR-TB) according to its prevalence, locally. Therefore, the development of screening approaches for DR/MDR-TB suspects would help to identify those should be tested by Xpert MTB/RIF. This work develops Artificial Neural Network (ANN) models considering data from presumed DR/MDR-TB subjects according to the National Guidelines at Rio de Janeiro/Brazil, attended in reference centers in Rio de Janeiro, from Feb 2011 and May 2013. Were evaluated subjects aged 18 years or older, and results were compared with models based on

Classification And Regression Trees (CART). Practical operation at different epidemiological scenarios are considered by constructing models using different variable selection criteria, so that environments with low resource conditions can be assisted. Among 280 presumed DR-TB cases included, 38 were DR-TB, 48-MDR, 32-Drug-Sensitive and 162 with no TB. Between DR-TB and non DR-TB, the sensitivity and specificity reached 95.1%(±5.0) and 85.0%(±4.9), respectively. The promising results of clinical score for DR/MDR-TB diagnosis indicate that this approach can be used in the evaluation of presumed DR/MDR-TB.

ÉVORA, L.H.R.A, SEIXAS, J.M., KRITSKI, A.L., "Artificial neural network models for diagnosis support of drug and multidrug resistant tuberculosis". *In2015 Latin America Congress on Computational Intelligence (LA-CCI)*, (pp. 1-5). IEEE. 2015.

**Abstract:** Tuberculosis (TB) remains as one of the major health problems in the world. The diagnosis of drug-resistant tuberculosis demands even more attention, leading to longer treatments and higher deceased rates. All methods available to do so have deficiencies in its detection rates, response time, or have a higher cost and need a complex infrastructure to setup. In this work, by using responses retrieved from anamnesis, it is proposed the development of a neural network model to support drug-resistant TB diagnosis. Data are originated from reference centers addressing drug-resistant TB in Rio de Janeiro, comprising exclusively patients presumed to have drug-resistant TB. For the classification between drug-resistant and drug-sensitive TB, the sensitivity and specificity reached  $95.15 \pm 4.13$  and  $85.47 \pm 5.68$ , respectively.

ÉVORA, L.H.R.A, SEIXAS, J.M., KRITSKI, A.L., "Modelos de redes neurais para apoio a triagem diagnóstica de tuberculose resistente e multirresistente". *V Workshop da Rede TB*. Maceió/AL, 22 a 24 de agosto de 2016. Poster.

**Resumo:** Tuberculose (TB) é a principal causa de mortalidade no mundo entre as doenças infecciosas. O diagnóstico de TB resistente (DR) e multirresistente (MDR) demanda ainda maior atenção, implicando em tratamentos mais longos e em um

aumento do número de mortes decorrentes da doença. Todos os métodos de diagnósticos disponíveis para TB-DR/MDR possuem baixa sensibilidade, elevado tempo para acesso do resultado laboratorial pela equipe de saúde, e demandam uma infraestrutura mais complexa em sua implantação, com aumento dos custos. Este trabalho propõe, a partir de informações obtidas na anamnese, analisar o desempenho de um escore clínico para triagem, na abordagem diagnóstica de TB-DR/MDR, por meio de modelos computacionais baseados em Redes Neurais Artificiais (RNA). Inspiradas na biologia, as RNA possuem a capacidade de aprendizado a partir da amostra de dados, sendo utilizadas com sucesso no reconhecimento de padrões em diferentes áreas, inclusive com aplicações no apoio ao diagnóstico médico. Informações sobre sintomas e dados históricos dos pacientes foram coletadas em centros de referência em TB-DR no Rio de Janeiro, Brasil, incluindo exclusivamente pacientes suspeitos de TB-DR/MDR. O estudo considerou diferentes cenários práticos de operação ao desenvolver modelos utilizando critérios distintos de seleção de variáveis, para que ambientes em condições de poucos recursos possam ser contemplados. Para a classificação entre TB-DR/MDR e pacientes com TB droga-sensível, a sensibilidade e especificidade alcançadas foram de 95,1%(±5,0) e 85,0%(±4,9), respectivamente. Os resultados promissores alcançados pelos escores clínicos indicam que essa abordagem pode ser utilizada na avaliação inicial de pacientes suspeitos de TB-DR/MDR.

ÉVORA, L.H.R.A, SEIXAS, J.M., KRITSKI, A.L., "A Supporting System for the Screening of Drug and Multidrug Resistant Tuberculosis in Brazil Based on Computational Models in a State and National-Wide Approach Comparison". To be submitted in 2018.

### **Resumo:**

*Background:* A positive RR result by new molecular diagnostics for Tuberculosis (TB), such as the Xpert MTB/RIF, should be further analyzed in order to take into consideration the risk of Multidrug-Resistant TB (MDR-TB). Nevertheless, in a continental-sized country such as Brazil, a unifying national guideline for screening may obfuscate potentially discriminating local characteristics.

*Objective:* Develop and evaluate the performance of localized clinical scores in



order to support screening of presumed DR-TB patients in Brazil, aiding the identification of those should be further tested by new molecular diagnostics.

*Method:* From Feb 2011 and May 2013, 502 presumed DR-TB subjects attended in five reference centers in Brazil were evaluated. Artificial Neural Networks (ANN) and Classification And Regression Tree (CART) models are generated, considering as input the patient's symptoms and historical factors comprising a total of 38 variables, and their results are compared. Evaluated diagnoses comprise in Drug-Sensitive (DS), Multi-Drug Resistant (MDR), Drug-resistant (DR) -TB and non-active TB. Specific scores by each State (Rio de Janeiro - RJ, Ceará - CE, Rio Grande do Sul - RS) are produced and the results are compared to a national-wide approach. In order to consider environments with low resource conditions, distinct operation scenarios are addressed by using different relevance criteria for variable selection.

*Results:* ANN presented better results than CART, and models developed by each State achieved an average sensitivity higher than 85% when screening RJ patients considering DR-TB from non DR-TB, against 82.7% using the national approach. When screening DR-TB and DS-TB the approach by States presented a sensitivity superior to 83%, while for MDR-TB discrimination the average was 77%. In these scenarios the sensitivity achieved nationally was 73.7% and 73.4%, respectively. Using only the four most relevant variables in localized models, a sensitivity of over 78.6% is still obtained, maintaining specificity over 74.5%.

*Conclusion:* Localized models show better performance than a national-wide approach, indicating that the clinical scores can be used for screening presumed DR-TB patients. Nevertheless, environments with low resource conditions can be attended by less complex models developed through variable selected by relevance criteria.