



UNIVERSIDADE ESTADUAL DO CEARÁ
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CUIDADOS CLÍNICOS EM
ENFERMAGEM E SAÚDE
MESTRADO ACADÊMICO EM CUIDADOS CLÍNICOS EM ENFERMAGEM E
SAÚDE

GEORGE JÓ BEZERRA SOUSA

USO DO MÉTODO CRISP-DM NA PREDIÇÃO DO ENCERRAMENTO DO
TRATAMENTO DA TUBERCULOSE COMO FERRAMENTA DE APOIO AO
CUIDADO

FORTALEZA-CEARÁ

2018

GEORGE JÓ BEZERRA SOUSA

USO DO MÉTODO CRISP-DM NA PREDIÇÃO DO ENCERRAMENTO DO
TRATAMENTO DA TUBERCULOSE COMO FERRAMENTA DE APOIO AO
CUIDADO

Dissertação apresentada ao Curso de Mestrado Acadêmico em Cuidados Clínicos em Enfermagem e Saúde do Programa de Pós-Graduação em Cuidados Clínicos em Enfermagem e Saúde do Centro de Ciências da Saúde da Universidade Estadual do Ceará, como requisito parcial à obtenção do título de Mestre em Cuidados Clínicos.

Área de Concentração: Cuidados Clínicos em Enfermagem e Saúde.

Orientadora: Prof^a. Dr^a. Maria Lúcia Duarte Pereira.

FORTALEZA - CEARÁ
2018

Dados Internacionais de Catalogação na Publicação

Universidade Estadual do Ceará

Sistema de Bibliotecas

Sousa, George Jó Bezerra.

Uso do método CRISP-DM na predição do encerramento do tratamento da tuberculose como ferramenta de apoio ao cuidado [recurso eletrônico] / George Jó Bezerra Sousa. - 2018.

1 CD-ROM: il.; 4 ¼ pol.

CD-ROM contendo o arquivo no formato PDF do trabalho acadêmico com 140 folhas, acondicionado em caixa de DVD Slim (19 x 14 cm x 7 mm).

Dissertação (mestrado acadêmico) - Universidade Estadual do Ceará, Centro de Ciências da Saúde, Programa de Pós-Graduação em Cuidados Clínicos em Enfermagem e Saúde, Fortaleza, 2018.

Área de concentração: Enfermagem.

Orientação: Prof.^a Ph.D. Maria Lúcia Duarte Pereira.

1. Tuberculose. 2. Mineração de Dados. 3. Sistemas de Apoio à Decisão. 4. Epidemiologia. 5. Enfermagem.
I. Título.

GEORGE JÓ BEZERRA SOUSA

USO DO MÉTODO CRISP-DM NA PREDIÇÃO DO ENCERRAMENTO DO
TRATAMENTO DA TUBERCULOSE COMO FERRAMENTA DE APOIO AO
CUIDADO

Dissertação de Mestrado apresentada ao
Curso de Mestrado Acadêmico em
Cuidados Clínicos em Enfermagem e
Saúde do Programa de Pós-Graduação
em Cuidados Clínicos em Enfermagem e
Saúde do Centro de Ciências da Saúde
da Universidade Estadual do Ceará, como
requisito à obtenção do título de Mestre
em Cuidados Clínicos.
Área de Concentração: Cuidados Clínicos
em Enfermagem e Saúde.

Aprovado em: 20 de dezembro de 2018.

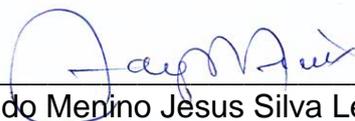
BANCA EXAMINADORA



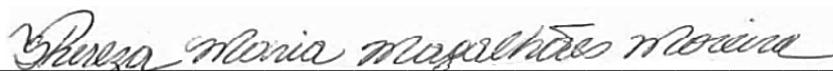
Prof.^a Dr.^a Maria Lúcia Duarte Pereira (Orientadora)
Universidade Estadual do Ceará – UECE



Prof. Dr. Jefferson Teixeira de Souza (1º membro efetivo)
Universidade Estadual do Ceará – UECE



Prof.^a. Dr.^a. Terezinha do Menino Jesus Silva Leitão (2º membro efetivo)
Universidade Federal do Ceará – UFC



Prof.^a. Dr.^a. Thereza Maria Magalhães Moreira (Membro suplente)
Universidade Estadual do Ceará – UECE

Dedico a todas as pessoas que acreditaram nesta pesquisa e àquelas cujo adoecer permitiu o seu desenvolvimento.

AGRADECIMENTOS

À Deus, pelo dom da vida, pela força dada e por nunca me fazer desistir.

Aos meus pais, que sempre me disseram que a educação é chave para uma vida melhor; e aos meus irmãos, por sempre estarem ao meu lado nessa caminhada.

A toda minha família, em especial aos meus padrinhos, que me acolheram como um de seus filhos e fizeram de sua casa mais um lar para mim.

A Universidade Estadual do Ceará (UECE), Programa de Pós-Graduação em Cuidados Clínicos em Enfermagem e Saúde (PPCCLIS): professores, coordenação e secretaria; e a Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) que me proporcionaram os meios para realizar essa pesquisa.

À minha orientadora, professora Lúcia Duarte (Lucinha) que aceitou participar desse caminho desafiador quando nem eu mesmo tinha domínio do assunto. Sua confiança em mim me fez seguir nesse processo sem olhar para trás. A senhora acreditou em mim e nessa pesquisa mais do que qualquer um nesse período.

À banca examinadora desse trabalho, professora Therezinha do Menino Jesus, especialista em tuberculose, professor Jerffeson Teixeira, especialista em computação e professora Thereza Moreira, especialista em estatística em saúde. A expertise de cada um, em conjunto, continuará contribuindo para que os resultados desse trabalho impactem positivamente na vida da população.

Ao grupo de pesquisa Clínica e Epidemiologia das Doenças Infecciosas e Parasitárias (CEDIP), especialmente à Lia e à Gizelly que acreditam no meu potencial e que são referências para mim. Espero sempre poder contar com os sequestros da Lia e aos conselhos da Gizelly.

A quatro amigos em especial: Bruna, Julio, Magno e Renata (ordem alfabética), que carregam em si um pedaço de mim, que me acolheram como irmão da vida e que só eu sei o quanto foram importantes durante esse processo.

Aos meus amigos/família do apartamento 102 (Ricardo, Ceará, Elias, Rafael, Thiago e minha sister Jéssica) por dividirem a vida comigo e aos professores Ricardo e Selma que ao ceder sua casa para nós, fizeram florescer os nossos sonhos.

Aos meus amigos de Itapajé que sabem que a distância física se fez necessária, mas amizade sempre continua.

Aos meus amigos do grupo do come-come e dos migles.

Aos amigos do programa de pós-graduação em saúde pública e do programa de pós-graduação de enfermagem da Universidade Federal do Ceará que tive o prazer de conhecer nesse período.

À décima terceira turma do mestrado (a minha T13), nunca poderia ter pensado que uma turma de mestrado poderia ser tão divertida e inteligente ao mesmo tempo! A eles acrescento alguns amigos da sexta turma do doutorado (T6), em especial o Eduardo que foi um dos melhores amigos que fiz no programa e você sabe que pode sempre contar comigo!

Aos professores do PPCCLIS que foram o alicerce para eu conseguir pensar e entender a enfermagem e saúde de uma maneira mais abrangente. Aos professores Marcial Fernandez, Thelmo Araújo do departamento de computação da UECE, que me ajudaram a entender melhor meu referencial metodológico. Aos professores Carlos Henrique Alencar e Larissa Fortunado do departamento de saúde pública da UFC, que são duas das minhas grandes referências em Epidemiologia. Ao professor Wellington Aguiar e a Liga de Big Data da Estácio que me guiaram nas primeiras leituras e entendimento desse novo ramo de estudo.

À Secretaria de Saúde do Estado do Ceará, especialmente à Dra. Kátia Alves por toda ajuda no processo burocrático de aquisição do banco de dados e pela disposição em ajudar os pesquisadores que a buscam.

Enfim, a todos que contribuíram para esse trabalho, o meu mais sincero obrigado! Esta pesquisa também é de vocês!

“Salve quebrada!
Vida longa aos vida louca!”
(Irmão do Jorel)

RESUMO

A tuberculose (TB) é um grave problema de saúde pública. Para garantir sucesso nos programas de manejo da doença, espera-se que, ao fim do tratamento, as pessoas apresentem cura e evitem desfechos negativos, como o abandono, óbito ou resistência medicamentosa. O uso de um método detalhado para analisar os dados que tais pessoas produzem pode ser uma ferramenta importante nesse contexto. O presente trabalho possui como objetivo aplicar o método CRISP-DM para identificar a epidemiologia da TB no Ceará e criar um modelo preditivo para os desfechos cura, abandono do tratamento, óbito por TB e resistência ao tratamento. Para tal foi desenvolvido um estudo transversal, realizado no Ceará, utilizando como referencial metodológico o CRISP-DM. Foi utilizado o banco de dados do Sistema de Informação de Agravos de Notificação para a Tuberculose (SINAN-TB) durante o período de 2001 a 2017, usando as variáveis sociodemográficas, epidemiológicas, clínico-laboratoriais e como desfecho a situação de encerramento do tratamento contidos nas fichas de notificação e acompanhamento do tratamento. O dataset foi analisado por três perspectivas: análise de tendência temporal, estudo epidemiológico analítico e estudo de predição com uso dos softwares Joinpoint Regression Program, Stata/SE e WEKA. As associações de variáveis foram consideradas significantes quando $p < 0,05$ e a força de associação foi verificada pela Razão de Prevalência. Pesquisa aprovada pelo Comitê de Ética em Pesquisa sob parecer número 2.687.046. O banco contou com 74.006 casos (71% curaram, 12,5% abandonaram o tratamento, 2% morreram e 0,5% adquiriu resistência ao tratamento). A análise de tendência temporal mostrou declínio da incidência, prevalência e taxa de cura da doença, assim como elevação sua resistência e mortalidade ($p < 0,05$). O modelo preditivo foi desenvolvido com uso do algoritmo de *machine learning* Redes Neurais Artificiais (RNA), composto por todas as variáveis significativas nos quatro modelos epidemiológicos como nós de entrada, sendo 43 nós na camada oculta e cinco nós de saída, identificando corretamente 74,21% de todos os casos e apresentando curva ROC $> 0,80$. Conclui-se que o método CRISP-DM pode ser utilizado na criação de modelos preditivos para o encerramento do tratamento e que eles podem ser uma ferramenta importante de apoio ao cuidado clínico de enfermagem e saúde às pessoas com tuberculose.

Palavras-chave: Tuberculose. Mineração de Dados. Sistemas de Apoio à Decisão. Epidemiologia. Enfermagem.

ABSTRACT

Tuberculosis (TB) is a serious public health problem. In order to guarantee success in disease management programs, people are expected to cure and avoid negative outcomes, such as default, death or drug resistance, at the end of treatment. The use of a detailed method for data analysis can be an important tool in the context. The present work aims to apply the CRISP-DM method to identify the TB epidemiology in Ceará and to create a predictive model for the outcomes of cure, treatment abandonment, TB death, and treatment resistance. A cross-sectional study was carried out in Ceará, using CRISP-DM as a methodological reference. The database of the Tuberculosis Notification Disease Information System (SINAN-TB) was used during 2001 to 2017, using sociodemographic, epidemiological, clinical-laboratory variables and as outcome the treatment conclusion situation contained on the notification and treatment records. The dataset was analyzed by three perspectives: time trend analysis, analytical epidemiological study and prediction study using the Joinpoint Regression Program, Stata / SE and WEKA software. Associations of variables were considered significant when $p < 0.05$ and the strength of association was verified by the Prevalence Ratio. Research approved by the Research Ethics Committee under opinion number 2.687.046. The dataset had 74,006 cases (71% cured, 12.5% abandoned treatment, 2% died and 0.5% acquired resistance to treatment). The temporal trend analysis showed a decline in the incidence, prevalence and cure rate of the disease, as well as an increase in resistance and mortality ($p < 0.05$). The predictive model was developed using the algorithm of machine learning Artificial Neural Networks (RNA), composed of all the significant variables in the four epidemiological models as input nodes, 43 nodes in the hidden layer and five nodes as output, correctly identifying 74.21% of all cases and presenting ROC curve > 0.80 . It is concluded that the CRISP-DM method can be used in the creation of predictive models for treatment conclusion and that they can be an important tool to support clinical nursing and health care for people with tuberculosis.

Keywords: Tuberculosis. Data Mining. Decision Supporting Systems. Epidemiology. Nursing.

LISTA DE FIGURAS

Figura 1 –	Representação de um neurônio humano	26
Figura 2 –	Representação de um neurônio humano em blocos ..	26
Figura 3 –	Modelo de um neurônio de RNA	27
Figura 4 –	Ciclo de vida de um projeto de mineração de dados de acordo com a metodologia CRISP-DM	36
Figura 5 -	Fases, tarefas e resultados propostos pela metodologia CRISP-DM	39
Figura 6 –	Matriz de confusão	47
Figura 7 –	Fluxograma de limpeza de dados	56
Figura 8 –	Padrão temporal da incidência da tuberculose	69
Figura 9 –	Padrão temporal da prevalência da tuberculose	70
Figura 10 –	Padrão temporal da taxa de cura da tuberculose	71
Figura 11 –	Padrão temporal da taxa de abandono da tuberculose	72
Figura 12 –	Padrão temporal da taxa de óbito (letalidade) da tuberculose	73
Figura 13 –	Padrão temporal do coeficiente de mortalidade da tuberculose	74
Figura 14 –	Padrão temporal do abandono do tratamento da tuberculose	74
Figura 15 –	Modelo de rede neural artificial proposto pelo pesquisador	100

LISTA DE QUADROS

Quadro 1 –	Decisão sobre uso de variáveis do banco de dados....	53
Quadro 2 –	Descrição das variáveis criadas para o estudo	65
Quadro 3 –	Valores reais e preditos dos desfechos do tratamento da tuberculose por um perceptron multicamadas.....	101
Quadro 4 –	Matriz de confusão das predições do desfecho cura do tratamento.....	101
Quadro 5 –	Matriz de confusão das predições do desfecho abandono do tratamento	102
Quadro 6 –	Matriz de confusão das predições do desfecho óbito por TB.....	102
Quadro 7 –	Matriz de confusão das predições do desfecho resistência ao tratamento da TB	103
Quadro 8 –	Valores reais e preditos dos desfechos do tratamento da tuberculose por validação cruzada (10-fold) de perceptron multicamadas	104

LISTA DE TABELAS

Tabela 1 –	Porcentagem de dados ausentes de cada variável do estudo.....	51
Tabela 2 –	Características sociodemográficas de pacientes com TB de acordo com o tipo de encerramento.....	57
Tabela 3 –	Características clínico-laboratoriais de pacientes com TB de acordo com o tipo de encerramento.....	59
Tabela 4 –	Características de tratamento e acompanhamento de pacientes com TB de acordo com o tipo de encerramento.....	62
Tabela 5 –	Indicadores Epidemiológicos dos desfechos da tuberculose no estado do Ceará, 2001-2017.....	68
Tabela 6 –	Padrão temporal dos indicadores epidemiológicos da tuberculose no estado do Ceará, 2001-2017.....	75
Tabela 7 –	Análise bivariada (p-valores) das variáveis preditoras com o desfecho cura do tratamento da tuberculose.....	76
Tabela 8 –	Análise bivariada (p-valores) das variáveis preditoras com o desfecho abandono do tratamento da tuberculose.....	79
Tabela 9 –	Análise bivariada (p-valores) das variáveis preditoras com o desfecho óbito por tuberculose.....	83
Tabela 10 –	Análise bivariada (p-valores) das variáveis preditoras com o desfecho resistência ao tratamento da tuberculose.....	86
Tabela 11 –	Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho cura da TB.....	91
Tabela 12 –	Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho abandono do tratamento da TB.....	94
Tabela 13 –	Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho óbito por TB.....	96
Tabela 14 –	Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho TB resistente a medicamentos.....	97
Tabela 15 –	Apresentação de critérios de avaliação dos modelos preditivos logísticos.....	103
Tabela 16 –	Apresentação de critérios de avaliação do modelo preditivo baseado em perceptron multicamadas.....	104

LISTA DE ABREVIATURA E SIGLAS

AAPC	Average Annual Percentage Change (Variação Percentual Anual Média)
AIDS	Síndrome da Imunodeficiência Adquirida
APC	Annual Percentual Change (Variação Percentual Anual)
BAAR	Bacilo Álcool Ácido Resistente
BCG	Bacilo Calmette-Guérin
BK	Bacilo de Koch
CAAE	Certificado de Apresentação para Apreciação Ética
CART	Classification and Regression Tree (Árvore de Classificação e Regressão)
CEP	Comitê de Ética em Pesquisa
CRISP-DM	Cross-Industry Standard Process for Data Mining (Processo Padrão Inter-Indústria para Mineração de Dados)
DOTS	Directly Observed Treatment Short-course (Tratamento Diretamente Observado de curta duração)
DPOC	Doença Pulmonar Obstrutiva Crônica
GUI	Graphical User Interface (Interface Gráfica do Usuário)
HIV	Vírus da Imunodeficiência Humana
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IC95%	Intervalo de Confiança de 95%
IRR	Incidence of Relative Risk (Incidência de Risco Relativo)
KNN	K-nearest neighbor
LM	Levenberg-Marquadt algoritmo
OMS	Organização Mundial da Saúde
OR	Odds Ratio (razão de chances)
PPD	Derivado Proteico Purificado (Teste Tuberculínico)
RNA	Rede Neural Artificial
ROC	Receiver Operator Characteristic (Caraterística de Operação do Receptor)
RP	Razão de Prevalência

SESA	Secretaria de Saúde do Estado do Ceará
SINAN	Sistema de Informação de Agravos de Notificação
SVM	Support Vector Machine (Máquina de Suporte ao Vetor)
TARV	Terapia Antirretroviral
TB	Tuberculose
TB-DR	Tuberculose Drogarresistente
TB-MR	Tuberculose Multirresistente
TB-XR	Tuberculose Extensivamente Resistente
TDOV	Tratamento Diretamente Observado por Vídeo
TIC	Tecnologia da Informação e Comunicação
UECE	Universidade Estadual do Ceará
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo
WEKA	Waikato Environment for Knowledge Analysis

SUMÁRIO

1	INTRODUÇÃO	17
2	HIPÓTESES	22
3	OBJETIVOS	23
3.1	GERAL.....	23
3.2	ESPECÍFICOS.....	23
4	REVISÃO DE LITERATURA	24
4.1	INTELIGÊNCIA ARTIFICIAL (IA) E <i>MACHINE LEARNING</i>	24
4.2	APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL (IA) NA EPIDEMIA DA TUBERCULOSE (TB).....	27
4.3	TECNOLOGIAS PARA O CUIDADO CLÍNICO DE ENFERMAGEM – SISTEMAS DE APOIO A DECISÃO	31
5	MÉTODO	35
5.1	REFERENCIAL METODOLÓGICO.....	35
5.2	TIPO DE ESTUDO.....	42
5.3	LOCAL DO ESTUDO.....	43
5.4	DADOS E FONTE DE DADOS	43
5.5	ANÁLISE DE DADOS.....	44
5.6	ASPECTOS ÉTICOS.....	48
6	RESULTADOS	49
7	DISCUSSÃO	108
7.1	EVOLUÇÃO TEMPORAL DOS INDICADORES EPIDEMIOLÓGICOS DA TUBERCULOSE	108
7.2	FATORES ASSOCIADOS A CURA DA TUBERCULOSE.....	111
7.3	FATORES ASSOCIADOS AO ABANDONO DO TRATAMENTO DA TUBERCULOSE.....	114
7.4	FATORES ASSOCIADOS AO ÓBITO POR TUBERCULOSE	118
7.5	FATORES ASSOCIADOS À RESISTÊNCIA AO TRATAMENTO DA TUBERCULOSE.....	120
7.6	APLICAÇÃO DO MÉTODO CRISP-DM NA PREDIÇÃO DOS DESFECHOS DA TB.....	122
8	CONCLUSÃO	124
	REFERÊNCIAS	127
	ANEXOS	136
	ANEXO A – FICHA DE NOTIFICAÇÃO DA TB.....	137
	ANEXO B – TELA DE ACOMPANHAMENTO TB.....	138
	ANEXO C – PARECER CONSUBSTANCIADO DO CEP	139

1 INTRODUÇÃO

A tuberculose (TB) é uma doença infectocontagiosa, transmitida pelo *Mycobacterium tuberculosis*, ou Bacilo de Koch (BK), nome dado graças a Robert Koch, cientista que detectou esses bacilos em 1882. Doença milenar, detectada em múmias egípcias, comprovando sua existência há mais de 3400 anos antes de Cristo. O *M. tuberculosis* é um bacilo ligeiramente curvo, imóvel, não esporulado, não encapsulado, que possui a propriedade de ser álcool-ácido resistente (BAAR). Mesmo sendo dificilmente corada pelo método de Gram, é considerada uma bactéria Gram-positiva devido os aspectos de sua parede celular. Sendo também chamada de bactéria não pigmentada de crescimento lento (CAMPOS, 2006).

A transmissão da doença ocorre por aerossóis, produzidos a partir de secreções respiratórias de pessoas com manifestação pulmonar ativa, que pode ser inalado por um contato susceptível. Geralmente, causa a TB pulmonar, mas também pode afetar outros órgãos, passando a se denominar TB extrapulmonar. Existem grupos de pessoas com mais chances de desenvolver a doença, como as vivem com o Vírus da Imunodeficiência Humana (HIV) ou Síndrome da Imunodeficiência Adquirida (Aids), pessoas privadas de liberdade e moradores de rua (CHURCHYARD, *et al.*, 2017). Contudo, deve-se atentar que tais grupos não são os únicos que podem se infectar. Por isso, é possível perceber que a abrangência da TB se dá em diversos eixos e grupos pelos quais se precisa trabalhar.

Segundo recente publicação da Organização Mundial da Saúde (OMS), a TB é a nona causa de morte no mundo e a primeira por doença infecciosa. Estimam-se que 10 milhões de pessoas tiveram TB no ano de 2017, sendo 90% adultos (≥ 15 anos) e 9% vivendo com HIV/Aids (WHO, 2018).

Além disso, outro grave problema de saúde pública para a TB é o desenvolvimento de formas resistentes a drogas padrão para o seu tratamento, que podem evoluir para uma forma conhecida como TB Multirresistente (TB-MR), caracterizando-se como resistência a, pelo menos, isoniazida e rifampicina. Além disso, a multirresistência pode progredir para uma resistência extensiva, conhecida como TB Extensivamente Resistente (TB-XR), definida como a TBMR mais a resistência a, pelo menos, uma fluoroquinolona e uma droga injetável de segunda linha (WHO, 2014).

No Brasil, em 2017 foram notificados 69.569 casos novos de TB. Além disso, o país apresenta incidência da doença de 33,5 casos/100.000 habitantes, mas esses índices não podem ser generalizados, pois os estados se mostram heterogêneos quanto à forma e à distribuição da doença. (BRASIL, 2017a; BRASIL, 2018).

No Ceará, em 2016 foram notificados 3.653 casos novos de TB. Como apresentado anteriormente, os dados do estado se mostram semelhantes aos brasileiros no que tange à queda da incidência da doença. Tal incidência ainda é considerada elevada, pois observou-se que em 2013 a mesma era de 49,4/100 mil habitantes e em 2017 de 38,3/100 mil habitantes (CEARÁ, 2018).

Nesse sentido, entender o desfecho do tratamento, ou situação de encerramento do tratamento da TB torna-se importante para a discussão da tendência temporal da TB. A OMS recomenda mínimo de 85% de cura e máximo de 5% de abandono do tratamento. O último relatório do desfecho mundial do tratamento mostra taxa global de cura de 82% (WHO, 2018).

Além disso, mundialmente ainda foi estimada uma taxa de 1,3 milhões de mortes para as pessoas somente com TB e 464 mil para as vivendo com HIV/Aids. Em relação à resistência, estima-se que 3,5% dos novos casos e 18% dos casos de retratamento tenham desenvolvido alguma das formas da doença (WHO, 2018).

No Brasil, dados de 2015 mostram taxa de 71,9% de cura dos casos novos e taxa de abandono de 10,4%. O último boletim epidemiológico do Brasil não mostra a mortalidade como porcentagem, mas como um índice comparado para 100 mil habitantes., observando-se um atual coeficiente de mortalidade de 2,2/100 mil habitantes. No Ceará constatou-se taxa de 59,7% de cura da doença, taxa de abandono de 10,8% e taxa de mortalidade de 4,5% (BRASIL, 2017a; CEARÁ, 2018).

Em relação à resistência, sabe-se que, mundialmente 480 mil pessoas desenvolveram TB multirresistente, com 190 mil mortes associadas e que nas Américas esse número chega a 7.700 casos. No Brasil, expressa-se que 1.077 pessoas desenvolveram formas resistentes da doença (BRASIL, 2017b).

Assim, é possível observar que as estratégias para o manejo e controle da TB no Ceará estão se tornando ineficientes, pois todos os indicadores se

mostram longe dos propostos pela OMS e pactuados pelo Ministério da Saúde brasileiro, exceto para mortalidade.

Considerando a situação descrita, vem-se buscando constantemente novas formas de lidar com esse problema de saúde pública. Devido a isso, a OMS, em seu programa para o combate da doença até 2035, traz no seu pilar três, que a pesquisa e inovação devem ser estimuladas. Nesse sentido, o uso de novas tecnologias, como as de informação e comunicação (TIC) são ferramentas que podem se mostrar importantes nesse campo (FALZON, *et al.*, 2016). Dentre elas, uma área das ciências da computação que vem ganhando destaque nos últimos anos nas ciências da saúde é a de Inteligência Artificial.

Inteligência artificial (IA) é um conceito amplo, que se refere aos sistemas poderiam realizarem tarefas humanas e, dessa forma, acelerarem as formas de trabalho e resolução de problemas. Ela pode ser dividida em IA geral e IA estreita. A IA geral se refere a uma máquina, que pode realizar muitas funções da mente humana como calcular, planejar, entender linguagem, reconhecer objetos, entre outros. Já a IA estreita se refere a uma máquina que faria bem uma das tarefas descritas anteriormente, mas não seria capaz de realizar outras. Esta é a forma como o aprendizado em máquina, ou *machine learning*, é pensado. Nele treina-se um algoritmo (sem ser amplamente programado) capaz de aprender por meios de padrões e realizar tarefas futuras, baseado no que foi aprendido (Associação Brasileira de Internet das Coisas – ABIC, 2017).

Vale ressaltar que, para se aplicar um algoritmo baseado em IA e *machine learning*, deve-se realizar um processo denominado de mineração de dados. Esse processo incorpora um trajeto para aquisição de conhecimento pela aquisição dos padrões, ao se analisar uma grande quantidade de dados. Para isso, são utilizados algoritmos que aprendem como esses dados detectam padrões para encontrar um conjunto de informações válidas a um objetivo específico. As últimas duas décadas demonstraram crescimento constante de pesquisas que envolvem esse assunto e, em saúde, mostra-se como um ramo promissor desde a última década (ALONSO *et al.* 2018). Para se extrair conhecimento por meio de mineração de dados, diversos são os referenciais metodológicos usados, dentre eles há o Cross Industry Standard Process for Data Mining (CRISP-DM) – traduzido livremente como Processo Padrão Inter-indústria para Mineração de Dados – ou o Knowledge-

discovery in database – traduzido como descoberta de conhecimento em base de dados.

Além disso, um editorial publicado em uma importante revista de pneumologia da Europa mostra diversos espaços no qual pesquisadores podem fazer uso de IA para abordar a temática TB, citando quatro pontos fundamentais: a) cuidado ao paciente, b) vigilância, c) gerenciamento de programas e d) *e-learning*. O cuidado ao paciente se refere a modelos que ajudem o paciente no seu tratamento como tratamento, diretamente observado por vídeo (TDOV) ou monitores eletrônicos que melhorem seu prognóstico, mas essa área ainda é onerosa (DOSHI, *et al.*, 2017).

A segunda aplicação é em vigilância. Nesse ponto, já existem sistemas desenvolvidos na China capazes de detectar pacientes em risco para a doença, avaliar os medicamentos a serem usados e informar diretamente o sistema nacional de doenças infecciosas sobre o status do paciente (DOSHI, *et al.*, 2017). A criação desses tipos de sistema de vigilância implicaria na detecção precoce de casos, além da identificação de possíveis desfechos negativos da doença.

O terceiro ponto de interesse se refere ao gerenciamento de programas, principalmente aqueles em que os pacientes moram em áreas remotas e de difícil acesso, fazendo o uso de diagnóstico por técnicas moleculares rápidas e de baixo custo, com respostas ligadas a sistemas de saúde. Por fim, o *e-learning* é um método que utiliza ferramentas de mídias digitais e eletrônicas para educação e monitoramento. O uso de IA nesse caso poderia ser com o uso de um emulador de um tutor humano para ajudar a diminuir os espaços em branco no conhecimento das pessoas, utilizando-se de aplicativos *mobile*, por exemplo (DOSHI, *et al.*, 2017).

Portanto, percebe-se existir amplo campo de atuação para a IA no cuidado à TB. Em relação à vigilância, campo mais acessível dos quatro anteriormente expostos, os atuais estudos com modelos de IA visam detectar os casos de TB na população, partindo de suas características clínicas, sociodemográficas e geográficas. Até o momento, dos estudos brasileiros encontrados, vê-se tendência pelo enfoque na detecção da doença em si, não sendo encontrados estudos brasileiros que considerem o desfecho do tratamento (ver revisão de literatura).

Este estudo, por sua vez, justifica-se por tomar como desfecho a situação de encerramento do tratamento para a TB, utilizando como preditores os dados clínicos, sociodemográficos e a forma da doença que a pessoa apresenta.

O presente estudo se mostra relevante, pois considera modelos de IA na predição dos diversos desfechos do tratamento da doença. Esta investigação também pode demonstrar que os modelos de IA, ao predizerem corretamente os desfechos do tratamento, podem ajudar os profissionais de saúde a acompanhar as pessoas em risco com mais afinco, contribuindo para a mudança da atual situação epidemiológica em relação ao desfecho da doença de pessoas previamente expostas.

No cuidado clínico de enfermagem, essa investigação contribuirá como um subsídio para à prática profissional, pois pode servir como um sistema de apoio à decisão. O modelo permitirá identificar casos com mais risco de desenvolver desfecho negativo da doença, permitindo a aplicabilidade de estratégias que modifiquem tal situação.

Tomando como base o atual panorama da TB nos países subdesenvolvidos e o papel que ela possui no processo saúde-doença das populações e o uso de modelos de IA como ferramenta de apoio a decisões em saúde, questiona-se: “Modelos de IA podem predizer o desfecho do tratamento da TB”?

2 HIPÓTESES

Hipótese Nula (H_0): O método CRISP-DM não pode ser aplicado para utilizar algoritmos de IA para prever os desfechos do tratamento da TB.

Hipótese Alternativa (H_1): O método CRISP-DM pode ser aplicado para utilizar algoritmos de IA para prever os desfechos do tratamento da TB.

3 OBJETIVOS

3.1 GERAL

Construir um modelo preditivo para os desfechos de TB (cura, abandono, óbito e resistência ao tratamento) a partir do método CRISP-DM.

3.2 ESPECÍFICOS

Caracterizar as causas de cura, abandono do tratamento, óbito por TB e resistência ao tratamento;

Analisar a tendência temporal da cura, abandono do tratamento, óbito por TB e resistência ao tratamento;

Estimar a prevalência das características associadas a cada desfecho da TB;

Desenvolver modelos preditivos para os desfechos da TB;

Criar modelos preditivos em uma rede neural artificial para os desfechos da TB;

Verificar acurácia desses modelos (sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo e área sobre a curva ROC).

4 REVISÃO DE LITERATURA

No campo da TB já existem diversos estudos que fazem uso de IA. Realizando uma busca no acervo de arquivos para ciências da saúde dos Estados Unidos (PubMed) é possível encontrar 144 artigos. Dos que realmente utilizaram-se de IA no estudo de TB, percebe-se que eles se dividem em duas grandes áreas: estudos para detecção da doença e estudos a níveis moleculares, geralmente relacionados à farmacologia.

Poucos desses estudos avaliam o desfecho. O desfecho em destaque é o da TB drogarresistente (TB-DR) expostos a um fármaco específico – amicacina (MODONGO, *et al.*, 2015; MODONGO, *et al.*, 2016). Entretanto, em uma busca mais ampla é possível encontrar alguns artigos publicados internacionalmente que já consideram o desfecho do tratamento da doença.

Nessa revisão, apresentar-se-á a definição de IA e aprendizado em máquina, mostrar-se-á como modelos de IA podem ser utilizados na investigação da TB e como a área da saúde, em especial a enfermagem, pode se beneficiar dessas tecnologias como ferramentas de apoio à decisão para o cuidado clínico de enfermagem.

4.1 INTELIGÊNCIA ARTIFICIAL (IA) E *MACHINE LEARNING*

Por toda a história, o homem se distinguiu dos demais animais pela sua capacidade de pensar, entender, prever e manipular (por conta de sua inteligência), intitulado-se nos dias de hoje como *Homo sapiens*. Entretanto, um dos mais novos campos da ciência e engenharia conhecido inteligência artificial (IA) vai além dessa capacidade, pois a IA busca não apenas entender, mas também criar entidades inteligentes (RUSSELL, NORVIG, 2010).

Atualmente, a IA está no cotidiano das pessoas, abrindo diariamente diversas possibilidades para seu uso. Ela está presente nos mais diversos setores que se pode pensar como “jogos de xadrez, provar teoremas matemáticos, escrever poesias, dirigir carros em ruas lotadas e diagnosticar doenças” (RUSSELL, NORVIG, 2010, p.1).

Além disso, vive-se em uma era onde dados são uma constante na vida. Tais dados encontram-se de forma estruturada e não estruturada, isto é, estão ou não organizados em bancos de dados. Daí surge um subcampo da IA chamado de aprendizado em máquina (do inglês *machine learning*).

Essa área usa algoritmos que aprendem sozinhos por meio desses dados e formulam regras que gradualmente aumentam sua performance como modelos preditivos e de decisão, facilitando a análise de conjuntos de dados grandes. Assim como IA, o *machine learning* já está aplicado no dia a dia dos seres humanos; por exemplo, os filtros de *spam* dos e-mails, *softwares* de reconhecimento de voz e ferramentas confiáveis de busca na web já integram a vida rotineira (RASCHKA, 2015).

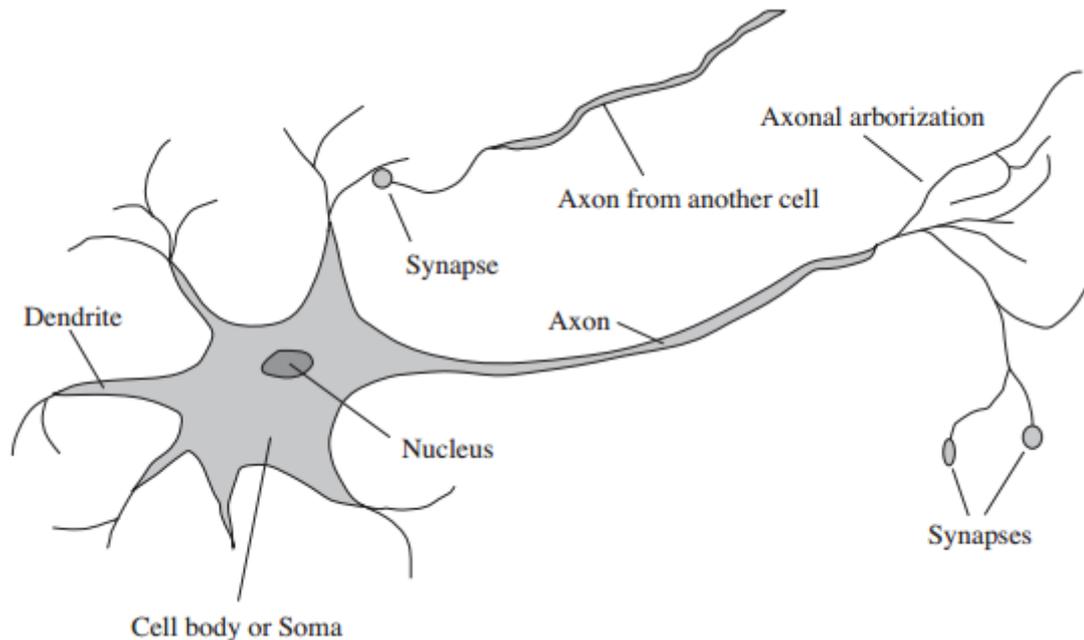
Esse aprendizado pode ser realizado de três diferentes formas: supervisionado, por reforço e não supervisionado. O termo supervisionado significa que a predição dos dados é dada por sinais (ou rótulos) já conhecidos; para a classificação, a predição de novas classes será dada por meio de um conjunto de dados do passado que servem como base para futuras predições. Eles são chamados dados de treino. A ideia de aprendizado por reforço é que o sistema (ou agente) melhore seu desempenho gradativamente por meio de interações com o ambiente. Esta melhora se dá por recompensas dadas quando a máquina corretamente prediz o desfecho. Por fim, o aprendizado não supervisionado pode ser realizado quando não se conhece os dados nem as interações que eles possuem (RASCHKA, 2015).

Tais tarefas realizadas pelas máquinas são estabelecidas por três contribuições da matemática: a lógica, a computação e a probabilidade. Tomando a computação como exemplo, entende-se que máquinas aprendem e desempenham suas funções por meio de diferentes algoritmos. Assim, um algoritmo é uma função pela qual a máquina lê e interpreta os dados. Essa ideia vem de al-Khowarazmi, um matemático persa do século IX, cujos escritos foram base para os números arábicos e álgebra na Europa (RUSSELL, NORVIG, 2010).

Um dos algoritmos mais utilizados pelos cientistas é o chamado Redes neurais Artificiais (RNA). As redes neurais recebem esse nome por serem pensadas

como neurônios humanos que, a partir de um estímulo, captam, processam e enviam uma mensagem resposta (Figura 01).

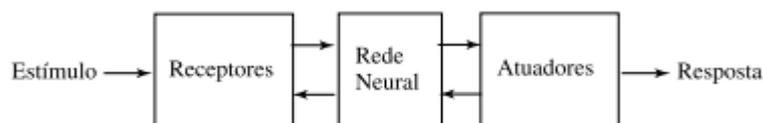
Figura 1 – Representação de um neurônio humano



Fonte: Russell, Norvig (2010, p. 11).

Os algoritmos de aprendizado dessas redes foram desenhados para representar a forma que o cérebro processa determinada tarefa (Figura 02). A função das RNA geralmente é implementada por um software, que aplica o algoritmo capaz de estabelecer a rede e conexões entre os nós de entrada dos neurônios.

Figura 2 – Representação de um neurônio humano em blocos

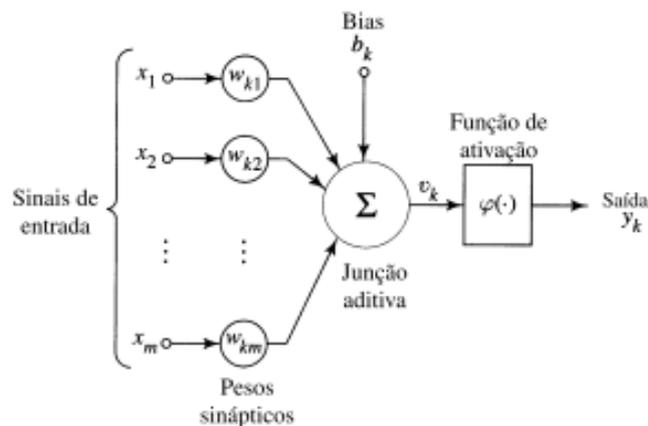


Fonte: Haykin (2009, p. 8).

O modelo gráfico a seguir (Figura 03) descreve um algoritmo de RNA. Segundo Haykin (2009), uma RNA é composta por três elementos básicos. Um conjunto de sinapses ou elos de conexão, cada uma possui um peso ou elo próprio.

Um somador que une os sinais de entrada pelas sinapses dos neurônios. Por fim, uma função de ativação que restringe a amplitude da saída neural, por isso também chamada como função restritiva. O intervalo de um neurônio, geralmente pode ser fechado $[0, 1]$ ou alternado $[-1, 1]$.

Figura 3 – Modelo de um neurônio de RNA



Fonte: Haykin (2009, p. 11).

Em suma, pode-se perceber que a IA pode estar presente no cotidiano das pessoas, as aplicações de algoritmos de aprendizado em máquina contribuem para a predição e detecção de diversos desfechos. A saúde, dessa forma, pode se beneficiar de tais sistemas como uma forma de melhor utilizar os dados gerados durante os diversos atendimentos, criando, assim, sistemas que possam ajudar os profissionais na tomada de decisões.

4.2 APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL (IA) NA EPIDEMIA DA TUBERCULOSE (TB)

Diversas são as aplicações de modelos baseados em IA no controle e combate à TB. Serão descritos, a seguir, como esses modelos vêm sendo utilizados e quais algoritmos são utilizados pelos pesquisadores.

Er, Yumusak e Termutas (2010) fizeram uso de redes neurais artificiais (RNA) numa amostra de 357 registros com 38 preditores para classificar doenças pulmonares, dentre elas a TB. Este estudo mostrou que, para essa doença, a melhor rede neural foi a multicamada com duas camadas ocultas usando algoritmo de

Levenberg-Marquardt (LM), apresentando uma acurácia de 90%. As demais doenças pulmonares (DPOC, pneumonia, asma, câncer de pulmão e sem doenças) também tiveram altos valores de acurácia, mas utilizando outras formas de redes neurais, como a regressão geral de redes neurais ou redes neurais com uma camada oculta.

Nesse sentido, outros modelos de IA baseados em aprendizado em máquina foram utilizados por Asha, Natarajan e Murthy (2011) na detecção da doença; dentre eles, support vector machine (SVM), k-nearest neighbors (KNN), classificador por naive Bayes, árvore de classificação e regressão (CART), bagging, AdaBoost e floretas aleatórias. Tal estudo demonstrou que a melhor acurácia pode ser obtida por meio de SVM, CART e florestas aleatórias.

O estudo de Saybani *et al.* (2015) melhorou um modelo de detecção da TB modificando o algoritmo de KNN para o de SVM. Com isso, obteve valores de sensibilidade e especificidade de 100% tão como área sob a curva ROC de valor um, mostrando que a troca de algoritmo trouxe mais poder de predição para o modelo. Os autores justificaram o uso dessas novas técnicas de detecção da doença em vista ao atraso do recebimento dos resultados de testes de baciloscopia de escarro e cultura, na qual pode apresentar valores pouco específicos em crianças. Dessa forma, tal modelo poderia apoiar a decisão do diagnóstico e início do tratamento.

Além destes já citados, estudos com base em reconhecimento de imagens têm se mostrado promissores no diagnóstico da TB. Um estudo mostrou o uso de IA na detecção de padrões de imagens, utilizando os resultados das radiografias pelo método de redes neurais convolucionais por *deep learning*. Nele, pode-se observar que os modelos que foram pré-treinados obtiveram melhores resultados que os que não tiveram nenhum tipo de treinamento, obtendo índices de sensibilidade, especificidade e acurácia superiores a 90%, mostrando que essa ferramenta pode identificar áreas pulmonares com lesões provenientes da infecção de *M. tuberculosis* (LAKHANI, SUNDARAM, 2017).

Modelos preditivos também foram utilizados para classificar o regime de tratamento que pode ser selecionado. Estudo utilizando árvores de classificação mostrou que o algoritmo consegue mostrar qual o tipo de paciente elegível para

determinado regime medicamentoso. Além disso, o software criado com esse algoritmo foi colocado para avaliação de médicos e enfermeiros, obtendo um score geral de 4,64/5. Dessa forma, o estudo mostrou-se capaz de fazer o que lhe foi proposto, além de se mostrar útil aos profissionais que fizeram e o manusearam, mostrando possuir uma aplicação prática (AVENIR, VILLOTA, ROSAS, 2015).

Diante do exposto, observa-se que a maioria dos estudos encontrados têm sido desenvolvidos no exterior. Entretanto, foi possível encontrar estudos realizados no Brasil que abordam o tema. Dos estudos brasileiros apresentados a seguir, observar-se-á que os autores preocuparam-se em desenvolver modelos de predição da TB.

Estudo realizado por Souza Filho, *et al.* (2016), mostrou uso de redes por diversos modelos que apresentaram altos índices de sensibilidade e especificidade para a detecção da tuberculose em 136 pacientes registrados no banco; entretanto, o uso de redes neurais artificiais foi o que mostrou melhor desempenho (sensibilidade de 100%, especificidade de 80%, acurácia de 88% e área sob a curva de 0.91). Mesmo assim os autores sugerem que pacientes HIV-positivos sejam melhor avaliados, pois para esse grupo, o modelo mostrou sensibilidade de 54% e especificidade de 76%.

Outro estudo brasileiro comparou CART e RNA na predição da TB. Identificou-se que o modelo de árvore de classificação, composto por 26 preditores, apresentou melhor especificidade e o de redes neurais, composto por 23 nós de entrada, apresentou melhor sensibilidade (SANTOS, *et al.*, 2003).

Foi possível encontrar um estudo que utilizou redes neurais para apoiar o diagnóstico de uma forma extrapulmonar da doença, a TB pleural. Os autores do estudo apontam que os modelos criados podem ajudar no diagnóstico da TB pleural, uma vez que testes histopatológicos mostraram as mesmas taxas de sensibilidade, além do mais, a inserção de marcadores bioquímicos aumenta a sensibilidade do teste para aproximadamente 99%. Entretanto, uma limitação do estudo é a amostra utilizada. Ele também sugere que trabalhos que envolvem esse tipo de métodos usem números maiores de registros nos bancos de dados (SEIXAS, *et al.*, 2013).

O último estudo brasileiro encontrado desenvolveu dois modelos baseados em redes neurais artificiais, o primeiro como um *perceptron* multicamadas

e o segundo como mapa auto organizado. O estudo alcançou sensibilidade de 96%, especificidade de 89%, conseguindo organizar o mapa em três grupos: baixo risco, risco moderado e alto risco. Em ambos, o resultado do exame radiológico foi o preditor mais importante. Entretanto, o HIV/Aids foi considerado um fator de confusão e não foi adicionado(a) no modelo final (AGUIAR, *et al.*, 2016).

Os estudos citados, como observado, tiveram como principal objetivo a detecção dos casos de TB (pulmonar ou extrapulmonar) baseado em alguns algoritmos de IA. Contudo, como esse estudo visa a predição dos desfechos, faz-se necessário expor os estudos que também se debruçaram sobre a mesma temática.

Primeiramente, como mostrado na introdução da revisão de literatura, estudos mostraram como a exposição a um determinado fármaco, como a amicacina, pode influenciar nos desfechos do tratamento da doença. Dentre eles identificou-se a TB resistente a medicamentos e a ototoxicidade como consequências graves do uso de determinadas concentrações de tal medicamento. Essas concentrações puderam ser identificadas pelos algoritmos utilizados (MODONGO, 2015; MODONGO, 2016).

Estudo de Kalhori e Zeng (2013) utilizou diversos algoritmos (CART, redes bayesianas, regressão logística, RNA, radial basis function e SVM) para prever os desfechos do tratamento da TB. Os desfechos avaliados foram cura, tratamento completo, abandono, falha e óbito, foram selecionados 38 preditores do registro para o programa “*Stop TB*” de base iraniana. Nesse estudo, as árvores de classificação se mostraram os melhores algoritmos com índices de acurácia por volta de 74,21%.

Modelos de regressão logística são atualmente utilizados em análise multivariada de dados em ciências da saúde. Estudo que objetivava buscar as características para a falha do tratamento por meio desse algoritmo conseguiu prever corretamente 81% dos dados apresentados no banco de teste, apresentando uma área sob a curva ROC de 0,70 (KALHORI, NASESHI, ZENG, 2010).

Um estudo de coorte retrospectiva também investigou os desfechos da TB extrapulmonar. Nesse estudo de coorte foram utilizados modelos de árvores de classificação, que conseguiram identificar, dentre outros, que quanto maior o tempo

de exposição ao tratamento da doença, maiores as probabilidades de se evoluir para um desfecho negativo. Quando feita a validação com os dados para treino, esse estudo ainda mostrou acurácia de 92% (PUSCH, *et al.*, 2014).

Por fim, o uso de árvores de classificação também foi utilizado para verificar o desfecho por meio da concentração das drogas do regime terapêutico no sangue. Esse estudo evidenciou que as três concentrações mais importantes são, em ordem de importância) as de pirazinamida (≤ 363 mg.h/L), rifampicina (≤ 13 mg.h/L) e isoniazida (≤ 52 mg.h/L). Uma limitação desse estudo é o pequeno número de pacientes HIV-positivos, o que torna difícil a generalização para esse grupo (PASIPANODYA, *et al.*, 2013).

Dessa forma, percebe-se que diversas podem ser as possibilidades para a aplicação de sistemas baseados em IA na linha de transmissão, identificação, cuidado e acompanhamento do tratamento da TB. O avanço científico e tecnológico são vitais nesse processo e a incorporação de áreas que à primeira vista seriam distantes da saúde são fundamentais para a inovação no processo de cuidar.

4.3 TECNOLOGIAS PARA O CUIDADO CLÍNICO DE ENFERMAGEM – SISTEMAS DE APOIO A DECISÃO

A enfermagem está diretamente ligada ao cuidado do paciente com TB. Segundo o Ministério da Saúde, dentre as competências que o enfermeiro possui, pode-se citar: a) realizar cuidado integral às pessoas e as famílias na Unidade Básica de Saúde - UBS; b) dar orientações a respeito da coleta do escarro; c) administrar a vacina BCG; d) realizar prova tuberculínica (se devidamente capacitado); e) realizar consultas de enfermagem de acordo com gestão local; f) solicitar exames, iniciar tratamento (se não houver médico no local) e prescrever medicações (esquema base da TB); g) planejar, coordenar e gerenciar atividades realizadas por técnicos, auxiliares e agentes comunitários de saúde (ACS); h) notificar casos de TB (BRASIL, 2011).

Como pode ser observado, o enfermeiro possui diversos papéis dentro da linha de cuidado ao paciente com TB. Devido à diversidade e abrangência das atividades e intervenções no ciclo vital dos indivíduos cuidados, podendo ser necessário suporte para a tomada de decisões clínicas. Nesse momento, a inserção

de modelos de apoio à decisão em saúde é uma alternativa que está cada vez mais presente nos dias de hoje.

Nesse sentido, um sistema de informação capaz de contribuir para a tomada de decisões é um sistema que contém dados que ajudam os enfermeiros a seguir o caminho correto durante o cuidado ao paciente. Por isso, é imprescindível que o sistema também seja criado por enfermeiros. Evidentemente esses modelos não substituem o conhecimento que o profissional possui, mas podem ser um aliado que contribua na segurança dos pacientes, tomada de decisões de enfermagem e melhoria da qualidade dos cuidados (TEIXEIRA, *et al.*, 2012). Alguns exemplos desses sistemas são demonstrados a seguir.

Sellmer *et al.* (2013) desenvolveram um sistema que apoia as decisões no cuidado à terapia tópica de úlceras venosas, de acordo com as características dessas úlceras. O sistema foi testado com o auxílio de uma enfermeira estomaterapeuta que elaborou casos clínicos para observar se o sistema se comportava da maneira como esperado. Vale ressaltar que o conhecimento sobre feridas, nesse caso, ainda permanece indispensável ao enfermeiro, pois as características cadastradas no sistema dependem do quanto o enfermeiro conhece tal morbidade.

Outro exemplo de sistema criado foi o de apoio para enfermagem obstétrica baseado em lógica *fuzzy* criado por Marques *et al.* (2005). Esse sistema baseava-se na tomada de decisão para a realização do procedimento de amniotomia, dado um conjunto de cinco variáveis obstétricas. A partir daí, quatro sugestões poderiam ser feitas pelo sistema, desde a contra-indicação do procedimento até a recomendação de fazer o mesmo.

Outro estudo da área da enfermagem desenvolveu um sistema para a identificação de diagnósticos relacionados à eliminação urinária e indicar os procedimentos mais importantes para o paciente em questão. Esse sistema foi baseado em entrevistas realizadas e depois transferido para um software que interpretava as decisões por meio de árvores binárias, ou seja, dada a característica definidora, o profissional deveria marcar sua presença ou ausência. Após sua criação, o sistema foi colocado a teste com o padrão-ouro, obtendo valor de concordância de 98,9% (LOPES, HIGA, 2005).

Outros sistemas foram abordados por Bakken *et al.* (2007), que subsidiam o cuidado de enfermagem. O primeiro sistema é um para busca por evidências via digital; o segundo é um sistema de avaliação do risco de queda e o terceiro é um assistente digital para identificação e gerenciamento de obesidade e depressão.

Segundo os autores citados, sistemas de apoio à decisão, construídos sob uma base de informática, garantem maior segurança ao paciente, além de uma enfermagem baseada em evidências, eles garantem que o cuidado prestado aos pacientes e as ações tomadas sejam mais embasadas. Devido a isso, esse tipo de infraestrutura sustenta o cuidado em quatro pontos essenciais: a) aquisição de dados, b) padrões de cuidados de saúde, c) repositório de dados e d) fontes digitais de evidências. Por fim, os autores concluem que, para que os sistemas de informação clínicos facilitem o cuidado de enfermagem, os enfermeiros devem possuir conhecimento em duas áreas, sendo elas a de prática baseada em evidências e a de informática.

Um dos estudos realizados por enfermeiros que usou algoritmos de IA é o de Silva *et al.* (2014). Nele foi usado aprendizado não supervisionado para detectar pacientes com risco para desenvolver pé diabético. O estudo se assemelhava com o proposto por esse projeto, pois foi realizado por enfermeiros e utilizou o algoritmo de *k means* para detectar padrões com base em 153 registros de pacientes e 30 variáveis.

Tal estudo entrou nessa sessão, pois como dito anteriormente, foi desenvolvido por enfermeiros para detectar um problema altamente presente no cotidiano da profissão, utilizando técnicas de IA para criar um modelo de apoio à decisão.

Dessa forma, uma revisão sistemática, com metanálise, que avaliou diversos sistemas de apoio à decisão para a enfermagem, mostrou que, para o sistema trombolítico avaliado, não houve diferença entre o uso de enfermeiros e médicos. O sistema de triagem reduziu significativamente a carga de trabalho de enfermeiros. Dentro da mesma revisão, um estudo mostrou que esses sistemas são benéficos para os pacientes, e somente um dos estudos mostrou resultados negativos para o paciente (RANDELL, *et al.*, 2007).

Diante disso, é imprescindível que o enfermeiro saiba lidar com os novos recursos que se fazem presentes em seu meio, como tecnologias de apoio à decisão em saúde. O conhecimento mínimo de informática é essencial para que isso aconteça. Dessa forma, manter-se atualizado com as novas tecnologias e como elas podem influenciar o cuidado de enfermagem é fundamental.

Essas novas tecnologias, como seu próprio nome diz, não substituirão o profissional, muito menos o seu cuidado. Elas serão, entretanto, mais uma ferramenta de suporte do profissional de saúde, oferecendo novas oportunidades no manejo da doença. Desse modo, os sistemas inteligentes ajudariam o enfermeiro a detectar um defeito a partir de um conjunto de características. Vale ressaltar que todo o cuidado e o relacionamento com o paciente é feito por meio do profissional.

5 MÉTODO

5.1 REFERENCIAL METODOLÓGICO

O referencial a seguir foi criado utilizando o material produzido por Chapman et al. (2000) “*CRISP-DM 1.0: step-by-step data mining guide*”. Apresenta-se abaixo o motivo da escolha de tal método e uma síntese de como ele foi aplicado nessa investigação.

O método *Cross Industry Standard Process for Data Mining* – CRIPS-DM (Processo Padrão Entre Indústrias para Mineração de Dados) foi criado em 1996 por três “veteranos” do ramo, na época ainda imaturo, de mineração de dados: Daimler-Chrysler (na época Daimler-Benz), SPSS (então ISL) e NCR. Os três trabalhavam com processos de mineração de dados e extração de conhecimento e, individualmente, esbarravam nos mesmos obstáculos e os solucionavam por meio do método tentativa e erro.

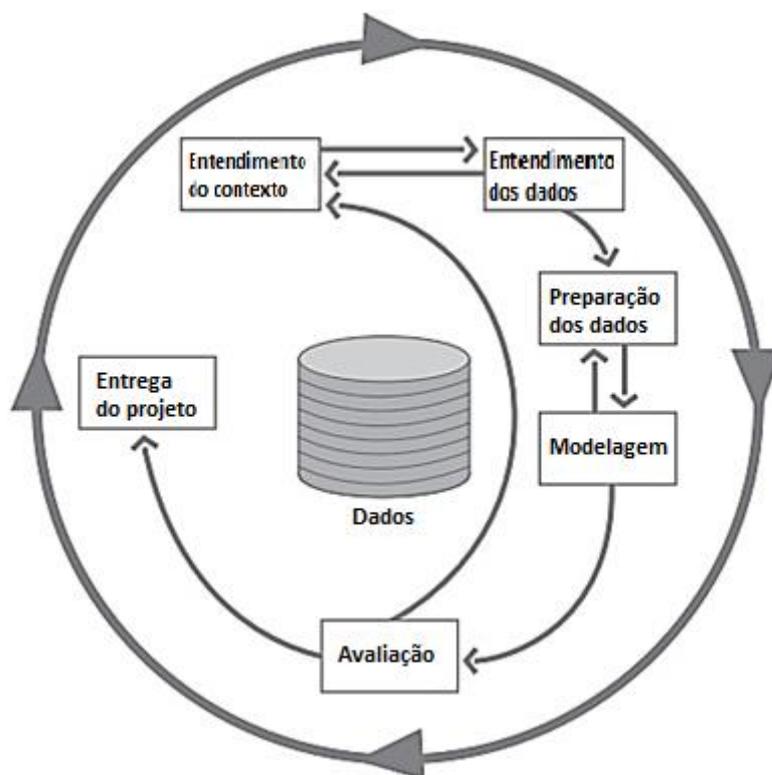
Dessa forma, com a perspectiva de avanço promissor da mineração de dados, viu-se a necessidade de estabelecer um método para padronizar o processo, de modo que fosse disponível livremente para as três empresas e todos os outros que tivessem o mesmo intuito. Dessa forma, por meio de um consórcio, as três formaram o CRISP-DM, convidando partes interessadas da indústria para opinar sobre o que o modelo final deveria possuir. No ano de 2000, o manual de uso do método foi distribuído livremente em meios eletrônicos.

Os criadores do método valorizam o CRISP-DM devido seus criadores e partes envolvidas serem do mercado e entenderem as necessidades do consumidor. Além disso, seu projeto de criação foi amplamente divulgado para a comunidade, inclusive com sua participação, dando caráter aberto ao seu processo.

Em relação ao método em si, suas seis fases são consideradas as etapas do ciclo vital de um projeto de mineração de dados. Cada uma dessas possui tarefas específicas e se relacionam com as outras fases. Tal ciclo vital pode ser observado na Figura 4. As fases são: entendimento do contexto (*business understanding*), entendimento dos dados (*data understanding*), preparação dos dados (*data preparation*), modelagem (*modeling*), avaliação (*evaluation*), entrega do projeto (*deployment*).

Os criadores do método ressaltam que ele não é rígido, havendo flexibilidade para pular entre fases. Ademais, todo projeto de mineração de dados é cíclico e não é encerrado quando uma solução é entregue, pois a resposta entregue pode gerar outros problemas que necessitam de novas estratégias de abordagem e, assim, sempre haverá necessidade de revisão do processo de mineração. Dessa forma, os projetos de mineração subsequentes se beneficiarão dos anteriores.

Figura 4 – Ciclo de vida de um projeto de mineração de dados conforme método CRISP-DM



Fonte: Traduzido de Chapman et al. (2000, p. 12).

É necessário lembrar que este tipo de método foi criado para empresas extraírem dados para suas atividades específicas. Por isso, a nomenclatura pode soar diferente para o ramo de ciências de saúde. Assim, é necessário entender que o CRISP-DM tem um método organizado para preparação dos dados, limpeza da base e aplicação de modelos inteligentes, quando necessários.

Outros projetos da área de ciências da saúde já usaram tal método de forma a identificar resultados positivos para suas áreas de estudo. Por exemplo, Hernández e Corales (2012) desenvolveram um sistema de apoio ao diagnóstico de

hipertensão em pacientes hospitalizados e Kebede, Zegeye e Zeleke (2017) utilizaram o método para identificar mudanças na contagem de linfócitos T-CD4 em pacientes HIV positivos.

Além disso, Rivo et al. (2012) criaram um sistema de apoio a decisão clínica para cirurgia de pulmão e Laureano, Caetano e Cortez (2014) fizeram uso de tais técnicas para previsão do tempo de estadia num hospital. Por fim, o método ainda foi usado para determinar parâmetros em pacientes com doença renal crônica (TAHMASEBIAN et al., 2017).

Ao entender a importância e aplicabilidade desse processo dentro das ciências da saúde, faz-se necessário brevemente explicar cada uma das fases que compõem o método. A primeira fase é chamada de entendimento do contexto, nela o pesquisador deve entender os objetivos do projeto e em qual momento a mineração de dados está inserida dentro desse processo. Dentre suas fases pode-se destacar: determinar os objetivos do projeto, avaliar a situação, determinar os objetivos da mineração de dados e criar o plano do projeto.

A segunda fase é o entendimento dos dados, nessa fase há a aquisição dos dados e familiarização com os mesmos, identificando potencialidades e obstáculos que os mesmos trazem. Os processos dessa fase são: coletar os dados iniciais, limpar os dados, construir dados, integrar dados e formatar os dados.

A terceira fase envolve a preparação dos dados, nela são inseridas as atividades para a criação do *dataset* final a partir dos dados iniciais. Nessa fase são incluídos selecionar os dados, limpar os dados, construir dados, integrar dados, formatar dados.

A quarta fase é a de modelagem. Em tal fase serão aplicadas as ferramentas de modelagem dos dados e os parâmetros calibrados a fim de resultar valores ótimos. As ferramentas são escolhidas de acordo com o tipo de dado e os objetivos da modelagem, daí a necessidade de que a etapa de preparação de dados seja cuidadosamente realizada. As etapas dessa fase são: selecionar as técnicas de modelagem, gerar o desenho do teste, criar o modelo e avaliar o modelo.

Avaliação é a quinta fase do CRISP-DM. Ela necessita que o modelo seja cuidadosamente avaliado para que atinja os resultados propostos pela mineração de

dados. Nela, pode ser determinado quais procedimentos foram tomados e se eles precisam ser revistos. Dentre suas fases, consideram-se: avaliar os resultados, revisar o processo e determinar os próximos passos.

A sexta, e última, fase é a entrega do projeto. Essa fase é inserida por entender que a entrega do modelo não encerra o projeto em si, pois o mesmo precisa ser colocado em prática e observado o seu desempenho. As suas fases são: planejar a entrega, planejar o monitoramento e manutenção, planejar o relatório final e revisar o projeto.

Essas fases e processos podem ser observadas na Figura 5 que mostra o CRISP-DM de forma detalhada. Todas elas foram seguidas pelo pesquisador e podem ser observadas na sessão de resultados dessa pesquisa. As fases do projeto estão destacadas no cabeçalho, cada uma delas com suas tarefas (em negrito) e resultados (em itálico).

Figura 5 - Fases, tarefas e resultados propostos pelo método CRISP-DM

Entendimento do contexto	Entendimento dos dados	Preparação dos dados	Modelagem	Avaliação	Entrega do projeto
Determine os objetivos do projeto <i>Conhecimento prévio</i> <i>Objetivos do projeto</i> <i>Critérios para sucesso do projeto</i>	Colete os dados iniciais <i>Relatório de coleta de dados inicial</i>	Selecione os dados <i>Razão para inclusão ou exclusão</i>	Selecione a técnica de modelagem <i>Técnica de modelagem</i> <i>Suposições da modelagem</i>	Avalie o modelo <i>Avaliação dos resultados da mineração de dados em relação ao critério de sucesso</i> <i>Modelos aprovados</i>	Planeje a entrega <i>Plano de entrega</i>
Avalie a situação <i>Inventário de recursos</i> <i>Exigências, suposições, restrições</i> <i>Riscos e orçamento</i> <i>Terminologia</i> <i>Custos e benefícios</i>	Descreva os dados <i>Relatório de descrição dos dados</i>	Limpe os dados <i>Relatório de limpeza de dados</i>	Gere o desenho dos testes <i>Desenho do teste</i>	Revise o processo <i>Revisão do processo</i>	Planejamento para o monitoramento e manutenção <i>Plano de monitoramento e manutenção</i>
Determine os objetivos da mineração de dados <i>Objetivos da mineração de dados</i> <i>Critérios de sucesso da mineração de dados</i>	Explore os dados <i>Relatório de exploração dos dados</i>	Construa os dados <i>Atributos derivados</i> <i>Registros gerados</i>	Crie o modelo <i>Configuração dos parâmetros</i> <i>Modelos</i> <i>Descrição dos modelos</i>	Determine os próximos passos <i>Lista de possíveis ações</i> <i>Decisão</i>	Produza o relatório final <i>Relatório final</i> <i>Apresentação final</i>
Produza um plano de projeto <i>Plano do projeto</i> <i>Avaliação inicial de ferramentas e técnicas</i>	Verifique a qualidade dos dados <i>Relatório de qualidade dos dados</i>	Integre os dados <i>Dados unidos</i> Formate os dados <i>Dados reformulados</i> <i>Dataset</i> <i>Descrição do dataset</i>	Avalie o modelo <i>Avaliação do modelo</i> <i>Revisão das configurações</i>	Revise o projeto <i>Documentação da experiência</i>	

Fonte: Traduzido de Chapman et al. (2000, p. 14).

A seguir, mostrar-se-á uma breve descrição de cada uma dessas fases tarefas:

Entendimento do contexto

Determine os objetivos do projeto: essa tarefa requer que se conheça minuciosamente o que se pretende atingir.

Avalie a situação: conhecer, em detalhe, os recursos, as restrições, suposições e outros fatores que podem ser considerados para se atingir o sucesso da mineração de dados do projeto.

Determine os objetivos da mineração de dados: descrever os objetivos da mineração de dados utilizando terminologia própria.

Produza um plano de projeto: especificar um plano para os objetivos da mineração de dados. O plano deve constar cada passo que será tomado durante a execução do projeto, incluindo ferramentas e técnicas.

Entendimento dos dados

Colete os dados iniciais: adquirir o conjunto de dados necessário para prosseguir com o projeto, listando como se deu o processo.

Descreva os dados: examinar a parte “grosseira” ou “superficial” dos dados e reportar os resultados.

Explore os dados: iniciar processo de análise, verificar associações com atributo chave (variável desfecho) ou análises estatísticas simples.

Verifique a qualidade dos dados: verificar se os dados possuem qualidade, verificando erros, completude, número de dados ausentes e quão comum eles são.

Preparação dos dados

Selecione os dados: decidir quais dados serão usados na análise, quais são os atributos e observações que possuem relevância para cumprir os objetivos da mineração de dados.

Limpe os dados: aumentar o nível de qualidade dados por meio do processo de limpeza. Descrever que decisões embasaram o processo de limpeza de dados.

Construa os dados: construir e preparar operações como a produção de novos atributos por meio dos existentes ou transformar os atributos existentes com novos valores.

Integre os dados: combinar múltiplos conjuntos de dados, caso seja utilizado mais de um banco.

Formate os dados: utilizar transformações sintáticas para que os dados não mudem seus significados.

Modelagem

Selecione a técnica de modelagem: selecionar as técnicas necessárias. Caso seja utilizada mais de uma, faz-se o necessário o detalhamento de cada técnica.

Gere o desenho dos testes: gerar o procedimento de testar a qualidade e validade dos modelos.

Crie o modelo: rodar as ferramentas de modelagem no conjunto de dados, gerando um ou mais modelos.

Avalie o modelo: interpretar os resultados de acordo com os desenhos dos testes. Pode-se tentar ranquear os modelos a fim de definir o que melhor representa os objetivos do projeto.

Avaliação

Avalie o modelo: diferente do anterior, esse processo visa avaliar o quanto o modelo vai de encontro aos objetivos da mineração de dados e procurar determinar se há alguma razão para que o modelo possa se apresentar deficiente.

Revise o processo: revisar minuciosamente todo o processo para verificar se houve algum processo que mereça nova abordagem, foi pulado ou precise ser repetido.

Determine os próximos passos: o time precisa decidir se o projeto foi finalizado ou como ele deve proceder. Faz-se necessário listar ações futuras e o motivo de escolhê-las.

Entrega do projeto

Planeje a entrega: determinar estratégia de entrega do projeto. Pode-se acrescentar um resumo das estratégias selecionadas e como foram realizadas.

Planejamento para o monitoramento e manutenção: montar estratégia de monitoramento e manutenção do projeto para evitar períodos de uso incorreto dos resultados ou aplicações desenvolvidas. Faz-se necessário que ele seja detalhado para cada um dos objetivos propostos.

Produza o relatório final: escrever o relatório final com a inclusão e organização dos resultados. Pode-se solicitar uma reunião para apresentação final das conclusões com as partes envolvidas.

Revise o projeto: examinar os resultados positivos e negativos durante o projeto. Pode-se falar da experiência adquirida durante o projeto.

5.2 TIPO DE ESTUDO

Estudo epidemiológico, transversal, embasado na aplicação da mineração de dados e aplicação de IA em bancos de dados de saúde. Estudos transversais, ou seccionais, são aqueles que fornecem a informação da prevalência de determinado objeto de pesquisa em um momento, eles fornecem uma “foto” da situação investigada (OLIVEIRA FILHO, 2015).

Além disso, o uso de bancos de dados é mostrado como positivo no que tange à ampla cobertura populacional, o baixo custo para a coleta de informações, e a possibilidade de seguimento de informações (COELI, 2010).

Para Chiavegatto Filho (2015) o uso de grandes bancos de dados (*big data*) em ciências da saúde considera uma quantidade de dados suficientemente grande, que possibilite o uso de técnicas estatísticas não tradicionais. Essa área é

uma oportunidade para epidemiologistas que buscam novas formas de análises de seus dados.

5.3 LOCAL DO ESTUDO

O estudo foi realizado no estado do Ceará.

5.4 DADOS E FONTE DE DADOS

Esse estudo utilizou fontes de dados secundários provenientes do banco de dados do Sistema de Informação de Agravos de Notificação para TB (SINAN-TB). O banco de dados da TB é composto por diversas variáveis criadas a partir de duas fichas. A primeira delas é a ficha de notificação da TB (ANEXO A), na qual o profissional de saúde que identifica o caso com tal condição, além de encaminhar para exames e iniciar o tratamento, a preenche com os dados informados pela pessoa. Para essa pesquisa, foram utilizados os dados sociodemográficos, econômicos, clínicos e de morbidade da ficha. Tais variáveis foram consideradas preditoras.

A segunda ficha é a ficha de acompanhamento da TB (Anexo B). Nela, o profissional de saúde acompanha o progresso do paciente durante seu tratamento. Suas variáveis também entrarão como preditoras no modelo a ser criado. Além disso, na mesma ficha de acompanhamento há uma variável nomeada “situação de encerramento”, que descreve a forma do término do tratamento. Essa variável foi considerada o desfecho desse estudo.

A variável desfecho “situação de encerramento”, variável 62 da ficha de acompanhamento da TB, pode ser preenchida de dez formas diferentes. Sendo elas: (1) cura, (2) abandono, (3) óbito por TB, (4) óbito por outras causas, (5) transferência, (6) mudança de diagnóstico, (7) TB-DR, (8) mudança de esquema, (9) falência, (10) abandono primário.

Foram incluídas todas as fichas com registro do banco. Foram excluídos somente aqueles que não tiveram o campo “situação de encerramento” devidamente preenchido, já que ele foi considerado o desfecho do estudo.

5.5 ANÁLISE DE DADOS

Para a análise de dados, inicialmente foi necessário converter o arquivo recebido em .dbf (database file – arquivo de banco de dados) para um arquivo .dta (Stata data file) e um .csv (comma separated value – valor separado por vírgula).

Após ser colocado no formato correto, o arquivo foi manuseado nos softwares Stata/SE v. 12, *Joinpoint Regression Program* v.4.6.0.0 e WEKA v. 3.8. Todos esses softwares são amplamente utilizados pela comunidade científica de acordo com suas funções. O Stata é um dos softwares mais utilizados em análise de dados em saúde, o JoinPoint é utilizado para análise de séries temporais e o WEKA é utilizado como ferramenta intuitiva para mineração de dados e aplicação de algoritmos inteligentes.

Inicialmente, o programa *Joinpoint Regression* foi utilizado para verificar o padrão temporal de indicadores epidemiológicos da TB (incidência da TB, prevalência da TB, taxa de cura, taxa de abandono, taxa de óbito – letalidade, coeficiente de mortalidade e taxa de resistência). Para isso foram consideradas como variáveis dependentes as taxas e coeficientes estudados e o ano de ocorrência como variável independente. Para realizar essas análises, foi utilizado o banco de dados já limpo e somente com os valores agregados de cada encerramento, em cada ano, juntamente com os dados das populações do estado encontrados na página eletrônica do IBGE.

Para determinar a taxa de cura, abandono, óbito (letalidade) e resistência foi utilizado o número de pessoas que tiveram cada encerramento dividido pelo número total de casos. A incidência da TB foi calculada considerando o número de casos novos da doença ajustado para 100.000 habitantes, a prevalência foi calculada com o número total de casos por ano e também foi ajustado para 100.000 habitantes. O coeficiente de mortalidade calculado pelo número de óbitos por TB dividido pela população do estado e ajustado para uma população de 100.000 habitantes.

A análise que o programa *Joinpoint Regression* realiza é do tipo linear segmentada, com transformação logarítmica dos valores. Nele, é testado se um ou mais pontos devem ser adicionados ao modelo linear por meio do método de permutação de Monte Carlo, ou seja, é avaliado se uma linha com múltiplos

segmentos descreve o modelo melhor do que apenas uma linha reta. Por esse motivo a análise também é conhecida por regressão por pontos de inflexão. Além disso, calculou-se a variação percentual anual (Annual Percentage Change – APC), com um intervalo de confiança de 95% (IC95%), onde um valor negativo da APC indica tendência decrescente e um valor positivo, tendência crescente.

Ademais, ao final do período é possível obter a variação percentual anual média (Average Annual Percentage Change – AAPC), que demonstra como ocorreu a mudança durante todo o período estudado. Caso haja mais de um ponto de inflexão a AAPC considerará todos para o seu cálculo. Caso contrário, o valor da AAPC será igual ao da APC.

Cada ponto de inflexão adicionado ao modelo significa uma mudança da tendência linear. Ajustou-se o modelo, assumindo que o número de joinpoints era diferente, podendo variar de zero (apenas um segmento) até dois (três segmentos) ao decorrer dos anos. Estabeleceu-se um nível de significância de 5% para testar a hipótese nula de que a APC e a AAPC da série são iguais a zero. Por fim, foram gerados gráficos e tabelas inseridos nos resultados do estudo.

O próximo tipo de análise realizada foi a epidemiológica. Nela foram verificadas as frequências simples e relativas, medidas de tendência central e dispersão dos dados. Após esse passo, realizou-se a análise bivariada dos dados, considerando como cada variável, individualmente, se relaciona com cada um dos desfechos (cura, abandono, óbito e resistência). No caso desse estudo alguns testes foram aplicados a fim de verificar esse comportamento.

Para todas as variáveis foram utilizadas regressões bivariadas (logística, Poisson e binomial negativa), verificou-se apenas o p-valor uma vez que para esta pesquisa os resultados multivariáveis seriam os importantes. Dessa forma, todas as variáveis que apresentaram $p \leq 0,20$ foram consideradas para o próximo passo. Os valores de cada associação foram inseridos em tabelas por desfecho pesquisado.

A seguir, as variáveis que restaram foram incluídas em modelos multivariáveis, dessa forma foi verificado como todas as variáveis, em conjunto, influenciavam o desfecho. Assim, considerou-se significante aquelas que apresentaram $p < 0,05$. Sabendo que apenas o p-valor é uma medida que traz

pouca informação, analisou-se a força de associação por meio dos intervalos de confiança de 95% (IC95%).

Para a regressão logística interpretou-se o resultado da razão de chances (*Odds Ratio* – *OR*) como razão de prevalência (RP) uma vez que este estudo é do tipo transversal. Para as regressões de Poisson e binomial negativa interpretou-se a incidência de risco relativo (*Incidence of Relative Risk* – *IRR*) como RP, isso foi possível graças a inserir uma variável de tempo com valor igual a um para todas as observações.

Ademais, para os modelos de Poisson e binomial negativo foi incluído um fator de correção chamado de estimador de variância robusta. Ele foi inserido devido tais regressões serem amplamente utilizadas para desfechos numéricos ou ordinais. Esse fator, então, corrige os possíveis erros de tais regressões e ainda cooperam com resultados mais precisos e intervalos de confiança mais estreitos. Todas as análises epidemiológicas foram realizadas por meio do software Stata/SE. A comparação dos resultados dessas três regressões, por desfecho, pode ser observada em tabelas nos resultados do estudo.

Além de comparar qual resultado é mais fidedigno para ser utilizado como fatores que se relacionam com a prevalência de cada desfecho da TB, as análises multivadas foram utilizadas para selecionar quais variáveis seriam colocadas na fase final de análise de dados: a criação de modelo preditivo. Esse modelo foi considerando como desfecho todas as quatro situações anteriores mais uma denominada “outros”.

Assim, o banco foi exportado do Stata como um documento .csv e importado no programa WEKA para análise de conhecimento. Inicialmente houve uma fase de pré-processamento de dados, com a verificação da normalidade dos dados e discretização de dados, quando necessário.

Após isso, a etapa de criação de um modelo de classificação do desfecho foi selecionada. Denominou-se como desfecho a variável com as cinco classificações e as demais como preditoras. Utilizou-se o algoritmo de redes neurais artificiais. Esse algoritmo possui diversos tipos, dentre eles o *perceptron* multicamadas, no qual o modelo preditivo foi embasado. O *perceptron* contém como atributos de entrada (input) as variáveis preditoras e como saída (output) a decisão

de qual desfecho o paciente poderá apresentar. Os nós da camada oculta, assim, realizaram as análises por meio da função de ligamento sigmoide e o desfecho escolhido foi o resultado da interação entre esses nós.

Como funções do *perceptron* foram selecionados visualização da interface gráfica do usuário (do inglês *Graphical User Interface – GUI*), entrada dos dados por mini lotes (*minibatch*), *learning decay*, taxa de aprendizado (*learning rate*), *momentum*, e o número de épocas (*epochs*). Somente as funções de criação automática do modelo e número de camadas ocultas com seu número de nós foi deixado de modo automático.

Separou-se o banco em duas partes: treino e teste. Isso quer dizer que uma parte do banco foi utilizada para mostrar os padrões para a máquina e ela aprender com isso e a outra parte foi utilizada para testar a máquina e conferir se ela aprendeu a reconhecer os padrões e acertar nas predições. Para esse estudo foi adotado um percentual de 90% do banco para treino e 10% para teste.

Para avaliar a adequação dos modelos foram considerados os critérios de sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo. Esses critérios podem ser obtidos por meio de uma matriz de confusão, essa matriz indica quais dos reais valores foram corretamente preditos (Figura 6). Além desses também foi utilizada a avaliação da curva ROC (*Receiver Operator Characteristic – Curva de Características de Operação do Receptor*) e acurácia do modelo.

Figura 6 – Matriz de Confusão

		Valor Predito	
		Não	Sim
Valores Reais	Não	Verdadeiro Negativo	Falso Negativo
	Sim	Falso Positivo	Verdadeiro Positivo

Fonte: Próprio autor

O último teste para o modelo preditivo foi o de validação cruzada. Em tal teste, divide-se o banco em um número *n* de partes (*folds*) e esse é testado em relação as outras partes do banco. Decidiu-se utilizar o critério de 10-fold, ou seja, uma parte era testada com as outras nove para verificar se havia concordância dos

modelos e nesse caso não houve separação em treino, fazendo-se uso de todo o banco. Os resultados das predições, juntamente com as matrizes de confusão e uma imagem da abstração dessa rede neural podem ser observados nos resultados do estudo.

5.6 ASPECTOS ÉTICOS

Este projeto foi construído e conduzido de acordo com os aspectos éticos estabelecidos pela resolução 466/2012 do Conselho Nacional de Saúde, que trata de pesquisa envolvendo seres humanos (BRASIL, 2013). O projeto foi submetido na Plataforma Brasil e encaminhado para Comitê de Ética em Pesquisa (CEP) da Universidade Estadual do Ceará (UECE).

O projeto foi aprovado por tal comitê por meio do CAAE: 88396718.1.0000.5534 e parecer número: 2.687.046 (ANEXO C). Embora seja uma pesquisa com dados secundários, ela envolve seres humanos; então, a responsabilidade do pesquisador em manter o sigilo dos dados das pessoas cujos registros estão no banco de dados foi mantida durante todo o processo da pesquisa.

6 RESULTADOS

Os resultados desse estudo serão apresentados pela aplicação de todas as fases e processos propostos pelo referencial metodológico do CRISP-DM. Alguns desses são semelhantes aos apontados em outros capítulos deste relatório, assim, quando oportuno, será indicado o capítulo que a descrição se encontra.

Entendimento do contexto

Determine os objetivos do projeto

Como consta nos objetivos, essa pesquisa foi guiada pelo objetivo: aplicar o método CRISP-DM para identificar a epidemiologia da TB no Ceará e criar um modelo preditivo para os desfechos cura, abandono do tratamento, óbito por TB e resistência ao tratamento.

Avalie a situação

A atual situação epidemiológica da TB como problema de saúde pública local, nacional e global, além das potencialidades de métodos de mineração de dados com aplicação de algoritmos de IA podem ser verificadas nos capítulos de introdução e revisão desta investigação. A importância de cada desfecho também está previamente descrita, assim como o impacto que possuem no processo saúde-doença das populações.

Determine os objetivos da mineração dos dados

Dentre os objetivos de aplicar técnicas de mineração de dados para os desfechos da TB, três podem ser destacados, sendo eles:

- Melhorar o cuidado ao paciente com TB
- Diminuir os gastos no cuidado a esse paciente
- Focar nos pacientes que possuem mais chances de apresentar desfechos negativos durante o tratamento para a TB

Produza um plano de projeto

O plano desse projeto foi previamente descrito na sessão método nos tópicos 5.4 dados e fonte de dados e 5.5 análise de dados.

Entendimento dos dados

Colete dados iniciais

O banco de dados de TB do estado do Ceará foi coletado na Secretaria de Saúde do Estado do Ceará (SESA). Os dados disponíveis foram aqueles de 2001 a 2017.

Para que o banco fosse disponibilizado pela Secretaria, foi necessário o envio do projeto de pesquisa para a apreciação e assinatura dos termos de fiel depositário, anuência, termo de autorização do setor responsável. Após a aquisição desses termos o projeto foi submetido à plataforma Brasil para recebimento de parecer do Comitê de Ética em Pesquisa (CEP) da Universidade Estadual do Ceará.

Com os termos e autorização do CEP, o pesquisador dirigiu-se novamente à SESA, para a transferência do banco para um CD a fim de se iniciar as análises. Ressalta-se que nenhuma variável de identificação do paciente foi solicitada pelo pesquisador.

Descreva os dados

Os dados cedidos pela SESA encontravam-se em um único arquivo de formato .dbf, em forma de uma única tabela e o documento possuía cerca de 50MB. Para iniciar as análises, o arquivo foi transformado para o formato .dta, a fim de ser lido no software estatístico Stata/SE.

Ao ser carregado no Stata, identificou-se que o arquivo em sua versão original possuía 109 variáveis (atributos) com 75.948 observações (instâncias). Como grande parte das variáveis estavam em formato “string”, elas foram transformadas em “numeric” ou “float” de acordo com cada variável. Além disso, foram adicionados rótulos a algumas variáveis, conforme dicionário de variáveis de TB.

Mesmo o arquivo contendo um grande número de atributos, nem todos foram utilizados nesse projeto, por não satisfazerem a proposta; por exemplo, dados de georreferenciamento, dados de identificação de outros tipos de tratamento, e dados das instituições de saúde não foram considerados.

Explore os dados

Nessa fase, é necessário que se obtenha uma visão geral dos dados estudados e uma descrição da população estudada. Entretanto, para dar uma visão mais acurada dos dados, algumas técnicas de limpeza precisam ser aplicadas nos próximos passos, assim, decidiu-se mostrar esses valores após tais etapas.

Verifique a qualidade dos dados

Por meio da análise anterior foi possível identificar que muitas variáveis são preenchidas de forma incompleta, pois há muitos dados ausentes. Na tabela a seguir (Tabela 1), as variáveis estão listadas com a porcentagem de dados ausentes em cada uma delas. Sua qualidade foi avaliada segundo a porcentagem de dados ausentes apresentada.

Tabela 1 – Porcentagem de dados ausentes de cada variável do estudo. Fortaleza-CE, 2018.

Variável	Dados ausentes (%)
Ano	0,00
Data do diagnóstico	0,00
Idade	0,00
Sexo	0,03
Raça	12,41
Escolaridade	8,99
Zona	3,09
Tipo de entrada	0,04
Institucionalizado	51,97
Raio X de tórax	1,30
Teste tuberculínico	21,44
Forma	0,04
Agravo AIDS	44,65

Agravo Álcool	36,95
Agravo diabetes	38,83
Agravo doenças	38,88
Agravo outro	41,42
Baciloscopia de escarro 1	0,04
Baciloscopia de escarro 2	51,99
Baciloscopia de outro material	20,43
Cultura de escarro	0,04
Cultura de outro material	20,30
Exame de HIV	0,01
Exame histopatológico	2,53
Data de início do tratamento	0,00
Rifampicina	19,79
Isoniazida	19,80
Pirazinamida	19,92
Etambutol	21,77
Estreptomicina	22,92
Etionamida	23,03
Outras drogas	26,34
Tratamento Supervisionado	18,92
Número de contatos	29,59
Doença do trabalho	27,82
Baciloscopia 1º mês	38,89
Baciloscopia 2º mês	18,11
Baciloscopia 3º mês	41,27
Baciloscopia 4º mês	21,29
Baciloscopia 5º mês	43,53
Baciloscopia 6º mês	24,03
Tratamento Supervisionado 2	32,58
Número de contatos examinados	29,44
Situação de encerramento	2,56
Data de encerramento	0,00
População privada de liberdade	78,09

População de rua	78,15
Profissional de saúde	78,17
População imigrante	78,18
Beneficiário do governo	78,57
Agravo drogas	77,55
Agravo tabagismo	77,57
Teste molecular	80,52
Teste de sensibilidade	93,90
Uso de antirretrovirais	94,77
Baciloscopia após 6 meses	89,08

Fonte: SINAN-TB (CE)

Preparação dos dados

Selecione os dados

Para decidir quais variáveis utilizar nos modelos, considerou-se a porcentagem de dados ausentes da tabela anterior. A fim de evitar perda na qualidade dos resultados, optou-se por não utilizar as variáveis com mais que 25% de dados ausentes.

Assim, no quadro a seguir (Quadro 1) encontra-se a decisão do pesquisador acerca do uso de cada uma das variáveis. Ao final, 32 variáveis originais foram escolhidas para a criação dos modelos.

Quadro 1 – Decisão sobre uso de variáveis do banco de dados

Variável	Dados ausentes (%)	Decisão do pesquisador
Ano	0,00	Utilizar
Data do diagnóstico	0,00	Utilizar
Idade	0,00	Utilizar
Sexo	0,03	Utilizar
Raça	12,41	Utilizar
Escolaridade	8,99	Utilizar
Zona	3,09	Utilizar

Tipo de entrada	0,04	Utilizar
Institucionalizado	51,97	Não utilizar
Raio X de tórax	1,30	Utilizar
Teste tuberculínico	21,44	Utilizar
Forma	0,04	Utilizar
Agravo AIDS	44,65	Não utilizar
Agravo Álcool	36,95	Não utilizar
Agravo diabetes	38,83	Não utilizar
Agravo doenças	38,88	Não utilizar
Agravo outro	41,42	Não utilizar
Baciloscopia de escarro 1	0,04	Utilizar
Baciloscopia de escarro 2	51,99	Não utilizar
Baciloscopia de outro material	20,43	Utilizar
Cultura de escarro	0,04	Utilizar
Cultura de outro material	20,30	Utilizar
Exame de HIV	0,01	Utilizar
Exame histopatológico	2,53	Utilizar
Data de início do tratamento	0,00	Utilizar
Rifampicina	19,79	Utilizar
Isoniazida	19,80	Utilizar
Pirazinamida	19,92	Utilizar
Etambutol	21,77	Utilizar
Estreptomina	22,92	Utilizar
Etionamida	23,03	Utilizar
Outras drogas	26,34	Utilizar
Tratamento Supervisionado	18,92	Utilizar
Número de contatos	29,59	Não utilizar
Doença do trabalho	27,82	Não utilizar
Baciloscopia 1º mês	38,89	Não utilizar
Baciloscopia 2º mês	18,11	Utilizar
Baciloscopia 3º mês	41,27	Não utilizar
Baciloscopia 4º mês	21,29	Utilizar
Baciloscopia 5º mês	43,53	Não utilizar

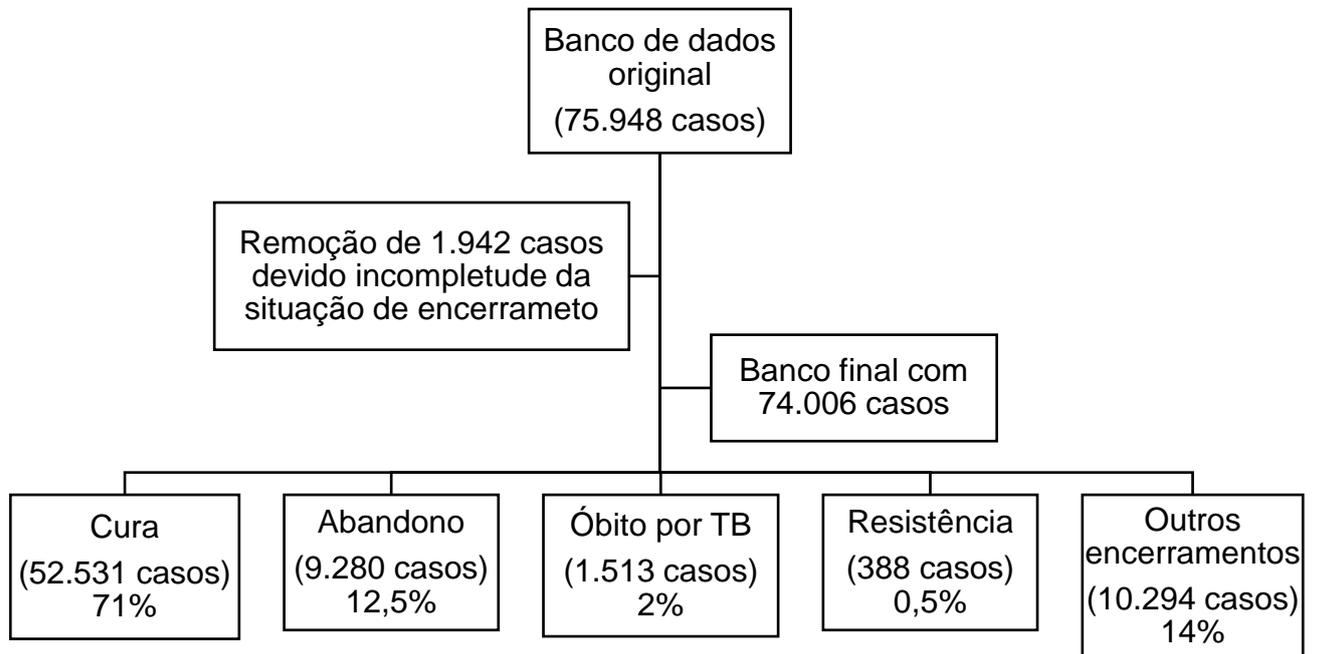
Baciloscopia 6 ^o mês	24,03	Utilizar
Tratamento Supervisionado 2	32,58	Não utilizar
Número de contatos examinados	29,44	Não utilizar
Situação de encerramento	2,56	Utilizar
Data de encerramento	0,00	Utilizar
População privada de liberdade	78,09	Não utilizar
População de rua	78,15	Não utilizar
Profissional de saúde	78,17	Não utilizar
População imigrante	78,18	Não utilizar
Beneficiário do governo	78,57	Não utilizar
Agravo drogas	77,55	Não utilizar
Agravo tabagismo	77,57	Não utilizar
Teste molecular	80,52	Não utilizar
Teste de sensibilidade	93,90	Não utilizar
Uso de antirretrovirais	94,77	Não utilizar
Baciloscopia após 6 meses	89,08	Não utilizar

Fonte: Próprio autor

Limpe os dados

Além da exclusão de variáveis com muitos *missing values*, também foram excluídos os casos sem situação de encerramento completa. Essa decisão foi tomada devido tal variável ser a mais importante do estudo e sua completude possibilitar a futura criação dos modelos preditivos. Portanto, foram excluídos 1.942 casos do banco de dados original, resultando em um banco com 74.006 casos, conforme a imagem a seguir (Figura 7).

Figura 7 – Fluxograma de limpeza de dados



Fonte: próprio autor

Como apontado no item de preparação dos dados, após limpo, o banco passou por algumas mudanças e, dessa forma, um detalhamento da população estudada pode ser dado. Na tabela 2, as características gerais e por desfecho podem ser observadas para cada uma das variáveis propostas. A população do estudo apresentou idade média de 39,77 ($\pm 17,85$) anos, com predominância do sexo masculino ($n = 46.537$; 62,9%), cor parda ($n = 42.768$; 66,1%), escolaridade da 5-8 série do fundamental incompleta ($n = 13.439$; 20%), morando na zona urbana ($n = 60.235$; 84%).

Tabela 2 – Características sociodemográficas de pacientes com TB segundo tipo de encerramento. Fortaleza (CE), 2018.

Variáveis	Cura (%)	Abandono (%)	Óbito por TB (%)	Resistência (%)	Outros (%)	Total (%)
Idade (média)	39,5 (±17,7)	36,6 (±15,49)	51,39 (±19,06)	40,73 (±14,22)	42,38 (±19,28)	39,77 (±17,85)
Sexo						
Feminino	20.467 (74,57)	2.862 (10,28)	432 (1,57)	144 (0,52)	3.582 (13,05)	27.447 (37,10)
Masculino	32.052 (68,87)	6.457 (13,87)	1.081 (2,32)	243 (0,52)	6.704 (14,41)	46.537 (62,90)
Raça/Cor						
Branca	7.840 (76,63)	864 (8,44)	164 (1,60)	34 (0,33)	1.329 (12,99)	10.231 (15,80)
Preta	3.571 (71,97)	667 (13,44)	97 (1,95)	50 (1,01)	577 (11,63)	4.692 (7,67)
Amarela	510 (70,25)	88 (12,12)	12 (1,65)	3 (0,41)	113 (15,56)	726 (1,12)
Parda	29.311 (68,53)	5.573 (13,03)	1.123 (2,63)	248 (0,58)	6.513 (15,23)	42.768 (66,07)
Indígena	190 (74,80)	22 (8,66)	6 (2,36)	3 (1,18)	33 (12,99)	254 (0,39)
Escolaridade						
Analfabeto	7.138 (72,17)	1.170 (11,83)	197 (1,99)	46 (0,47)	1.339 (13,54)	9.890 (14,66)
1-4 série fund. incomp.	8.811 (70,99)	1.478 (11,91)	298 (2,40)	75 (0,60)	1.749 (14,09)	12.411 (18,40)
1-4 série fund.	2.543	491	115	34	496	3.679

comp.	(69,12)	(13,35)	(3,13)	(0,92)	(1,48)	(5,46)
5-8 série fund.	9.763	1.818	158	73	1.627	13.439
incomp.	(72,65)	(13,53)	(1,18)	(0,54)	(12,11)	(19,93)
5-8 série fund.	2.224	376	47	18	395	3.060
comp.	(72,68)	(12,29)	(1,54)	(0,59)	(12,91)	(4,54)
Ens médio	4.031	495	38	23	523	5.110
incomp.	(78,88)	(9,69)	(0,74)	(0,45)	(10,23)	(7,58)
Ens médio	2.897	283	41	19	433	3.673
comp.	(78,87)	(7,70)	(1,12)	(0,52)	(11,79)	(5,45)
Ens superior	415	39	5	5	51	515
incomp.	(80,58)	(5,57)	(0,97)	(0,97)	(9,90)	(0,76)
Ens superior	1.145	81	15	5	190	1.436
comp.	(79,74)	(5,64)	(1,04)	(0,35)	(13,23)	(2,13)

Zona de Moradia

Urbana	42.015	8.436	1.266	315	8.203	60.235
	(69,75)	(14,01)	(2,10)	(0,52)	(13,62)	(83,97)
Rural	8.523	528	186	45	1.544	10.826
	(78,73)	(4,88)	(1,72)	(0,42)	(14,26)	(15,09)
Periurbana	333	31	6	2	103	475
	(70,11)	(6,53)	(1,26)	(0,42)	(21,68)	(0,66)

Fonte: SINAN-TB (CE)

Em relação as características clínico-laboratoriais, os usuários deram entrada, principalmente, como casos novos (n = 60.614; 82%), apresentando TB pulmonar (n = 64.524; 87,2%). A maioria apresentou radiografia de tórax suspeita (n = 53.577; 73,3%), não realizou teste tuberculínico (n = 49.006; 83,2%), apresentou baciloscopia de escarro positiva (n = 43.539; 58,8%), não realizou baciloscopia de outro material (n = 56.394; 94,5%), não estavam em critérios para a realização de cultura de escarro (n = 60.043; 81,2%) nem para outro material (n = 56.246; 94,1%). O exame de HIV não foi realizado pela maioria das pessoas (n = 41.270; 55,8%) assim como o exame histopatológico (n = 62.261; 86,2%) (Tabela 3).

Tabela 3 – Características clínico-laboratoriais de pacientes com TB segundo tipo de encerramento. Fortaleza (CE), 2018.

Variáveis	Cura (%)	Abandono (%)	Óbito por TB (%)	Resistência (%)	Outros (%)	Total (%)
Tipo de entrada						
Caso novo	44.905 (74,08)	6.402 (10,56)	1.120 (1,85)	210 (0,35)	7.977 (13,16)	60.614 (81,93)
Recidiva	3.269 (65,02)	728 (14,48)	131 (4,61)	93 (1,85)	807 (16,05)	5.028 (6,80)
Reingresso pós-abandono	1.755 (40,79)	1.712 (39,80)	128 (2,98)	50 (1,16)	657 (15,27)	4.302 (5,81)
Não sabe	509 (60,38)	164 (19,45)	14 (1,66)	5 (0,49)	151 (17,91)	843 (1,14)
Transferência	2.077 (65,79)	273 (8,65)	86 (2,72)	30 (0,95)	691 (21,89)	3.157 (4,27)
Forma da doença						
Pulmonar	45.875 (71,10)	8.265 (12,81)	1.322 (2,05)	364 (0,56)	8.698 (13,48)	64.524 (87,21)
Extra-pulmonar	5.881 (71,79)	827 (10,10)	133 (1,62)	16 (0,20)	1.335 (16,30)	8.192 (11,07)
Mista	760 (59,75)	187 (14,70)	58 (4,56)	8 (0,63)	259 (20,36)	1.272 (1,72)
Radiografia de tórax						
Suspeito	38.165 (71,23)	6.523 (12,18)	1.145 (2,14)	283 (0,53)	7.461 (13,93)	53.577 (73,30)
Normal	2.525 (74,84)	297 (8,80)	32 (0,95)	12 (0,36)	508 (15,06)	3.374 (4,62)
Outra patologia	427 (65,69)	52 (8,0)	27 (4,15)	2 (0,31)	142 (21,85)	650 (0,89)

Não realizado	10.780 (69,58)	2.330 (15,04)	284 (1,83)	86 (0,56)	2.012 (12,99)	15.492 (21,19)
Teste tuberculínico (PPD)						
Suspeito	1.923 (69,93)	314 (11,42)	55 (2,0)	3 (0,11)	455 (16,55)	2.750 (4,67)
Normal	875 (75,89)	118 (10,23)	14 (1,21)	5 (0,43)	141 (12,23)	1.153 (1,96)
Outra Patologia	4.728 (78,71)	548 (9,12)	38 (0,63)	20 (0,33)	673 (11,20)	6.007 (10,20)
Não realizado	35.392 (72,22)	6.181 (12,61)	971 (1,98)	247 (0,50)	6.215 (12,68)	49.006 (83,18)
Baciloscopia de escarro						
Positiva	31.933 (73,34)	5.410 (12,43)	780 (1,79)	285 (0,65)	5,131 (11,78)	43.539 (58,85)
Negativa	10.436 (72,0)	1.398 (9,64)	266 (1,84)	43 (0,30)	2.352 (16,23)	14.495 (19,59)
Não realizada	9.845 (64,08)	2.368 (15,41)	444 (2,89)	50 (0,33)	152 (25,76)	15.364 (20,77)
Não se aplica	302 (51,19)	103 (17,46)	23 (3,90)	10 (1,69)	152 (25,76)	590 (0,80)
Baciloscopia de outro material						
Positiva	1.123 (72,83)	157 (10,18)	34 (2,20)	6 (0,39)	222 (14,40)	1.542 (2,58)
Negativa	1.210 (69,74)	195 (11,24)	39 (2,25)	5 (0,29)	286 (16,48)	1.753 (2,91)
Não realizada	41.071 (72,83)	6.898 (12,23)	1.024 (1,82)	272 (0,48)	7.129 (12,64)	56.394 (94,51)
Cultura de escarro						

Positiva	5.228	972	130	190	923	7.443
	(70,24)	(13,06)	(1,75)	(2,55)	(12,40)	(10,06)
Negativa	3.097	283	69	19	568	4.036
	(76,73)	(7,01)	(1,71)	(0,47)	(14,07)	(5,45)
Não realizada	1.553	255	36	16	606	2.466
	(62,98)	(10,34)	(1,46)	(0,65)	(24,57)	(3,33)
Não se aplica	42.638	7.769	1.278	163	8.195	60.043
	(71,01)	(12,94)	(2,13)	(0,27)	(13,65)	(81,15)
Cultura de outro material						
Positivo	892	130	22	6	154	1.204
	(74,09)	(10,80)	(1,83)	(0,50)	(12,79)	(2,01)
Negativo	719	129	24	4	176	1.052
	(68,35)	(12,26)	(2,28)	(0,38)	(16,73)	(1,76)
Não realizada	884	143	22	8	206	1.263
	(69,99)	(11,32)	(1,74)	(0,63)	(16,31)	(2,11)
Não se aplica	40.993	6.842	1.036	264	7.111	56.246
	(72,88)	(12,16)	(1,84)	(0,47)	(12,64)	(94,11)
HIV						
Positivo	1.755	894	201	21	1.082	3.953
	(44,40)	(22,62)	(5,08)	(0,53)	(27,37)	(5,34)
Negativo	19.975	2.447	480	221	3.110	26.233
	(76,14)	(9,33)	(1,83)	(0,84)	(11,86)	(35,45)
Em andamento	1.816	230	35	12	456	2.549
	(71,24)	(9,02)	(1,37)	(0,47)	(17,89)	(3,44)
Não realizado	28.984	5.709	797	134	5.646	41.270
	(70,23)	(13,83)	(1,93)	(0,32)	(13,68)	(55,77)
Histopatológico						
BAAR positivo	2.767	332	79	15	501	3.694
	(74,91)	(8,99)	(2,14)	(0,41)	(13,56)	(5,12)
Sugestivo de	2.623	315	43	2	432	3.415

TB	(76,81)	(9,22)	(1,26)	(0,06)	(12,65)	(4,73)
Não sugestivo de TB	342	51	16	5	99	513
Em andamento	(66,67)	(9,94)	(3,12)	(0,97)	(19,30)	(0,71)
	1.593	268	56	9	390	2.316
	(68,78)	(11,57)	(2,42)	(0,39)	(16,84)	(3,21)
Não realizado	43.912	8.162	1.281	352	8.554	62.261
	(70,53)	(13,11)	(2,06)	(0,57)	(13,74)	(86,22)

Fonte: SINAN-TB (CE)

Quanto ao tratamento, a Tabela 4 mostra que os medicamentos mais utilizados foram a rifampicina (n = 58.882; 97,9%), a isoniazida (n = 58.771; 97,8%) e a pirazinamida (n = 58.463; 97,3%). Pouco mais da maioria realizou tratamento diretamente observado (n = 33.603; 55,3%). As baciloscopias bimestrais apresentaram, principalmente, resultado negativo no 2º (n = 29.281; 47,6%), 4º (n = 28.778; 48,6%) e 6º mês (n = 29.616; 51,67%).

Tabela 4 – Características de tratamento e acompanhamento de pacientes com TB segundo tipo de encerramento. Fortaleza (CE), 2018.

Variáveis	Cura (%)	Abandono (%)	Óbito por TB (%)	Resistênci a (%)	Outros (%)	Total (%)
Tratamento						
Rifampicina	43.280	7.127	1.021	242	7.212	58.882
	(73,50)	(12,10)	(1,73)	(0,41)	(12,25)	(97,87)
Isoniazida	43.200	7.130	1.017	237	7.187	58.771
	(73,51)	(13,13)	(1,73)	(0,40)	(12,23)	(97,71)
Pirazinamida	42.851	7.132	1.019	272	7.189	58.463
	(73,30)	(12,20)	(1,74)	(0,47)	(12,30)	(97,33)
Etambutol	15.249	3.274	560	180	3.297	22.560
	(67,59)	(14,51)	(2,48)	(0,80)	(14,61)	(38,45)
Estreptomicina	594	141	41	29	220	1.025
	(57,95)	(13,76)	(4,0)	(2,83)	(21,46)	(1,77)

Etionamida	759	163	41	25	216	1.204
	(63,04)	(13,54)	(3,41)	(2,08)	(17,94)	(2,08)
Tratamento						
Supervisionado (DOTS)						
Sim	24.302	3.976	644	178	4.530	33.630
	(72,26)	(11,82)	(1,91)	(0,53)	(13,47)	(55,32)
Não	11.359	2.096	389	79	2.121	16.044
	(70,80)	(13,06)	(2,42)	(0,49)	(13,22)	(26,39)
Ignorado	8.536	1.268	94	27	1.190	11.115
	(76,80)	(11,41)	(0,85)	(0,24)	(10,71)	(18,28)
Baciloscopia de 2º mês						
Positivo	2.984	469	56	84	499	4.092
	(72,92)	(11,46)	(1,37)	(2,05)	(12,19)	(6,66)
Negativo	25.839	1.528	138	86	1.690	29.281
	(88,24)	(5,22)	(0,47)	(0,29)	(5,77)	(47,65)
Não realizado	15.868	4.889	700	139	4.101	25.697
	(61,75)	(1,03)	(2,72)	(0,54)	(15,96)	(41,82)
Baciloscopia de 4º mês						
Positivo	512	122	13	70	142	859
	(59,60)	(14,20)	(1,51)	(8,15)	(16,53)	(1,45)
Negativo	26.880	893	76	72	857	28.778
	(93,40)	(3,10)	(0,26)	(0,25)	(2,98)	(48,63)
Não realizado	16.881	5.083	714	140	4.344	27.162
	(62,15)	(18,71)	(2,63)	(0,52)	(15,99)	(45,90)
Baciloscopia de 6º mês						
Positivo	199	73	19	93	121	505
	(39,41)	(14,46)	(3,76)	(18,42)	(23,96)	(0,88)
Negativo	28.779	320	44	45	428	29.616
	(97,17)	(1,08)	(0,15)	(0,15)	(1,45)	(51,67)

Não realizado	14.692	4.991	710	124	4.300	24.817
	(59,20)	(20,11)	(2,86)	(0,50)	(17,33)	(43,30)

Fonte: SINAN-TB (CE)

Em relação aos desfechos, alguns dados merecem destaque. Por exemplo, das 27.447 mulheres com TB, 74,6% alcançaram a cura do tratamento, comparado a 68,9% dos homens (32.052/46.537). O sexo masculino apresentou maiores taxas de abandono e óbito por TB. Indivíduos de raça branca foram os que alcançaram maiores taxas de cura (76,6%), já as maiores taxas de abandono ficaram com os indivíduos de cor preta (13,4%) e o óbito com indivíduos pardos (2,6%). As maiores taxas de cura estão com indivíduos de maior escolaridade, e as menores taxas com os desfechos negativos (abandono e óbito) (Tabela 2).

Além disso, maiores taxas de cura (78,7%) são detectadas em indivíduos da zona rural e maiores taxas de abandono (14%), óbito (2,1%) e resistência (0,52%) em indivíduos na zona urbana (Tabela 2). Quanto ao tipo de entrada, casos novos apresentam mais taxas de cura (74,1%), enquanto casos de reingresso pós-abandono apresentam mais taxas de um novo abandono do tratamento (39,8%). O óbito foi mais comum (4,6%) em casos de recidiva, assim como resistência (1,85%) (Tabela 3).

Em relação à entrada, a porcentagem de cura foi semelhante entre casos da com forma pulmonar (71,1%) ou extrapulmonar (71,8%) da doença. A forma mista esteve mais presente em casos que abandonaram (14,7%), foram a óbito (4,56%) ou apresentaram resistência ao tratamento (0,63%). Um dado que chamou atenção foi em relação ao exame de HIV. Identificou-se que 55% das pessoas não realizaram tal exame, tendo alta prevalência de abandono do tratamento (13,83%) e, mais importante, de resistência ao tratamento (0,32%). Os demais percentuais podem ser observados na Tabela 3 e 4.

Construa os dados

Nesta etapa foram criadas variáveis a partir de variáveis originais, a fim de melhorar o processo de análise de dados. O quadro 2, a seguir, mostra as variáveis criadas e de onde elas foram originadas.

Quadro 2 – Descrição das variáveis criadas para o estudo

Variável criada	Variáveis originais e/ou expressões de criação
Idade	A variável idade foi gerada uma vez que o banco de dados do SINAN a mostra em códigos. Por exemplo, para a idade em anos, é acrescentado o código 40 ao seu valor. A nova variável foi criada respeitando as codificações existentes e para aquelas que lidavam com dias e meses os dados foram transformados em zero, pois todas as idades estavam em anos completos.
Tempo = 1	Foi criada a variável tempo com valor 1 para todas as observações. Esse método é aplicado a fim de uniformizar o tempo e os resultados das regressões de Poisson e binomial negativa apresentem Incidência de Risco Relativo (IRR) interpretáveis como Razão de Prevalência (RP).
Faixa etária (1 - 12)	A partir da variável idade foram criadas 12 faixas etárias de acordo com classificação do IBGE.
Raça (1 - 6)	Criada uma variável dummy para cada uma das raças existentes na ficha de notificação
Escolaridade (1 - 11)	Criada uma variável dummy para cada uma das escolaridades existentes na ficha de notificação
Zona (1 - 4)	Criada uma variável dummy para cada uma das zonas de moradia existentes na ficha de notificação
Tipo de entrada (1 - 6)	Criada uma variável dummy para cada um dos tipos de entrada existentes na ficha de notificação
Forma (1 - 3)	Criada uma variável dummy para cada uma das formas de entrada existentes na ficha de notificação
Exame raio-x (1 - 4)	Criada uma variável dummy para cada um dos resultados do exame de raio-x
Exame de PPD (1 - 4)	Criada uma variável dummy para cada um dos resultados do exame de PPD
Exame de baciloscopia 1 (1 - 4)	Criada uma variável dummy para cada um dos resultados da primeira baciloscopia de escarro

Exame de baciloscopia 2 (1 - 3)	Criada uma variável dummy para cada um dos resultados da segunda baciloscopia de escarro
Exame de cultura de escarro (1 - 4)	Criada uma variável dummy para cada um dos resultados cultura de escarro
Exame HIV (1 - 4)	Criada uma variável dummy para cada um dos resultados do exame de HIV
Exame histopatológico (1 - 6)	Criada uma variável dummy para cada um dos resultados do exame histopatológico
Tratamento diretamente observado (1 - 3)	Criada uma variável dummy para o acompanhamento do tratamento do paciente
Tipo de entrada (1 - 6)	Criada uma variável dummy para cada um dos tipos de entrada existentes na ficha de notificação
Situação de encerramento (1 - 10)	Criada uma variável dummy para cada um dos tipos de encerramento existentes na ficha de acompanhamento
Encerramento (1-5)	Criada variável para os cinco encerramentos a serem analisados pela rede neural artificial no software WEKA (cura, abandono, óbito, resistência, outros)

Fonte: Próprio autor

Integre os dados

Esta etapa pede para fazer a integração de diversas tabelas ou de dados para criar novas tabelas ou dados. Entretanto, não foi necessário utilizar essa integração, uma vez que a Secretaria de Saúde forneceu uma só tabela com todos os dados dos pacientes notificados com TB.

Formate os dados

De acordo com o manual do CRISP-DM, essa fase refere-se a transformações sintáticas preliminares sem a mudança de significado. Tal procedimento também se dá devido à alguma ordem de alocação dos dados, sendo necessária sua randomização. Devido não haver ordem de seleção ou alocação dos pacientes no SINAN-TB, também não houve necessidade de seguir essa etapa.

Modelagem

Selecione a técnica de modelagem

As técnicas de modelagem das séries temporais, fatores associados aos desfechos da tuberculose e a rede neural está descrita no tópico 5.5 análise de dados, dentro da sessão método.

Gere os desenhos dos testes

Para as análises de séries temporais, estudos de fatores associados a cada um dos encerramentos estudados e modelagem de redes neurais foram selecionadas técnicas que podem ser observadas no tópico 5.5 análise de dados, dentro da sessão método.

Crie o modelo

Para a criação do modelo final, diversos pontos foram considerados a fim de entender o comportamento dos dados em relação aos encerramentos. Portanto, as seções a seguir demonstram como cada um dos desfechos foi tratado, seu comportamento ao longo do tempo, que variáveis se associaram a eles e como os modelos preditivos foram selecionados e criados.

Parte 1: Análise de Séries Temporais

A análise de séries temporal considerou diferentes indicadores epidemiológicos da TB, como incidência, prevalência da doença, taxa de cura, taxa de abandono, letalidade, mortalidade, taxa de resistência medicamentosa do ano de 2001 a 2017 (Tabela 5). Por meio da regressão por pontos de inflexão, identificou-se em que ano essas mudanças ocorreram e com qual intensidade.

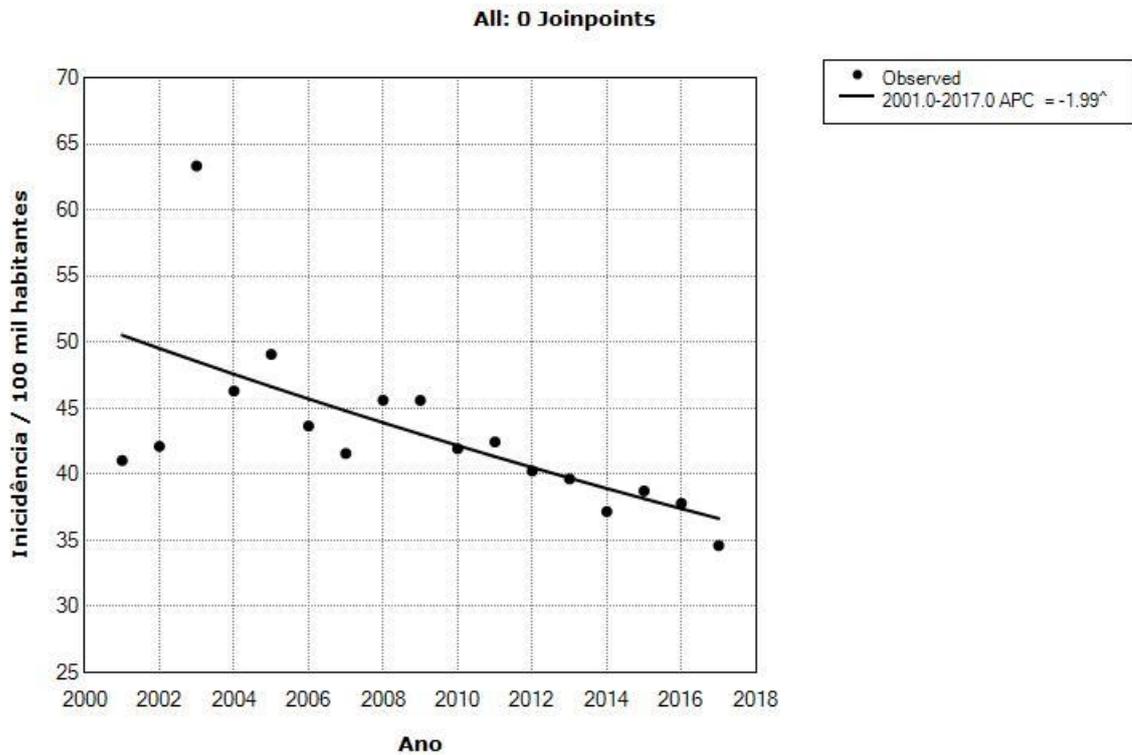
Tabela 5 – Indicadores Epidemiológicos dos desfechos da tuberculose no estado do Ceará, 2001-2017. Fortaleza-CE, 2018.

Ano	Inc. /100 mil hab	Prev. /100 mil hab	Cura %	Aband. %	Letal. %	Mortal. /100 mil hab	TB-DR %
2001	41,06	49,39	78,57	11,94	0,47	0,23	0,18
2002	42,13	51,66	74,04	15,90	0,42	0,22	0,27
2003	63,36	72,92	72,59	14,75	0,45	0,33	0,36
2004	46,32	53,49	76,48	12,41	0,21	0,11	0,28
2005	49,09	56,75	76,49	9,09	0,45	0,26	0,41
2006	43,67	51,30	76,57	8,95	0,83	0,42	0,54
2007	41,59	49,56	76,98	9,50	3,00	1,49	0,51
2008	45,62	53,86	74,55	9,95	3,22	1,74	0,22
2009	45,61	55,27	71,01	10,84	3,75	2,07	0,40
2010	41,95	50,67	71,99	11,22	2,90	1,47	0,67
2011	42,46	51,18	69,84	12,25	2,13	1,09	0,57
2012	40,26	49,54	68,44	12,63	3,01	1,49	0,76
2013	39,66	48,91	66,60	13,67	2,70	1,32	0,63
2014	37,19	45,68	63,29	16,44	2,97	1,36	0,84
2015	38,75	47,87	63,52	14,38	2,25	1,08	0,84
2016	37,82	46,72	64,59	14,97	2,77	1,29	0,43
2017	34,62	44,25	61,20	14,50	3,58	1,59	1,08

Fonte: SINAN (TB), IBGE

A incidência da TB foi calculada segundo recomendação do DATASUS, considerando novo o caso classificado com tipo de entrada “caso novo”, “não sabe” e “pós óbito”. Ele foi calculado utilizando dados da população do estado do Ceará disponibilizado pelo website do IBGE. Em 2001, a doença representava um coeficiente de incidência de 41,06 novos casos para cada 100 mil habitantes no estado e 34,62 novos casos para 100 mil habitantes em 2017. Esse decréscimo representou uma AAPC = -2,0 (IC95% = -3,1 – -0,8; $p < 0,001$) como pode ser observado na Tabela 6 e Figura 8).

Figura 8 – Padrão temporal da incidência da tuberculose

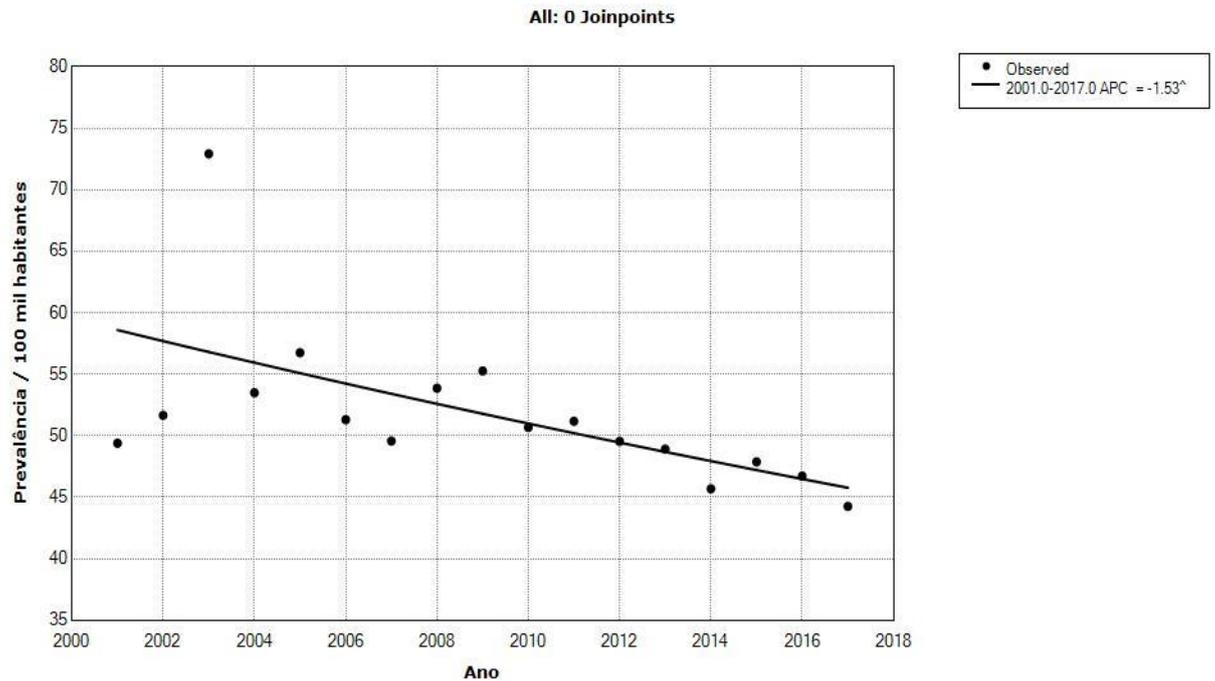


^ Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 0 Joinpoints.

Fonte: SINAN-TB (CE)

Para a prevalência da doença todos os casos foram considerados no cálculo assim como as populações do estado. Identificou-se decréscimo semelhante ao da incidência, onde em 2001 haviam 49,39 casos para 100 mil habitantes e em 2017 haviam 44,25 casos para 100 mil habitantes. Tal decréscimo representou AAPC = -1,5% ao ano (IC95%: -2,5 – -0,5; $p > 0,001$), como mostram a Figura 9 e Tabela 6.

Figura 9 – Padrão temporal da prevalência da tuberculose

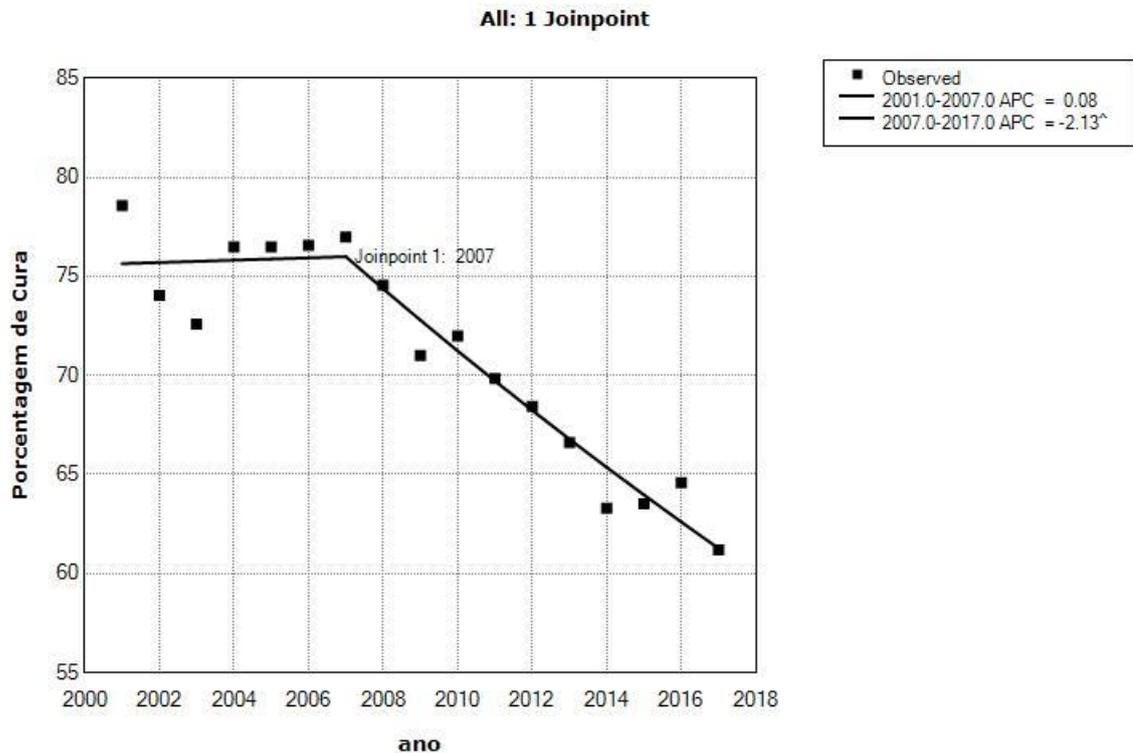


[^] Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 0 Joinpoints.

Fonte: SINAN-TB (CE)

A cura da doença apresentou uma mudança significativa decrescente ao longo dos anos. Em 2001 observou-se que 78,57% dos casos curaram e que essa medida permaneceu constante durante seis anos, pois a partir de 2007 houve um decréscimo nesse valor de 2,1% ao ano, chegando ao valor de apenas 61,20% dos casos evoluírem para cura no estado do Ceará ($p < 0,001$). Considerando o período como um todo, o modelo demonstrou uma AAPC de -1,3% ao ano (IC95%: -1,8 – -0,08; $p < 0,001$), visto na Figura 10 e Tabela 6).

Figura 10 – Padrão temporal da taxa de cura da tuberculose

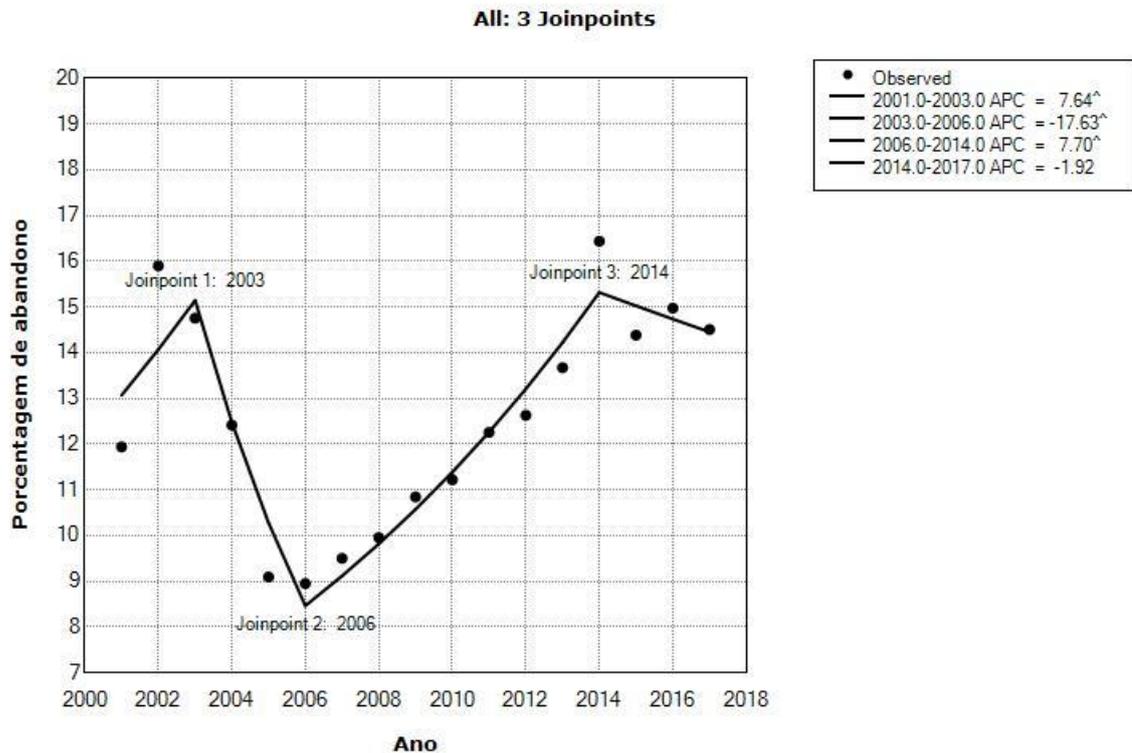


[^] Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
 Final Selected Model: 1 Joinpoint.

Fonte: Próprio autor

O abandono da doença foi o desfecho que mostrou mais variação ao longo da série temporal. Foi possível identificar acréscimo de 7,64% por ano ($p < 0,001$) no período de 2001-2003. Entre 2003 a 2006 esse valor reduziu 17,6% ao ano ($p < 0,001$), já entre 2006 e 2014 houve aumento na porcentagem de casos com abandono de tratamento (7,7%, $p < 0,001$). Por fim, o período de 2014 a 2017 apresentou novamente decréscimo na porcentagem de abandono dos casos (-1,9%), porém não significativo ($p = 0,20$). Ao observar toda a série, houve 11,94% de abandono do tratamento no ano de 2001 e 14,50% em 2017. A AAPC da série mostrou acréscimo de 0,6% ao ano (IC95%:-0,7 – -2,0; $p = 0,30$), dado observado na Figura 11 e Tabela 6.

Figura 11 – Padrão temporal da taxa de abandono da tuberculose

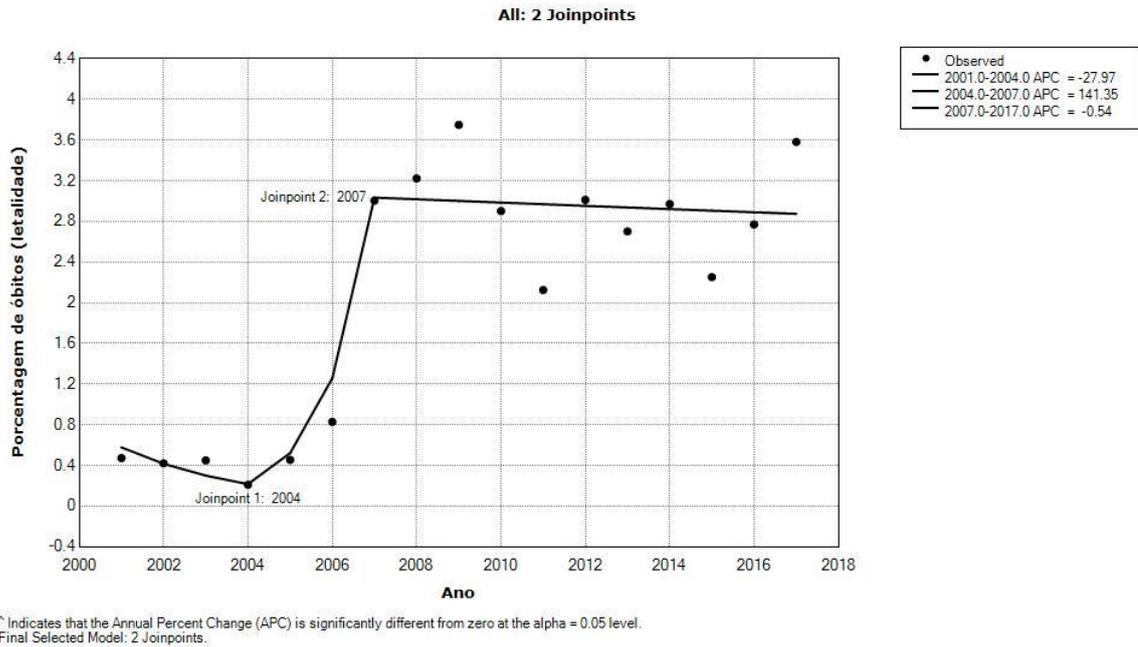


^ Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 3 Joinpoints.

Fonte: SINAN-TB (CE)

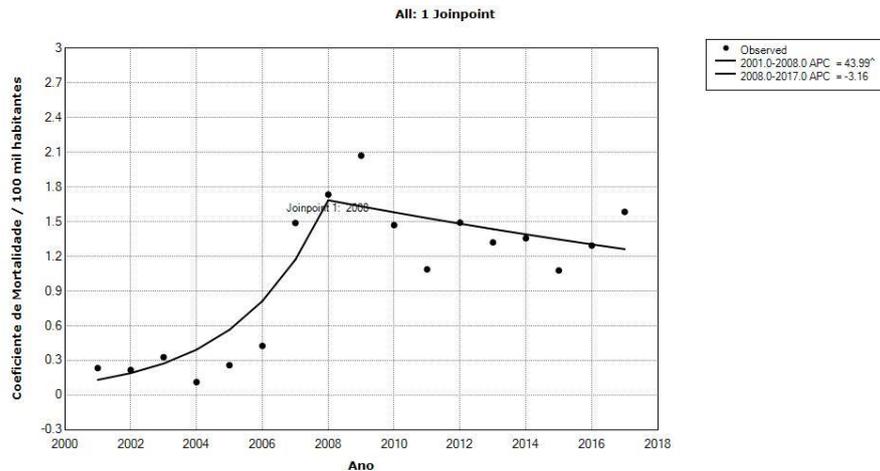
Em relação a letalidade, a série mostrou que em 2001 0,47% dos casos evoluíram para o óbito, já em 2017 esse valor subiu para 3,58% e pela regressão por joinpoints identificou-se uma série com três segmentos. O primeiro período (de 2001 a 2004) apresentou um declínio de 28% ao ano na letalidade ($p = 0,30$), o segundo período (de 2004 a 2007) apresentou um crescimento de 141,3% da letalidade da doença ao ano ($p = 0,10$). O último segmento apresentou-se graficamente de forma estacionária com $APC = -0,5$ ($p = 0,80$). Assim como cada segmento não apresentou significância estatística, a série como um todo também se comportou da mesma forma com $AAPC = 10,6$ ($IC95\%: -12,0 - 39,0$; $p = 0,40$), mas com uma elevação visualmente perceptível (Figura 12, Tabela 6).

Figura 12 – Padrão temporal da taxa de óbito (letalidade) da tuberculose



A mortalidade por TB foi o terceiro indicador que precisou ser calculado. Considerando a população do estado do Ceará, esse dado foi disponibilizado pelo website do IBGE e em cada ano essa informação foi acrescida. Notou-se que no período de 2001 a 2008 houve crescimento de 44% no número de casos por 100 mil habitantes ($p < 0,001$), já entre 2008 e 2017 houve tendência de decréscimo em 3,2%, mas sem significância estatística ($p = 0,30$). Durante o período de 2001 a 2017 observou-se que a doença representava 0,23 óbitos para 100 mil habitantes, e ao fim, letalidade de 1,59 óbitos para 100 mil habitantes, significando crescimento da tendência de mortalidade com AAPC = 15,2% (IC95% = 6,3 – 24,8; $p < 0,001$), como visto na Figura 13 e Tabela 6.

Figura 13 – Padrão temporal do coeficiente de mortalidade da tuberculose

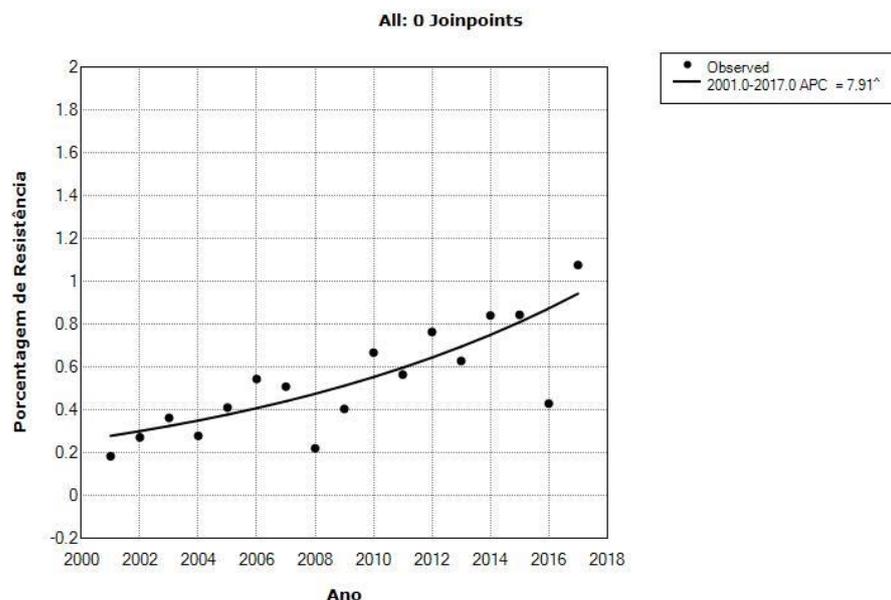


[^] Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 1 Joinpoint.

Fonte: próprio autor

Por fim, a resistência ao tratamento da TB mostrou tendência quase linear durante toda a série estudada, pois não houve necessidade de serem acrescentados pontos de inflexão para o período. Dessa forma, em 2001, observou-se que 0,18% dos casos adquiriram resistência medicamentosa, já em 2017 houve 1,08% de resistência de todos os casos. Isso representou AAPC = 7,9% (IC95% = 4,7 -11,2; $p < 0,01$) ao ano (Figura 14, Tabela 4).

Figura 14 – Padrão temporal do abandono do tratamento da tuberculose



[^] Indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level.
Final Selected Model: 0 Joinpoints.

Fonte: próprio autor

Tabela 6 – Padrão temporal dos indicadores epidemiológicos da tuberculose no estado do Ceará, 2001-2017. Fortaleza (CE), 2018.

	Período 1		Período 2		Período 3		Período 4	Período completo
	APC (IC95%)	Joinpoint	APC (IC95%)	Joinpoint	APC (IC95%)	Joinpoint	APC (IC95%)	AAPC (IC 95%)
Incidência	-2,0 (-3,1 – -0,8)*							-2,0 (-3,1 – -0,8)*
Prevalência	-1,5 (-2,5 – -0,5)*							-1,5 (-2,5 – -0,5)*
Cura	0,1 (-1,1 – 1,2)	2007	-2,1 (-2,8 – -1,5)*					-1,3 (-1,8 – -0,8)*
Abandono	7,6 (1,6 – 14,5)*	2003	-17,6 (-23,0 – -11,9)*	2006	7,7 (6,8 – 8,6)*	2014	-1,9 (-4,7 – 1,0)	0,6 (-0,7 – 2,0)
Letalidade	-28,0 (-63,4 – 41,9)	2004	141,3 (-29,2 – 722,2)	2007	-0,5 (-4,6 – 3,6)			10,6 (-12,0 – 39,0)
Mortalidade	44,0 (19,6 – 73,3)*	2008	-3,2 (-9,3 – 3,4)					15,2 (6,3 – 24,8)
Resistência	7,9 (4,7 – 11,2)*							7,9 (4,7 – 11,2)*

* p<0,05 Fonte: SINAN-TB (CE); IBGE, 2018

Parte 2: Análise bivariada

Para decidir quais variáveis utilizar nos modelos preditivos, utilizou-se uma série de análises bivariadas. Decidiu-se optar, inicialmente, pelo p-valor como uma medida que expressa relações entre as variáveis preditoras e desfecho; entretanto, sabe-se da importância de avaliar a variável dentro do seu contexto epidemiológico e da necessidade do uso de medidas mais fiéis como as razões de prevalência com seus respectivos intervalos de confiança. Tal atitude foi tomada nos modelos multivariáveis, posteriormente demonstrados.

Para o desfecho cura do tratamento da tuberculose, seis variáveis não foram utilizadas devido $p \leq 0,20$ sendo elas: raça amarela, escolaridade 1-4 série do ensino fundamental incompleta, zona de moradia periurbana, baciloscopia de outro material positiva, cultura de outro material positiva e exame de HIV em andamento (Tabela 7). Todas as demais variáveis permaneceram como possíveis preditoras da cura da tuberculose a serem incluídas nos três modelos multivariáveis.

Tabela 7 – Análise bivariada (p-valores) das variáveis preditoras com o desfecho cura do tratamento da tuberculose. Fortaleza (CE), 2018.

Variáveis	Bivariado Logístico	Bivariado Poisson	Bivariado Binomial Negativo
Idade	<0,001	<0,001	<0,001
Sexo			
Feminino	1	1	1
Masculino	<0,001	<0,001	<0,001
Raça			
Branca	1	1	1
Preta	0,004	0,003	0,003
Amarela*	0,957	0,957	0,957
Parda	<0,001	<0,001	<0,001
Indígena	0,106	0,078	0,078
Escolaridade			
Analfabeto	1	1	1
1-4 fundamental incomp.*	0,973	0,973	0,973
1-4 fundamental comp.	0,011	0,013	0,013
5-8 fundamental incomp.	<0,001	<0,001	<0,001

5-8 fundamental comp.	0,034	0,029	0,029
Ens médio incomp.	<0,001	<0,001	<0,001
Ens médio comp.	<0,001	<0,001	<0,001
Ens superior incomp.	<0,001	<0,001	<0,001
Ens superior comp.	<0,001	<0,001	<0,001
Zona de Moradia			
Urbana	1	1	1
Rural	<0,001	<0,001	<0,001
Periurbana*	0,643	0,648	0,648
Tipo de entrada			
Caso novo	1	1	1
Recidiva	<0,001	<0,001	<0,001
Reingresso pós-abandono	<0,001	<0,001	<0,001
Não sabe	<0,001	<0,001	<0,001
Transferência	<0,001	<0,001	<0,001
Forma da doença			
Pulmonar	1	1	1
Extra-pulmonar	0,087	0,083	0,083
Mista	<0,001	<0,001	<0,001
Raio-x			
Suspeito	0,022	0,023	0,023
Normal	1	1	1
Outra patologia	0,003	0,006	0,006
Não realizado	<0,001	<0,001	<0,001
Teste tuberculínico			
Suspeito	<0,001	0,001	0,001
Normal	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de escarro			
Positiva	<0,001	<0,001	<0,001
Negativa	1	1	1
Não se aplica	<0,001	<0,001	<0,001
Baciloscopia de outro			

material			
Positiva*	0,937	0,937	0,937
Negativa	1	1	1
Não realizada	0,041	0,046	0,046
Cultura de escarro			
Positiva	0,139	0,143	0,143
Negativa	1	1	1
Não realizada	<0,001	<0,001	<0,001
Cultura de outro material			
Positivo*	0,298	0,286	0,286
Negativo	1	1	1
Não realizada	0,025	0,033	0,033
HIV			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Em andamento*	0,767	0,766	0,766
Não realizado	<0,001	<0,001	<0,001
Histopatológico			
BAAR positivo	<0,001	<0,001	<0,001
Sugestivo de TB	<0,001	<0,001	<0,001
Não sugestivo de TB	1	1	1
Em andamento	0,019	0,023	0,023
Não realizado	<0,001	<0,001	<0,001
Tratamento			
Rifampicina	<0,001	<0,001	<0,001
Isoniazida	<0,001	<0,001	<0,001
Pirazinamida	<0,001	<0,001	<0,001
Etambutol	<0,001	<0,001	<0,001
Estreptomicina	<0,001	<0,001	<0,001
Etionamida	<0,001	<0,001	<0,001
DOTS			
Sim	0,006	0,006	0,006
Não	1	1	1

Ignorado	<0,001	<0,001	<0,001
Baciloscopia de 2º mês			
Positivo	<0,001	0,001	0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 4º mês			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 6º mês			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001

* Variável eliminada Fonte: SINAN-TB (CE)

Em relação ao abandono do tratamento da TB, sete variáveis foram excluídas: raça amarela, 5-8 série fundamental completa, teste tuberculínico suspeito, baciloscopia de escarro positiva, rifampicina, isoniazida e baciloscopia de 2º mês positiva. As demais foram incluídas nos modelos multivariáveis (Tabela 8).

Tabela 8 – Análise bivariada (p-valores) das variáveis preditoras com o desfecho abandono do tratamento da tuberculose. Fortaleza (CE), 2018.

Variáveis	Bivariado Logístico	Bivariado Poisson	Bivariado Binomial Negativo
Idade	<0,001	<0,001	<0,001
Sexo			
Feminino	1	1	1
Masculino	<0,001	<0,001	<0,001
Raça			
Branca	1	1	1
Preta	0,120	0,119	0,119
Amarela*	0,618	0,619	0,619
Parda	0,002	0,002	0,002
Indígena	0,053	0,058	0,058

Escolaridade

Analfabeto	1	1	1
1-4 fundamental incomp.	0,005	0,006	0,006
1-4 fundamental comp.	0,197	0,195	0,195
5-8 fundamental incomp.	0,001	0,001	0,001
5-8 fundamental comp.*	0,528	0,529	0,529
Ens médio incomp.	<0,001	<0,001	<0,001
Ens médio comp.	<0,001	<0,001	<0,001
Ens superior incomp.	0,001	0,001	0,001
Ens superior comp.	<0,001	<0,001	<0,001

Zona de Moradia

Urbana	1	1	1
Rural	<0,001	<0,001	<0,001
Periurbana	<0,001	<0,001	<0,001

Tipo de entrada

Caso novo	1	1	1
Recidiva	<0,001	<0,001	<0,001
Reingresso pós-abandono	<0,001	<0,001	<0,001
Não sabe	<0,001	<0,001	<0,001
Transferência	<0,001	<0,001	<0,001

Forma da doença

Pulmonar	1	1	1
Extra-pulmonar	<0,001	<0,001	<0,001
Mista	0,019	0,018	0,018

Raio-x

Suspeito	<0,001	<0,001	<0,001
Normal	1	1	1
Outra patologia	<0,001	0,001	0,001
Não realizado	<0,001	<0,001	<0,001

Teste tuberculínico

Suspeito*	0,226	0,228	0,228
Normal	1	1	1
Não realizado	<0,001	<0,001	<0,001

Baciloscopia de escarro			
Positiva*	0,256	0,256	0,256
Negativa	1	1	1
Não se aplica	<0,001	<0,001	<0,001
Baciloscopia de outro material			
Positiva	0,017	0,018	0,018
Negativa	1	1	1
Não realizada	0,011	0,012	0,012
Cultura de escarro			
Positiva	0,155	0,154	0,154
Negativa	1	1	1
Não realizada	0,001	0,001	0,001
Cultura de outro material			
Positivo	0,155	0,158	0,158
Negativo	1	1	1
Não realizada	0,380	0,382	0,382
HIV			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Em andamento	<0,001	<0,001	<0,001
Não realizado	<0,001	<0,001	<0,001
Histopatológico			
BAAR positivo	<0,001	<0,001	<0,001
Sugestivo de TB	<0,001	<0,001	<0,001
Não sugestivo de TB	1	1	1
Em andamento	0,155	0,117	0,117
Não realizado	<0,001	<0,001	<0,001
Tratamento			
Rifampicina*	0,302	0,299	0,299
Isoniazida*	0,796	0,796	0,796
Pirazinamida	0,004	0,004	0,004
Etambutol	<0,001	<0,001	<0,001

Estreptomina	0,146	0,143	0,143
Etionamida	0,177	0,174	0,174
DOTS			
Sim	0,034	0,034	0,034
Não	1	1	1
Ignorado	0,017	0,017	0,017
Baciloscopia de 2º mês			
Positivo*	0,650	0,651	0,651
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 4º mês			
Positivo	0,001	0,001	0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 6º mês			
Positivo	0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001

* Variável eliminada Fonte: SINAN-TB (CE)

Já para a variável óbito por TB foram excluídas 13 variáveis, sendo elas: raça amarela, raça indígena, zona de moradia periurbana, tipo de entrada desconhecido, teste tuberculínico suspeito, baciloscopia de outro material positiva, cultura de outro material (positiva e não realizada), histopatológico com BAAR positivo, histopatológico não realizado, tratamento supervisionado realizado, baciloscopia de 2º mês positiva e baciloscopia de 4º mês positiva (Tabela 9). As 43 variáveis restantes foram incluídas nos três modelos multivariáveis.

Tabela 9 – Análise bivariada (p-valores) das variáveis preditoras com o desfecho óbito por tuberculose. Fortaleza (CE), 2018.

Variáveis	Bivariado Logístico	Bivariado Poisson	Bivariado Binomial Negativo
Idade	<0,001	<0,001	<0,001
Sexo			
Feminino	1	1	1
Masculino	<0,001	<0,001	<0,001
Raça			
Branca	1	1	1
Preta	0,120	0,121	0,121
Amarela*	0,263	0,265	0,265
Parda	<0,001	<0,001	<0,001
Indígena*	0,922	0,922	0,922
Escolaridade			
Analfabeto	1	1	1
1-4 fundamental incomp.	0,001	0,001	0,001
1-4 fundamental comp.	<0,001	<0,001	<0,001
5-8 fundamental incomp.	<0,001	<0,001	<0,001
5-8 fundamental comp.	0,056	0,056	0,056
Ens médio incomp.	<0,001	<0,001	<0,001
Ens médio comp.	<0,001	<0,001	<0,001
Ens superior incomp.	0,099	0,100	0,100
Ens superior comp.	0,010	0,010	0,010
Zona de Moradia			
Urbana	1	1	1
Rural	0,012	0,012	0,012
Periurbana*	0,237	0,239	0,239
Tipo de entrada			
Caso novo	1	1	1
Recidiva	0,004	0,004	0,004
Reingresso pós-abandono	<0,001	<0,001	<0,001
Não sabe*	0,429	0,430	0,430

Transferência	0,006	0,006	0,006
Forma da doença			
Pulmonar	1	1	1
Extra-pulmonar	0,004	0,004	0,004
Mista	<0,001	<0,001	<0,001
Raio-x			
Suspeito	0,001	0,001	0,001
Normal	1	1	1
Outra patologia	<0,001	<0,001	<0,001
Não realizado	0,044	0,045	0,045
Teste tuberculínico			
Suspeito*	0,495	0,495	0,495
Normal	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de escarro			
Positiva	<0,001	<0,001	<0,001
Negativa	1	1	1
Não se aplica	<0,001	<0,001	<0,001
Baciloscopia de outro material			
Positiva*	0,278	0,278	0,278
Negativa	1	1	1
Não realizada	0,089	0,088	0,088
Cultura de escarro			
Positiva	0,055	0,056	0,056
Negativa	1	1	1
Não realizada	0,038	0,038	0,038
Cultura de outro material			
Positivo*	0,958	0,958	0,958
Negativo	1	1	1
Não realizada*	0,779	0,779	0,779
HIV			
Positivo	<0,001	<0,001	<0,001

Negativo	1	1	1
Em andamento	0,016	0,016	0,016
Não realizado	0,015	0,015	0,015
Histopatológico			
BAAR positivo*	0,672	0,672	0,672
Sugestivo de TB	0,001	0,001	0,001
Não sugestivo de TB	1	1	1
Em andamento	0,195	0,194	0,194
Não realizado*	0,483	0,483	0,483
Tratamento			
Rifampicina	<0,001	<0,001	<0,001
Isoniazida	<0,001	<0,001	<0,001
Pirazinamida	<0,001	<0,001	<0,001
Etambutol	<0,001	<0,001	<0,001
Estreptomicina	<0,001	<0,001	<0,001
Etionamida	<0,001	<0,001	<0,001
DOTS			
Sim*	0,215	0,215	0,215
Não	1	1	1
Ignorado	<0,001	<0,001	<0,001
Baciloscopia de 2º mês			
Positivo*	0,408	0,408	0,408
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 4º mês			
Positivo*	0,828	0,828	0,828
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001
Baciloscopia de 6º mês			
Positivo	0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado	<0,001	<0,001	<0,001

* Variável eliminada Fonte: SINAN-TB (CE)

Finalmente, para o desfecho resistência ao tratamento da TB 28 variáveis apresentaram $p \leq 0,20$ e, conseqüentemente foram retiradas. São elas: idade, sexo, raça amarela, raça parda, 1-4 série do fundamental incompleta, 5-8 série fundamental incompleta, 5-8 série fundamental completa, ensino médio incompleto, ensino médio completo, ensino superior completo, zona de moradia periurbana, tipo de entrada desconhecida, tuberculose forma mista, raio-x (suspeito, outra patologia, não realizado), baciloscopia de outro material (positiva, não realizada), cultura de escarro não realizada, cultura de outro material (positivo, não realizado), HIV positivo, HIV em andamento, histopatológico com BAAR positivo, histopatológico em andamento, baciloscopia de 2º mês não realizada, baciloscopia de 4º mês não realizada, baciloscopia de 6º mês não realizada (Tabela 10). Para o modelo multivariável desse desfecho restaram 28 variáveis preditoras.

Tabela 10 – Análise bivariada (p-valores) das variáveis preditoras com o desfecho resistência ao tratamento da tuberculose. Fortaleza (CE), 2018.

Variáveis	Bivariado Logístico	Bivariado Poisson	Bivariado Binomial Negativo
Idade*	0,288	0,178	0,178
Sexo			
Feminino	1	1	1
Masculino*	0,964	0,964	0,964
Raça			
Branca	1	1	1
Preta	<0,001	<0,001	<0,001
Amarela*	0,584	0,584	0,584
Parda*	0,493	0,493	0,493
Indígena	0,200	0,199	0,199
Escolaridade			
Analfabeto	1	1	1
1-4 fundamental* incomp.	0,314	0,314	0,314
1-4 fundamental comp.	0,001	0,001	0,001
5-8 fundamental incomp.*	0,986	0,986	0,986
5-8 fundamental comp.*	0,735	0,735	0,735
Ens médio incomp.*	0,343	0,343	0,343
Ens médio comp.*	0,820	0,820	0,820

Ens superior incomp.	0,193	0,192	0,192
Ens superior comp.*	0,312	0,312	0,312
Zona de Moradia			
Urbana	1	1	1
Rural	0,146	0,146	0,146
Periurbana*	0,791	0,791	0,791
Tipo de entrada			
Caso novo	1	1	1
Recidiva	<0,001	<0,001	<0,001
Reingresso pós-abandono	<0,001	<0,001	<0,001
Não sabe*	0,781	0,781	0,781
Transferência	0,001	0,001	0,001
Forma da doença			
Pulmonar	1	1	1
Extra-pulmonar	<0,001	<0,001	<0,001
Mista*	0,603	0,603	0,603
Raio-x			
Suspeito*	0,793	0,793	0,794
Normal	1	1	1
Outra patologia*	0,448	0,449	0,449
Não realizado*	0,545	0,545	0,545
Teste tuberculínico			
Suspeito	0,010	0,010	0,010
Normal	1	1	1
Não realizado	0,004	0,004	0,004
Baciloscopia de escarro			
Positiva	<0,001	<0,001	<0,001
Negativa	1	1	1
Não se aplica	<0,001	<0,001	<0,001
Baciloscopia de outro material			
Positiva*	0,623	0,623	0,623
Negativa	1	1	1

Não realizada*	0,237	0,238	0,238
Cultura de escarro			
Positiva	<0,001	<0,001	<0,001
Negativa	1	1	1
Não realizada*	0,385	0,385	0,385
Cultura de outro material			
Positivo*	0,892	0,892	0,892
Negativo	1	1	1
Não realizada*	0,399	0,398	0,398
HIV			
Positivo*	0,950	0,950	0,950
Negativo	1	1	1
Em andamento*	0,704	0,704	0,704
Não realizado	<0,001	<0,001	<0,001
Histopatológico			
BAAR positivo*	0,287	0,287	0,287
Sugestivo de TB	0,001	0,002	0,002
Não sugestivo de TB	1	1	1
Em andamento*	0,342	0,342	0,342
Não realizado	0,001	0,001	0,001
Tratamento			
Rifampicina	<0,001	<0,001	<0,001
Isoniazida	<0,001	<0,001	<0,001
Pirazinamida	0,106	0,106	0,106
Etambutol	<0,001	<0,001	<0,001
Estreptomicina	<0,001	<0,001	<0,001
Etionamida	<0,001	<0,001	<0,001
DOTS			
Sim	0,013	0,013	0,013
Não	1	1	1
Ignorado	<0,001	<0,001	<0,001
Baciloscopia de 2 ^o mês			
Positivo	<0,001	<0,001	<0,001

Negativo	1	1	1
Não realizado*	0,433	0,433	0,433
Baciloscopia de 4 ^o mês			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado*	0,384	0,384	0,384
Baciloscopia de 6 ^o mês			
Positivo	<0,001	<0,001	<0,001
Negativo	1	1	1
Não realizado*	0,353	0,353	0,353

* Variável eliminada Fonte: SINAN-TB (CE)

Parte 3: Regressões

A partir dos dados anteriores foram criados modelos multivariáveis para os quatro desfechos por três métodos distintos: logístico, Poisson robusto e binomial negativo robusto. Todos os modelos foram criados a partir de modelos cheios, considerando todas as variáveis que apresentaram $p \leq 0,20$ dos resultados bivariados; assim, uma a uma foram eliminadas as variáveis que não apresentavam significância estatística. Dessa forma, os modelos finais foram ajustados para todas as variáveis estatisticamente significantes apresentadas a seguir.

Para a cura, observou-se que, pelo modelo logístico, a prevalência do desfecho foi aumentada pela escolaridade de ensino médio incompleto (RP = 1,24), ensino médio completo (RP = 1,48), ensino superior incompleto (RP = 2,24) e ensino superior incompleto (RP = 1,66), a tuberculose apresentar-se de forma extrapulmonar (RP = 1,28), exame histopatológico não sugestivo de TB (RP = 1,69), uso de medicamentos como a rifampicina (RP = 1,83), isoniazida (RP = 1,93) e a estreptomicina (RP = 1,59) (Tabela 11).

A prevalência da cura do tratamento mostrou-se diminuída quando associada ao avançar da idade (RP = 0,996), ao sexo masculino (RP = 0,80), raça parda (RP = 0,87), entrada como recidiva (RP = 0,80), reingresso pós abandono (RP = 0,32) ou transferência (RP = 0,77), teste tuberculínico suspeito (RP = 0,61) ou não realizado (RP = 0,56), exame de HIV positivo (RP = 0,28) ou não realizado (RP =

0,67), realizar tratamento supervisionado (RP = 0,92), fazer uso de etambutol (RP = 0,87). Em relação às baciloscopias bimestrais, todas se mostraram com diminuição da prevalência de cura, sendo elas a de 2º mês positiva (RP = 0,66) e não realizada (RP = 0,82), a de 4º mês positiva (RP = 0,37) e não realizada (RP = 0,46), e a de 6º mês positiva (RP = 0,04) e não realizada (RP = 0,09) (Tabela 11).

Os modelos de Poisson robusto e binomial negativo robusto apresentaram resultados semelhantes e com razões de prevalência menos expressivas, assim como intervalos de confiança mais estreitos; ademais, as variáveis do exame de radiografia entraram para o modelo final. Assim, em ambos os modelos, a prevalência de cura foi aumentada pela escolaridade de ensino médio incompleto (RP = 1,03), ensino médio completo (RP = 1,05), ensino superior incompleto (RP = 1,10) e ensino superior completo (RP = 1,06), a TB apresentar-se na forma extrapulmonar (RP = 1,02), exame histopatológico não sugestivo de TB (RP = 1,08), uso de rifampicina (RP = 1,17), isoniazida (RP = 1,14) e estreptomicina (RP = 1,09) (Tabela 11).

A prevalência da cura do tratamento mostrou-se diminuída quando associada ao sexo masculino (RP = 0,97), raça parda (RP = 0,97), entrada como recidiva (RP = 0,96), reingresso pós abandono (RP = 0,71), não saber (RP = 0,79) ou transferência (RP = 0,95), radiografia suspeita (RP = 0,96), outra doença (RP = 0,91) ou não realizado (RP = 0,96), teste tuberculínico suspeito (RP = 0,92) ou não realizado (RP = 0,92), exame de HIV positivo (RP = 0,71) ou não realizado (RP = 0,94), realizar tratamento supervisionado (RP = 0,98), fazer uso de etambutol (RP = 0,98). Em relação as baciloscopias bimestrais, todas se mostraram com diminuição da prevalência de cura, sendo elas a de 2º mês positiva (RP = 0,95) e não realizada (RP = 0,97), a de 4º mês positiva (RP = 0,84) e não realizada (RP = 0,84), e a de 6º mês positiva (RP = 0,45) e não realizada (RP = 0,67) (Tabela 11).

Como modelos preditivo, o modelo logístico utilizou 34.621 observações e 28 variáveis apresentando pseudo $R^2 = 0,2867$. Já os modelos de Poisson e binomial negativo robustos utilizaram 34.508 observações de 31 variáveis com pseudo $R^2 = 0,0306$.

Tabela 11 – Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho cura da TB. Fortaleza (CE), 2018.

Variáveis	Logístico	Poisson Robusto	Binomial Negativo Robusto
	RP (IC95%)	RP (IC95%)	RP (IC95%)
Idade	0,996 (0,994 – 0,997)	-	-
Sexo masculino	0,80 (0,75 – 0,85)	0,97 (0,96 – 0,98)	0,97 (0,96 – 0,98)
Raça Parda	0,87 (0,81 – 0,93)	0,97 (0,96 – 0,98)	0,97 (0,96 – 0,98)
Ens. médio incomp.	1,24 (1,10 – 1,40)	1,03 (1,02 – 1,05)	1,03 (1,02 – 1,05)
Ens. médio comp.	1,48 (1,28 – 1,72)	1,05 (1,03 – 1,07)	1,05 (1,03 – 1,07)
Ens. superior incomp.	2,24 (1,43 – 3,52)	1,10 (1,05 – 1,16)	1,10 (1,05 – 1,16)
Ens. superior incomp.	1,66 (1,32 – 2,09)	1,06 (1,04 – 1,09)	1,06 (1,04 – 1,09)
Zona rural	1,25 (1,13 – 1,38)	1,02 (1,01 – 1,04)	1,02 (1,01 – 1,04)
Entrada recidiva	0,80 (0,71 – 0,90)	0,96 (0,94 – 0,98)	0,96 (0,94 – 0,98)
Entrada reingresso pós abandono	0,32 (0,28 – 0,36)	0,71 (0,68 – 0,74)	0,71 (0,68 – 0,74)
Entrada não sabe	-	0,79 (0,72 – 0,87)	0,79 (0,72 – 0,87)
Entrada transferência	0,77 (0,67 – 0,89)	0,95 (0,93 – 0,98)	0,95 (0,93 – 0,98)
Forma extrapulmonar	1,28 (1,15 – 1,42)	1,05 (1,03 – 1,08)	1,05 (1,03 – 1,08)
Radiografia suspeita	-	0,96 (0,94 – 0,99)	0,96 (0,94 – 0,99)
Radiografia outra	-	0,91	0,91

patologia		(0,86 – 0,98)	(0,86 – 0,98)
Radiografia não realizada	-	0,96 (0,94 – 0,99)	0,96 (0,94 – 0,99)
PPD suspeito	0,61 (0,52 – 0,71)	0,92 (0,89 – 0,95)	0,92 (0,89 – 0,95)
PPD não realizado	0,56 (0,51 – 0,62)	0,90 (0,89 – 0,92)	0,90 (0,89 – 0,92)
HIV positivo	0,28 (0,24 – 0,33)	0,71 (0,67 – 0,74)	0,71 (0,67 – 0,74)
HIV não realizado	0,67 (0,62 – 0,71)	0,94 (0,93 – 0,94)	0,94 (0,93 – 0,94)
Histopatológico sugestivo de TB	1,69 (1,43 – 1,99)	1,08 (1,05 – 1,11)	1,08 (1,05 – 1,11)
DOTS realizado	0,92 (0,86 – 0,98)	0,98 (0,97 – 0,99)	0,98 (0,97 – 0,99)
Rifampicina	1,83 (1,35 – 2,48)	1,17 (1,09 – 1,26)	1,17 (1,09 – 1,26)
Isoniazida	1,93 (1,44 – 2,59)	1,14 (1,07 – 1,21)	1,14 (1,07 – 1,21)
Etambutol	0,87 (0,81 – 0,93)	0,98 (0,97 – 0,99)	0,98 (0,97 – 0,99)
Estreptomomicina	1,59 (1,20 – 2,10)	1,09 (1,03 – 1,16)	1,09 (1,03 – 1,16)
Baciloscopia 2º mês positiva	0,66 (0,57 – 0,77)	0,95 (0,93 – 0,97)	0,95 (0,93 – 0,97)
Baciloscopia 2º mês não realizada	0,82 (0,75 – 0,89)	0,97 (0,95 – 0,98)	0,97 (0,95 – 0,98)
Baciloscopia 4º mês positiva	0,37 (0,28 – 0,48)	0,84 (0,79 – 0,90)	0,84 (0,79 – 0,90)
Baciloscopia 4º mês não realizada	0,46 (0,42 – 0,53)	0,89 (0,88 – 0,90)	0,89 (0,88 – 0,90)
Baciloscopia 6º mês positiva	0,04 (0,03 – 0,05)	0,45 (0,39 – 0,52)	0,45 (0,39 – 0,52)

Baciloscopia 6º mês não realizada	0,09 (0,08 – 0,10)	0,67 (0,66 – 0,68)	0,67 (0,66 – 0,68)
--------------------------------------	-----------------------	-----------------------	-----------------------

Fonte: SINAN-TB (CE)

Em relação ao abandono do tratamento, o modelo logístico identificou que os fatores de diminuição de sua prevalência são a idade (RP = 0,98), o ensino médio incompleto (RP = 0,80), ensino médio completo (RP = 0,56), ensino superior incompleto (RP = 0,56), ensino superior completo (RP = 0,56). Além desses, a zona de moradia rural (RP = 0,43), tipo de entrada transferência (RP = 0,70) e forma da doença extrapulmonar (RP = 0,72) (Tabela 12).

O abandono é mais presente em pessoas do sexo masculino (RP = 1,34), entrada como recidiva (RP = 1,42) ou reingresso pós abandono (RP = 3,63), não realizar o exame de radiografia (RP = 1,19) ou teste tuberculínico (RP = 1,45). O HIV eleva o abandono quando o exame acusa positivo (RP = 1,92) ou não é realizado (RP = 1,64) assim como não realizar o exame histopatológico (RP = 1,27). Quanto as baciloscopias bimestrais, a baciloscopia de 4º mês positiva (RP = 2,05) ou não realizada (RP = 1,79) tão quanto a de 6º mês positiva (RP = 6,97) e não realizada (RP = 10,81) aumentam a prevalência de abandono do tratamento de TB (tabela 12).

Nos modelos de Poisson e binomial negativo as mesmas variáveis acima permaneceram significantes, porém com valores reduzidos. Por exemplo, o tipo de entrada de reingresso pós abandono que, no modelo logístico, apresentou um aumento de 3,63 vezes, no modelo Poisson mostrou aumento de 2,27 vezes. A não realização da baciloscopia de 6º mês demonstrava um aumento em 10,81 vezes a prevalência do abandono no modelo logístico reduziu para 8,79 vezes no modelo de Poisson. As demais variáveis e seus valores podem ser observados na tabela 12.

Tratando como modelo preditivo, tanto a regressão logística quanto a de Poisson e binomial negativa utilizaram 20 variáveis e 40.473 observações. No entanto, o modelo logístico apresentou pseudo $R^2 = 0,2292$ e os demais $R^2 = 0,1962$.

Tabela 12 – Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho abandono do tratamento da TB. Fortaleza (CE), 2018.

Variáveis	Logístico	Poisson Robusto	Binomial Negativo Robusto
	RP (IC95%)	RP (IC95%)	RP (IC95%)
Sexo masculino	1,34 (1,24 – 1,45)	1,25 (1,18 – 1,32)	1,25 (1,18 – 1,32)
Idade	0,98 (0,98 – 0,99)	0,992 (0,990 – 0,993)	0,992 (0,990 – 0,993)
Ens. médio incomp.	0,80 (0,69 – 0,91)	0,84 (0,75 – 0,93)	0,84 (0,75 – 0,93)
Ens. médio comp.	0,56 (0,45 – 0,68)	0,61 (0,51 – 0,73)	0,61 (0,51 – 0,73)
Ens. superior incomp.	0,49 (0,27 – 0,90)	0,54 (0,31 – 0,93)	0,54 (0,31 – 0,93)
Ens. superior comp.	0,56 (0,33 – 0,63)	0,52 (0,39 – 0,69)	0,52 (0,39 – 0,69)
Zona rural	0,43 (0,38 – 0,50)	0,50 (0,44 – 0,56)	0,50 (0,44 – 0,56)
Entrada recidiva	1,42 (1,25 – 1,62)	1,31 (1,19 – 1,43)	1,31 (1,19 – 1,43)
Entrada reingresso pós abandono	3,63 (3,23 – 4,07)	2,27 (2,12 – 2,43)	2,27 (2,12 – 2,43)
Entrada transferência	0,70 (0,57 – 0,87)	0,75 (0,62 – 0,89)	0,75 (0,62 – 0,89)
Forma extrapulmonar	0,72 (0,63 – 0,81)	0,77 (0,69 – 0,85)	0,77 (0,69 – 0,85)
Radiografia não realizada	1,19 (1,09 – 1,30)	1,13 (1,06 – 1,21)	1,13 (1,06 – 1,21)
PPD não realizado	1,45 (1,31 – 1,60)	1,34 (1,23 – 1,45)	1,34 (1,23 – 1,45)
HIV positivo	1,92 (1,65 – 2,24)	1,62 (1,45 – 1,82)	1,62 (1,45 – 1,82)
HIV não realizado	1,64	1,46	1,46

	(1,51 – 1,79)	(1,37 – 1,56)	(1,37 – 1,56)
Histopatológico não realizado	1,27	1,21	1,21
	(1,12 – 1,43)	(1,09 – 1,34)	(1,09 – 1,34)
Baciloscopia 4º mês positiva	2,05	1,81	1,81
	(1,52 – 2,76)	(1,42 – 2,30)	(1,42 – 2,30)
Baciloscopia 4º mês não realizada	1,79	1,60	1,60
	(1,60 – 1,99)	(1,46 – 1,75)	(1,46 – 1,75)
Baciloscopia 6º mês positiva	6,97	6,29	6,29
	(4,91 – 9,90)	(4,63 – 8,55)	(4,63 – 8,55)
Baciloscopia 6º mês não realizada	10,81	8,79	8,79
	(9,47 – 12,35)	(7,79 – 9,92)	(7,79 – 9,92)

Fonte: SINAN-TB (CE)

Para o encerramento óbito por TB, o modelo logístico identificou que somente o exame histopatológico sugestivo de TB e fazer uso de isoniazida apareceram como fatores que diminuem a prevalência do desfecho, RP = 0,46 e RP = 0,30 respectivamente. Fatores como idade (RP = 1,03), raça parda (RP = 1,67), radiografia suspeita (RP = 3,18), outra patologia (RP = 4,93) ou radiografia não realizada (RP = 2,31), HIV positivo (RP = 2,95) aumentam a prevalência do óbito. Em relação as baciloscopias bimestrais, o desfecho mostrou-se mais prevalente quando não há a realização da de 2º mês (RP = 1,71), de 4º mês e de 6º mês (RP = 7,01), assim como, principalmente, o resultado positivo da de 6º mês (RP = 18,6) (Tabela 13).

Os modelos de Poisson e binomial negativo encontraram as mesmas variáveis em suas análises. Destacam-se as mudanças nos valores da radiografia suspeita (RP = 3,05) ou outra patologia (RP = 4,50) e a baciloscopia de 6º mês positiva (RP = 17,40) ou não realizada (RP = 6,81). As demais mudanças podem ser observadas na tabela 13. Todas as regressões utilizaram 39.123 observações de 13 variáveis. A regressão logística apresentou pseudo $R^2 = 0,1807$ e a de Poisson e binomial negativa, pseudo $R^2 = 0,1758$.

Tabela 13 – Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho óbito por TB. Fortaleza (CE), 2018.

Variáveis	Logístico	Poisson Robusto	Binomial Negativo Robusto
	RP (IC95%)	RP (IC95%)	RP (IC95%)
	1,035	1,03	1,03
Idade	(1,031 – 1,04)	(1,02 – 1,04)	(1,02 – 1,04)
	1,67	1,63	1,63
Raça parda	(1,38 – 2,03)	(1,35 – 1,96)	(1,35 – 1,96)
	3,18	3,05	3,05
Radiografia positiva	(1,62 – 6,22)	(1,55 – 6,01)	(1,55 – 6,01)
	4,93	4,50	4,50
Radiografia outra patologia	(1,82 – 13,30)	(1,75 – 11,58)	(1,75 – 11,58)
	2,31	2,25	2,25
Radiografia não realizada	(1,14 – 4,66)	(1,11 – 4,56)	(1,11 – 4,56)
	2,58	2,48	2,48
PPD não realizado	(1,90 – 3,49)	(1,84 – 3,33)	(1,84 – 3,33)
	2,95	2,76	2,76
HIV positivo	(2,25 – 3,87)	(2,15 – 3,55)	(2,15 – 3,55)
	0,46	0,47	0,47
Histopatológico sugestivo de TB	(0,25 – 0,82)	(0,26 – 0,83)	(0,26 – 0,83)
	0,30	0,34	0,34
Isoniazida	(0,22 – 0,42)	(0,25 – 0,46)	(0,25 – 0,46)
	1,71	1,66	1,66
Baciloscopia 2º mês não realizada	(1,34 – 2,18)	(1,35 – 2,04)	(1,35 – 2,04)
	2,01	1,96	1,96
Baciloscopia 4º mês não realizada	(1,46 – 2,75)	(1,54 – 2,50)	(1,54 – 2,50)
	18,6	17,40	17,40
Baciloscopia 6º mês positiva	(10,30 – 33,73)	(9,89 – 30,60)	(9,89 – 30,60)
	7,01	6,81	6,81
Baciloscopia 6º mês não realizada	(4,90 – 10,05)	(5,03 – 9,21)	(5,03 – 9,21)

Fonte: SINAN-TB (CE)

Por fim, para a resistência do tratamento da TB, o modelo logístico considerou a não realização do exame HIV (RP = 0,64) e o uso de rifampicina (RP =

0,42) como características que diminuem sua prevalência. Ela foi aumentada pelos fatores de ensino superior incompleto (RP = 4,09), entrada como recidiva (RP = 3,42), reingresso pós abandono (RP = 2,66) ou transferência (RP = 1,97), a cultura de escarro positiva (RP = 4,32), o exame histopatológico não realizado (RP = 1,87) e as baciloscopias de 4º e 6º mês positivas (RP = 1,89 e RP = 41,20, respectivamente) (Tabela 14).

Nos modelos de Poisson e binomial negativo três variáveis foram retiradas por não apresentarem significância estatística: ensino superior incompleto, exame histopatológico não realizado e baciloscopia de 4º mês positiva. Como nos demais modelos, essas regressões demonstraram RP menores que os valores da logística, com destaque à baciloscopia de 6º mês positiva, que apresentou RP = 38,24 (demais valores disponíveis na tabela 14).

No modelo logístico, 42.487 observações de dez variáveis foram utilizadas com valor de pseudo $R^2 = 0,2739$. Nos demais modelos foram utilizadas 47.303 observações de sete variáveis, obtendo-se valor de pseudo $R^2 = 0,2556$.

Tabela 14 – Modelos logístico, Poisson robusto e binomial negativo robusto do desfecho TB resistente a medicamentos. Fortaleza (CE), 2018.

Variáveis	Logístico	Poisson Robusto	Binomial Negativo Robusto
	RP (IC95%)	RP (IC95%)	RP (IC95%)
Ens. superior incompleto	4,09 (1,26 – 13,23)	-	-
Entrada recidiva	3,42 (2,30 – 5,08)	2,77 (1,93 – 3,97)	2,77 (1,93 – 3,97)
Entrada reingresso pós abandono	2,66 (1,61 – 4,39)	2,28 (1,46 – 3,55)	2,28 (1,46 – 3,55)
Entrada transferência	1,97 (1,10 – 3,52)	1,68 (1,02 – 2,78)	1,68 (1,02 – 2,78)
Cultura de escarro positiva	4,32 (3,10 – 6,02)	3,77 (2,76 – 5,15)	3,77 (2,76 – 5,15)
HIV não realizado	0,64 (0,46 – 0,88)	0,68 (0,51 – 0,90)	0,68 (0,51 – 0,90)
Histopatológico	1,87	-	-

não realizado	(1,04 – 3,36)		
	0,42	0,54	0,54
Isoniazida	(0,25 – 0,69)	(0,35 – 0,86)	(0,35 – 0,86)
Baciloscopia 4 ^o mês positiva	1,89 (1,13 – 3,16)	-	-
Baciloscopia 6 ^o mês positiva	41,20 (26,56 – 63,97)	38,24 (26,96 – 54,24)	38,24 (26,96 – 54,24)

Fonte: SINAN-TB (CE)

Parte 4: Rede Neural Artificial

A partir dos modelos anteriores, selecionaram-se 20 atributos que contemplassem os quatro desfechos. Algumas delas estão presentes em modelos e em outros não, mas possuem importância em, pelo menos, um deles, sendo elas: sexo, idade, escolaridade, raça, zona de moradia, forma da doença, tipo de entrada, exame de baciloscopia de escarro, HIV, histopatológico, radiografia de tórax, PPD, medicações do tratamento (rifampicina, isoniazida, pirazinamida, etambutol), DOTS, baciloscopia de 2^o, 4^o e 6^o mês. Além disso, decidiu-se unir os desfechos em uma única variável denominada “encerramento” dividida em 5 fatores: cura, abandono, óbito, resistência e outros.

Esses modelos foram processados no programa WEKA versão 3.8, utilizando a função Multilayer Perceptron. Nessa função selecionou-se a visualização da GUI do modelo, o modo de entrada de dados em pequenos lotes “mini batches” de 20 instâncias, optou-se por usar a função de decaimento do aprendizado “learning decay”, a fim de não superestimar os dados.

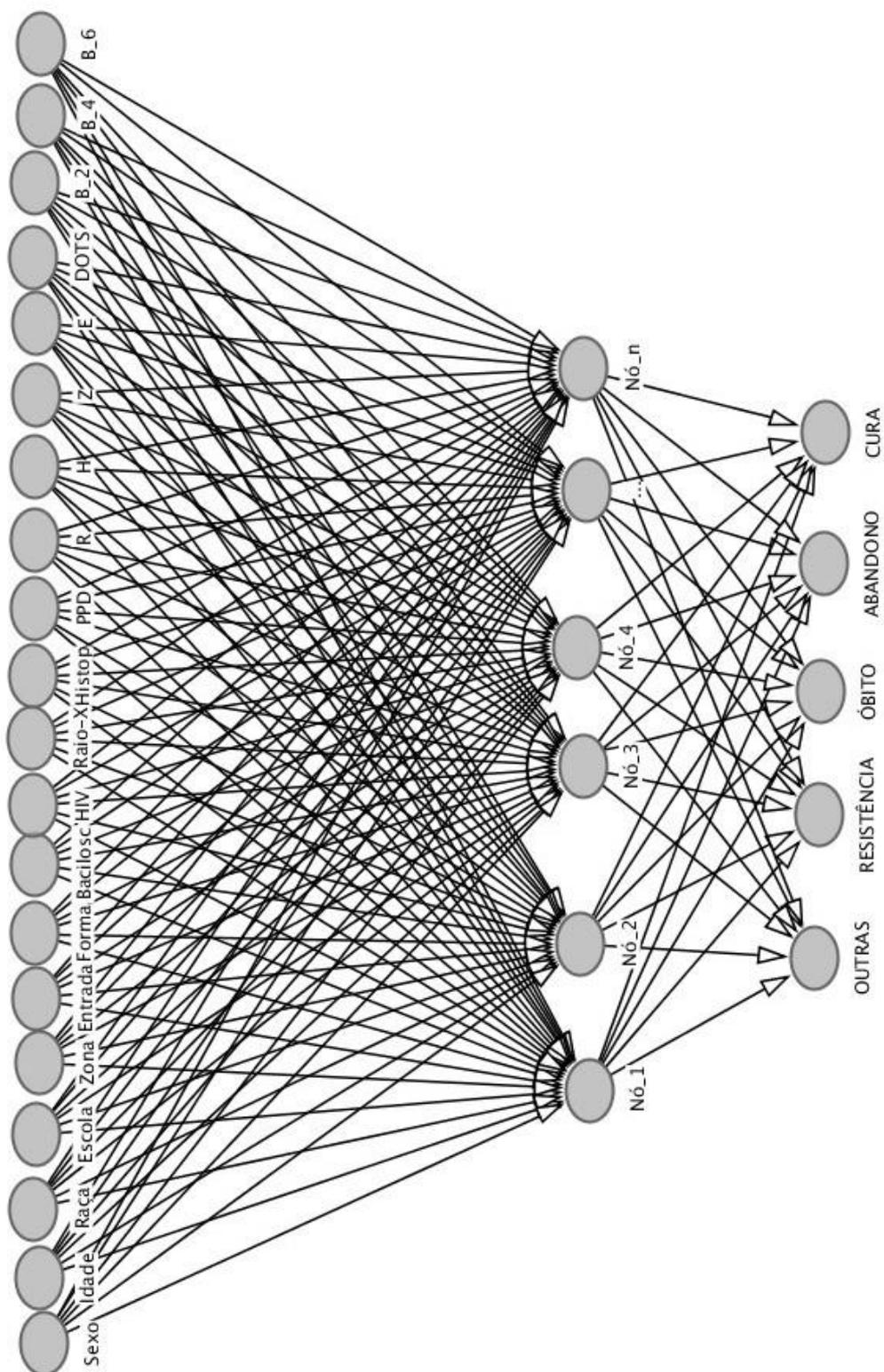
Além disso, utilizou-se como learning rate 0,3 e como momentum 0,2 e definiu-se 200 épocas para a criação, treinamento e predição do modelo. Os nós da camada oculta foram ativados por funções sigmóides. As únicas opções automáticas que permaneceram foram as de criação automática do modelo e ligações e do número de camadas ocultas e nós em cada uma delas.

O modelo foi criado utilizando 90% dos dados para treino e os outros 10% para teste, ou seja, das 74.006 instâncias, 67.002 foram utilizadas para treino e

7.004 para teste (escolhidos aleatoriamente pela máquina). A análise foi realizada em uma máquina com processador Intel® Core™ i7-3537U com sistema operando em 64-bit, contanto com 1TB de armazenamento, 8GB de memória RAM. Para ajudar na análise, a máquina ainda conta com 2GB extras dedicados a gráficos, por meio de uma placa de vídeo *AMD Radeon Graphics Processor*.

A máquina gerou um modelo preditivo com uma camada oculta e 43 nós nessa camada em aproximadamente 35 minutos para cada etapa. Para as três etapas (criação do modelo, aplicação no treino e teste) cerca de 1:30h foram necessários. Esse modelo considerou cada categoria de cada um dos nós de entrada (uma abstração resumida desse modelo pode ser observada na Figura 15). O menor erro por época foi considerado na época 200 (0,06558), entretanto a partir da época 100 o modelo já apresentou erro de 0,066. O modelo foi previamente treinado com diversos parâmetros para se chegar aos aqui exibidos.

Figura 15 – Modelo de rede neural artificial proposto pelo pesquisador



Fonte: próprio autor.

Considerando os cinco encerramentos, esse modelo apresentou uma área sob a curva ROC de 0,83. No quadro 3 é possível observar a matriz de confusão entre os dados reais e os preditos realizados por meio do teste nos 10% de dados restantes.

Quadro 3 – Valores reais e preditos dos desfechos do tratamento da tuberculose por um perceptron multicamadas

		VALORES PREDITOS					TOTAL
		CURA	ABANDONO	ÓBITO	TB-DR	OUTRO	
VALORES REAIS	CURA	4808	83	1	1	333	5226
	ABANDONO	561	186	1	1	180	929
	ÓBITO	74	12	2	1	53	142
	TB-DR	20	6	0	3	14	43
	OUTRO	607	72	2	3	377	1061
	TOTAL	6070	359	6	9	957	7401

Fonte: SINAN-TB(CE)

Avalie o modelo

Para avaliar os modelos multivariados foram usados critérios de sensibilidade, especificidade, valor preditivo positivo, valor preditivo negativo e área sob a curva ROC. Por meio dos modelos logísticos foi possível identificar os critérios descritos para cada um dos desfechos a seguir; entretanto foi realizado somente o do modelo logístico devido à possibilidade de estabelecer predições.

O modelo da cura apresentou uma concordância em relação a 25.751 casos com os valores reais identificados como positivos e de 2.622 casos identificados como não cura, assim como apresentado no banco de dados (Quadro 4). A tabela 15 apresenta os critérios de avaliação dos modelos e pode-se observar que ele apresentou alta sensibilidade (94,54%), especificidade de 35,21% e uma área sob a curva ROC de 0,85.

Quadro 4 – Matriz de confusão das predições do desfecho cura do tratamento

		Predição de cura		Total
		Sim	Não	
Caso de cura	Sim	25751	1487	27238
	Não	4761	2622	7383
	Total	30512	4109	34621

Fonte: SINAN-TB (CE)

O modelo logístico para o abandono do tratamento identificou 397 casos que realmente abandonaram o tratamento e 35.980 que não o fizeram. Esse modelo demonstrou um grande número de casos classificados como falso negativos (o modelo predisse 3.815 casos que não abandonaram o tratamento, mas eles o fizeram) (Quadro 5). Esse modelo mostrou-se pouco sensível (9,43%) e bastante específico (99,23%) e área sob a curva ROC de 0,83 (Tabela 15).

Quadro 5 – Matriz de confusão das predições do desfecho abandono do tratamento

		Predição de abandono		Total
		Sim	Não	
Caso de abandono	Sim	397	3815	4.212
	Não	281	35.980	36.261
	Total	678	39.795	40.473

Fonte: SINAN-TB (CE)

O modelo que apresentou mais problemas de predição foi o para o óbito por TB. Ele não conseguiu identificar nenhum caso como abandono, falseando 565 resultados que evoluíram para esse desfecho (Quadro 6). Dessa forma ele apresentou sensibilidade de 0% e especificidade de 100%, apresentando uma área sob a curva ROC de 0,58 (Tabela 15).

Quadro 6 – Matriz de confusão das predições do desfecho óbito por TB

		Predição de óbito por TB		Total
		Sim	Não	
Caso de óbito por TB	Sim	0	565	565
	Não	0	38.558	38.558
	Total	0	39.123	39.123

Fonte: SINAN-TB (CE)

O último modelo, o da resistência, classificou corretamente 18 pessoas com TB resistente assim como 42.275 que não a apresentaram ao fim do tratamento (Quadro 7). Foi possível identificar que o modelo é pouco sensível (9,38%) e bastante específico (99,95%); além disso, ele apresentou uma área sob a curva ROC de 0,86 (Tabela 15).

Quadro 7 – Matriz de confusão das predições do desfecho resistência ao tratamento da TB

		Predição de TB resistente		Total
		Sim	Não	
Caso de TB resistente ao tratamento	Sim	18	174	192
	Não	20	42.275	42.295
	Total	38	42.275	42.487

Fonte: SINAN-TB (CE)

Tabela 15 – Apresentação de critérios de avaliação dos modelos preditivos logísticos. Fortaleza (CE), 2018.

	Cura	Abandono	Óbito	Resistência
Sensibilidade	94,54%	9,43%	0,0%	9,38%
Especificidade	35,51%	99,23%	100,0%	99,95%
Valor Preditivo Positivo	84,40%	58,55%	0,0%	47,37%
Valor Preditivo Negativo	63,81%	90,41%	98,56%	99,59%
Área Sob a Curva ROC	0,85	0,83	0,58	0,86
Corretamente classificados (acurácia)	81,95%	89,88%	98,56%	99,54%

Fonte: SINAN-TB (CE)

Para avaliar o modelo baseado na RNA de um perceptron multicamadas, os mesmos critérios foram utilizados; entretanto, foi utilizado o critério de validação cruzada com 10-fold. O modelo foi criado, aplicado e testado pelos mesmos processos utilizados na sua criação. A única diferença foi que ele considerou todas as 74.006 instâncias para validar. Cada parte do modelo (fold) demorou 20 minutos para ser treinada. Assim, para as 10 partes necessárias, foi estimado um tempo de 3 horas e 20 minutos para a criação desse modelo de avaliação.

Na matriz de confusão a seguir (quadro 8) é possível identificar que 74,21% das instâncias foram corretamente classificadas, o erro por época gerado foi semelhante ao do modelo 90/10 (0,065). Por outro lado, a estatística Kappa mostrou concordância relativamente baixa (0,33).

Quadro 8 – Valores reais e preditos dos desfechos do tratamento da tuberculose por validação cruzada (10-fold) de perceptron multicamadas

VALORES REAIS		VALORES PREDITOS					
		CURA	ABANDONO	ÓBITO	TB-DR	OUTRO	TOTAL
	CURA	49002	1335	7	0	2187	52531
	ABANDONO	5508	2526	4	0	1242	9280
	ÓBITO	762	161	34	0	556	1513
	TB-DR	225	59	0	2	102	388
	OUTRO	5836	1082	16	1	3359	10294
	TOTAL	61333	5163	61	3	7446	74006

Fonte: SINAN-TB (CE)

Além disso, o modelo mostrou-se sensível para detectar casos de cura (0,799). Mostrou bons valores preditivos positivo para todos os desfechos, em especial para a cura e resistência. As áreas sob curva ROC de cada desfecho foi superior a 0,80 em todos os desfechos e a rede neural apresentou capacidade de corretamente classificar 74,21% de todos os casos (Tabela 14).

Tabela 16 – Apresentação de critérios de avaliação do modelo preditivo baseado em perceptron multicamadas. Fortaleza (CE), 2018.

	Cura	Abandono	Óbito	Resistência	Outro
Sensibilidade	0,799	0,279	0,022	0,005	0,326
Especificidade	0,626	0,952	0,999	0,999	0,926
Valor Preditivo Positivo	0,932	0,489	0,557	0,666	0,451
Valor Preditivo Negativo	0,193	0,885	0,973	0,993	0,881
Área Sob a Curva ROC	0,859	0,825	0,849	0,804	0,813
Corretamente classificados (acurácia)	74,21%	74,21%	74,21%	74,21%	74,21%

Fonte: SINAN-TB (CE)

Avaliação dos modelos

Avalie resultados

Diferentemente da etapa anterior, essa etapa não foca em determinar a acurácia dos modelos, mas em determinar o quanto eles vão de encontro ao objetivo proposto. Dentre os resultados provenientes dos modelos epidemiológicos pode-se observar que a regressão logística tende superestimar os resultados das razões de chances (entendidas aqui como razões de prevalência) e a fornecer intervalos de confiança mais largos.

Assim, como opção surgem os modelos de Poisson ou binomial negativo. Entretanto, deve-se ressaltar que, quando aplicado a desfechos não numéricos, eles devem ser acrescidos do estimador de variância robusta a fim de reduzir o erro produzido por desfechos categóricos. Dentre os dois, preferiu-se o modelo de Poisson robusto devido ao consumo de tempo menor que o binomial negativo.

Além disso, o modelo preditivo baseado em IA se mostrou com uma boa capacidade de prever os desfechos do tratamento da TB. Dessa forma, entende-se que os modelos criados estão de acordo com os objetivos propostos e, principalmente, o criado por meio de algoritmo inteligente insere-se como uma alternativa de classificação comparado a modelos epidemiológicos baseados em regressões.

Revise o processo

A etapa de revisão surge como etapa para repensar os passos tomados e o que poderia ter sido feito para melhorar o desempenho dos modelos criados. Nesse sentido, entende-se que talvez uma mudança no número de lotes de entrada, na taxa de aprendizado e no momentum poderiam produzir modelos mais acurados. Entretanto, essas etapas consomem bastante tempo e podem ser refeitas no futuro.

Além disso, a escolha dos atributos se deu baseado em modelos epidemiológicos também criados pelo pesquisador, talvez o uso de técnica como a análise dos componentes principais possa ser uma alternativa viável. Tornar o modelo menos complexo também seja uma alternativa, mas é um desafio tendo em vista que a TB em si é um problema de saúde pública complexo de ser abordado.

Determine os próximos passos

Devido ao prazo do pesquisador o modelo criado irá direto para a fase de entrega do projeto. No futuro, etapas como testar esses modelos em pacientes que apresentem a doença para determinar se a máquina acerta os desfechos seja uma etapa a ser realizada. Além disso, reajustar o modelo a bases com menos dados ausentes é essencial para que ele possa ser amplamente utilizado.

Entrega do projeto

Plano de entrega

O projeto foi entregue por meio de defesa pública de dissertação de Mestrado apresentado no Programa de Pós-graduação Cuidados Clínicos em Enfermagem e Saúde e avaliado por banca examinadora composta por professores doutores das diversas áreas que esse projeto envolve: infectologia, computação, análise estatística aplicada à enfermagem e saúde. Em relação aos resultados obtidos, eles ficarão armazenados no computador do pesquisador, assim como em armazenamento em nuvem.

Plano para monitoramento e manutenção

Após considerações de todos os membros da banca examinadora sobre o processo de criação dos modelos e arguição do pesquisador ele, modificações podem ser realizadas a fim de melhorar a qualidade dos resultados e desempenho dos modelos desenvolvidos. Eles podem passar por novos processos de treino e teste com a inserção de argumentos e processos que resultem em maior acurácia tanto da parte epidemiológica, como da computacional que essa pesquisa envolve.

Produza o relatório final

O método CRISP-DM, no qual essa pesquisa se embasou, envolve uma série de passos distintos, como a produção de relatórios antes, durante e após a modelagem dos dados. Por mais que não apareçam dentro dessa sessão, eles estão incluídos na dissertação de mestrado que está sendo entregue. A mesma possui capítulos específicos sobre o motivo de essa investigação ter sido desenvolvida e como os resultados obtidos se encontram dentro do que já foi

produzido nacional e internacionalmente. Portanto, esse documento como um todo se encaixa como o relatório final de pesquisa.

Revise o projeto

Desde a decisão de que o projeto iria se embasar em um método de mineração de dados e aplicação de algoritmos para criação de modelos inteligentes, o pesquisador está sendo desafiado devido à nova abordagem proposta. Estudou-se assuntos relacionados à epidemiologia, bioestatística e computação, como programação e apreensão de conceitos até então desconhecidos, além de debruçar-se cada vez mais dentro do objeto de pesquisa relacionado ao encerramento do tratamento da TB.

Experiência foi adquirida durante o desenvolvimento de tal projeto. Entretanto, entende-se que o período de dois anos que o mestrado produz ainda se fez reduzido diante da complexidade do projeto proposto. Acredita-se que modelos menos complexos, com menos desfechos, sejam uma alternativa para pesquisadores iniciantes no campo de mineração dos dados.

Em relação aos modelos, também é considerável refletir sobre a quantidade de atributos que se pretende utilizar, pois cada um possui um peso que deve ser mensurado dentro dos modelos. Em relação a isso, futuros projetos devem considerar o peso de cada ligação aos nós das camadas ocultas e também os pesos das variáveis nos modelos epidemiológicos. Entende-se que a ausência desse processo como uma possível limitação desse projeto. Além disso, é importante considerar o tempo consumido para determinar as camadas ocultas, nós, entrada de dados, taxa de aprendizado, momentum, épocas, treino e teste dos modelos.

7 DISCUSSÃO

Um conjunto de resultados puderam ser obtidos por meio desta investigação. Assim, faz-se necessário discutir diversos aspectos que eles trazem. Devido a isso, decidiu-se dividir a discussão desse trabalho em tópicos menores para abordar melhor cada uma das temáticas estudadas que, a seguir, serão apresentadas.

7.1 EVOLUÇÃO TEMPORAL DOS INDICADORES EPIDEMIOLÓGICOS DA TUBERCULOSE

Este estudo evidenciou a evolução temporal de alguns dos principais indicadores da tuberculose. Em relação à incidência e prevalência da doença, pode-se constatar significativa diminuição no decorrer do período de 2001 a 2017. Corroborando com essa pesquisa, estudo realizado em Araraquara (SP) identificou que incidência de TB caiu de 30,24/100 mil habitantes no período 2002-2006 para 23,6/100 mil habitantes no período de 2007-2011 (TELAROLLI JUNIOR, LOFFREDO, GASPARETTO, 2017).

Já o estado de São Paulo apresentou coeficientes de incidência estáveis para o período de 2006 a 2013, sendo observado que eles tendem a ser duas vezes maiores, quando observados no sexo masculino (PINTO, P.F.P.S, et al., 2017). Em consonância com essas estimativas, em Belo Horizonte (MG) também houve redução da incidência da TB entre os anos de 2002 a 2008, especialmente no período 2004-2007 (REIS et al., 2013). Internacionalmente, pode-se observar queda na incidência da TB em Buenos Aires (Argentina) de 49,7/100 mil habitantes em 2000 para 37,9/100 mil habitantes em 2011, indicando queda na variação percentual anual de 2,71% ao ano (CHIRICO et al., 2016).

Para o indicador da taxa de cura, pode-se identificar decréscimo de 1,3% ao longo dos anos. Já em um município de São Paulo foram obtidas taxas de cura maiores que 70% durante o período de 2002-2011 (TELAROLLI JUNIOR, LOFFREDO, GASPARETTO, 2017). Em um município cearense, identificou-se taxa de cura estimada em 77% (PINTO, M.L., et al., 2017).

Ainda em relação à cura, em pesquisa de Heck, Costa e Nunes (2013) não houve mudança significativa dela no período de 2000-2008 em um município do

Rio Grande do Sul. Finalmente, em estudo realizado em país africano, pode-se evidenciar aumento de 75% para 88% na taxa de cura (DEMEMEWE et al., 2016).

A taxa de abandono do tratamento é um indicador importante por se identificar como os programas de manejo da doença estão fazendo os pacientes aderirem ao tratamento. Por meio desse estudo pode-se identificar flutuação desse indicador que hora cresce e hora decresce. Em outras investigações que estudaram o abandono do tratamento foi possível identificar que a tendência se deu diferente da cearense, pois o primeiro apresentou diminuição das suas taxas durante o período de 2001 a 2014 (SOARES et al., 2017).

Em Salvador, a taxa de abandono permaneceu com poucas alterações durante 2006 a 2015 (SANTOS, MARTINS, 2018). Em Araraquara (SP), a tendência mostrou-se crescente ao longo dos anos (TELAROLLI JUNIOR, LOFFREDO, GASPARETTO, 2017). No Crato (CE), a tendência de abandono esteve presente em 12,6% dos casos (PINTO, M.L., et al., 2017).

Outra pesquisa apontou mudança significativa nas alterações dessa taxa em Sapucaia do Sul (RS), evidenciando anos com mais de 13% de abandono do tratamento (HECK, COSTA, NUNES, 2013). Já em estudo realizado na Etiópia pode-se identificar diminuição de 2% para 0,9% na taxa de abandono de casos notificados de TB (DEMEMEWE et al., 2016).

Outros indicadores importantes relacionados aos programas de TB são os relacionados ao óbito: letalidade e mortalidade. Dos dois, este estudo pode evidenciar aumento significativo da mortalidade em 15,2% ao ano. Estudo de Ceccon et al. (2017) evidenciou que, das capitais nacionais, as que apresentam maiores coeficientes de mortalidade são Manaus (6,4/100.000 habitantes), Belém (5,1/100.000 habitantes) e Recife (4,9/100.000 habitantes).

A mortalidade por TB mostrou decréscimo em São Paulo, assim como em cidade do Nordeste onde a taxa de óbito esteve em torno de 4,5% (TELAROLLI JUNIOR, LOFFREDO, GASPARETTO, 2017; PINTO, M.L., et al., 2017). Além desses, em Minas Gerais tendência similar também pode ser identificada entre o período de 2002 a 2008, com média anual de 1,6 casos por 100 mil habitantes. Da mesma forma, o estado do Paraná seguiu tal tendência (REIS et al., 2013; CECILIO, 2018).

Em contrapartida, estudo realizado em Santa Catarina não identificou mudanças significativas na mortalidade por TB durante o período de 2007-2011. Os mesmos resultados foram obtidos em pesquisa realizada no Rio Grande do Sul (VILELA et al., 2018; HECK, COSTA, NUNES, 2013).

Poucos estudos evidenciam os padrões temporais de formas resistentes ao tratamento. Por esse motivo, esse indicador foi adicionado nessa investigação, onde foi possível identificar que ela tende a crescer 7,9% ao ano. Estudo de Dalal et al. (2015) demonstrou que, embora não significativo, há decréscimo no padrão temporal de formas multirresistentes de TB ao passo que há aumento nas formas extensivamente resistentes. Assim, faz-se necessário o uso de novos testes para a detecção do desfecho resistência no contexto da epidemia mundial da TB (BARREIRA, 2018).

Dessa forma, é importante mencionar que a falta de investimento nos programas de manejo da TB pode prejudicar resultados do desfecho do tratamento. Não é possível, por exemplo, entender que uma região apresenta bons indicadores de cura, mas os seus indicadores de abandono e óbito também crescem (LIMA et al., 2016). Dessa forma, entender o panorama por completo, e não somente parte dele, garante que o mascaramento de informações não ocorra.

Como forma de intervenção, a literatura sugere que medidas de proteção social, como seguridade social e salarial, provisão de alimentos, apoio psicossocial e acesso universal a sistemas de saúde são efetivamente importantes nos programas de TB. Estudos mostram que estes programas aumentam as taxas de cura, ao passo que diminuem as de abandono e mortalidade pela doença (ANDRADE et al., 2018; PALHA et al., 2012).

Outro plano fundamental para a mudança nesse panorama é a estratégia “End TB”, que possui a ousada proposta de eliminar a epidemia da doença até 2035, reduzindo a incidência em 90% e a mortalidade em 95%. Ela possui metas que somente serão atingidas com novas propostas dentro desse programa, trazendo em seu pilar 3 a intensificação da pesquisa e inovação na área. Ressalta-se a capacidade do Brasil de atingir tal meta, pois, além de ser o país com as menores taxas de incidência e mortalidade dentre os 30 países com alta carga de TB, possui

um sistema de saúde universal, que oferece gratuitamente todo o aporte terapêutico à pessoa doente (BARREIRA, 2018).

Portanto, ao observar decréscimo nos coeficientes de incidência, prevalência da doença não é o único ponto a se considerar. Indicadores importantes, como taxa de cura, abandono, óbito e resistência continuam crescendo ou, quando não, apresentam valores ainda elevados. Dessa forma, ao entender o comportamento da doença ao longo dos anos, faz-se vital investir em estratégias que retomem o aumento da cura e reduzam a proporção de todos os outros desfechos desfavoráveis.

7.2 FATORES ASSOCIADOS A CURA DA TUBERCULOSE

A cura é o desfecho favorável do tratamento da tuberculose e, dessa forma, este estudo evidenciou que o sexo masculino é mais associado com a diminuição da prevalência de cura no estado do Ceará (62% de todos os casos eram do sexo masculino e 31,12% destes não curaram, $RP = 0,8$). Estudo realizado no Rio de Janeiro identificou que, dos 4.208 casos estudados, 63,1% eram do sexo masculino e, destes, 68,7% ($n = 774$) não alcançaram a cura da doença (PRADO JUNIOR, VIRGÍLIO e MEDRONHO, 2016). Estudo realizado no Rio de Janeiro também encontrou associação do sexo masculino com menores chances de cura ($OR = 0,71$) da doença (SANTOS et al., 2018).

Ainda em relação à diferença entre os sexos, estudo realizado na Colômbia aponta que ambos experenciam a doença de maneira semelhante, mostrando pouco conhecimento antes da infecção. Entretanto, tal estudo evidencia que mulheres tendem a subestimar seus sintomas, devido ao seu papel dentro das famílias e os homens tendem a superestimá-los. Mesmo assim, a procura do sexo masculino por serviços de saúde é menor e somente realizada quando influenciada por algum membro da família (VÉLEZ, MONTOYA, 2015).

Nessa investigação foi possível identificar que a raça parda está associada à diminuição da prevalência da cura da TB. Estudo de Mendes et al. (2016) identificou que no estado do Rio Grande do Sul a raça parda também estava associada com taxas menores desse desfecho (66,2%). Nesse mesmo estudo

identificou-se que tal desfecho estava mais associado com pessoas autodeclaradas brancas, considerada como categoria de referência nessa pesquisa. Da mesma forma, outra investigação identificou que raças não brancas diminuem em 25% as chances de cura da doença (PRADO JUNIOR, VIRGÍLIO e MEDRONHO, 2016). Santos et al. (2018) também identificaram menores chances de desfecho positivo da doença quando o paciente é declarado preto (OR = 0,62) ou pardo (0,73).

Numa análise bivariada, estudo de Prado Junior, Virgílio e Medronho (2016) identificou que a escolaridade a partir do ensino médio aumenta as chances de cura de pessoas com TB (OR = 1,7 para o ensino médio e OR = 2,1 para ensino superior). Quando comparado aos analfabetos, Santos et al. (2018) identificaram que, a partir de 9 anos de estudos, as chances de cura da doença são maiores. Corroborando com esses achados, Silva et al. (2017, p. 116) apontam que “a baixa escolaridade sugere baixo nível socioeconômico e dificuldade de acesso aos serviços de saúde”.

Diferentemente dos resultados dessa pesquisa, que apontaram maior prevalência de cura em pessoas da zona rural (RP = 1,25), estudo de Santos et al. (2048) evidenciou que, no Rio de Janeiro, estas chances também são maiores para desfechos desfavoráveis (OR = 2,27).

Esse estudo encontrou que qualquer forma de entrada diferente de caso novo diminui a prevalência da cura do tratamento da tuberculose. Destaca-se a importância da completude do tratamento. Assim, entrada como recidiva ou reingresso pós abandono devem ser consideradas, pois diminuem a cura da doença. Corroborando com essa investigação, no Rio de Janeiro, qualquer outra forma de entrada diferente de caso novo apresentou menores chances de cura (SANTOS et al., 2018). Além disso, estudo realizado em Fortaleza enfatiza a importância de casos novos no aumento da cura ou completude do tratamento (PIRES NETO et al., 2012).

Nesse estudo, a forma extrapulmonar foi identificada como fator de aumento da prevalência de cura (RP = 1,05). Estudo de Pires Neto (2012) identificou que, das formas extrapulmonares coinfectadas com HIV, 40% evoluíram para cura ou completude do tratamento.

O HIV desempenha grande papel na coinfeção com a TB, estando associado com formas extrapulmonares. Nesse estudo identificou-se que o exame de HIV, quando positivo ou não realizado, diminuem significativamente a prevalência da cura da TB. Assim como essa investigação, estudo de Cardoso et al. (2017) identificou que as chances de cura da doença são diminuídas em 66%, caso o paciente tenha HIV e não esteja em terapia antirretroviral (TARV) ou em 35% caso faça uso dessa terapia.

Além disso, Gaspar et al. (2016) encontraram que durante o período de 2002 a 2012 a taxa nacional de cura da TB encontrava-se em 71,10%. Entretanto, quando associada com o HIV, essa taxa caiu 58%, chegando à proporção de apenas 50,74% de cura dos casos. Em estudo realizado na China, pacientes com HIV que apresentaram TB obtiveram uma taxa de cura da segunda de 69,2% (162/234) e outros 23 tiveram melhora de sintomas, mas não apresentaram cura da doença (MAIMAITI et al., 2017).

Em relação ao tratamento medicamentoso, esse estudo mostrou que fármacos do esquema básico, como a rifampicina e a isoniazida, elevam a prevalência de cura da doença. O etambutol a diminui e a estreptomicina (utilizada em esquemas especiais) também contribui para a cura da doença. Dessa forma, Ferreira et al. (2013) afirmam que a dose fixa combinada (DFC) dos medicamentos para a TB (rifampicina-R, isoniazida-H, pirazinamida-Z e etambutol-E) pode não estar fazendo diferença nos regimes na situação de encerramento da doença.

Ademais, a inclusão do etambutol nesse esquema se deu, justamente, pelo aumento de casos resistentes ao tratamento inicial. Entretanto, percebeu-se que seu uso dentro dos esquemas não possui alteração nos casos de cura, pois a literatura já apontava que os casos de falência do esquema RHZ era baixo (RABAHI, SILVA JUNIOR, CONDE, 2017). Além desses, estudo realizado com 208 pacientes no Rio de Janeiro não encontrou diferenças significativas entre as taxas de cura de paciente que utilizavam o esquema RHZ e RHZE/DFC (SILVA, MELLO, FIGUEREDO, 2017).

Além desses dados, faz-se importante destacar o apoio familiar, de amigos ou rede de apoio como essencial durante o curso do tratamento para TB. Dessa forma, a adesão ao tratamento da TB se dá por aspectos que envolvem não

apenas a tomada diária das medicações, mas também perpassa sob o próprio usuário se reconhecer com a doença, da sua responsabilidade em relação a si mesmo, anseio por autocuidado, do seu desejo de curar e também do apoio recebido por familiares (RODRIGUES et al., 2017; PALHA et al., 2012; VÉLEZ, MONTOYA, 2015).

Por fim, deve-se dar a devida importância ao acompanhamento dos pacientes, para que estes atinjam a cura da TB. Nesse estudo foi possível identificar pequena diminuição da prevalência de cura nos pacientes em tratamento supervisionado (RP = 0,98) e que, quando positivas ao longo dos meses, as baciloscopias bimestrais também diminuem tal prevalência.

Paz et al. (2012) destacam que a implantação de programas diretamente observados ainda possui diversas barreiras dentro dos serviços de saúde e que os usuários continuam realizando tratamento autoadministrado. Como alternativa a esse problema, é sugerido que tratamentos supervisionados sejam realizados apenas em casos com maiores chances de abandonar o tratamento e que o tratamento autoadministrado seja realizado nos demais casos, com provisão de medicação de doses para 30 dias e a cada retorno seja realizada consulta priorizando o paciente, valorizando suas queixas, controle de baciloscopia e a entrega de medicações.

Assim, pode-se identificar os diversos aspectos epidemiológicos que influenciam na cura da TB. São necessárias, então, estratégias que foquem em fatores que aumentem a prevalência de tal desfecho, assim como incentivo e investimento a programas de manejo da doença, numa esfera local, nacional e global.

7.3 FATORES ASSOCIADOS AO ABANDONO DO TRATAMENTO DA TUBERCULOSE

O abandono do tratamento da TB ainda é um obstáculo no sucesso dos programas anti-TB e, por isso, necessita de profunda investigação. Assim, esta pesquisa identificou que ser do sexo masculino aumenta em 25% as chances de prevalência de abandono do tratamento. Estudo de Cardoso et al. (2017) encontrou

que o sexo feminino diminui as chances de abandono (OR = 0,27), resultado semelhante encontrado por Gomes et al. (2015). Estudo realizado em Pernambuco identificou que a maior proporção de abandono também se dá em pessoas do sexo masculino (SOARES et al., 2017); resultado semelhante é encontrado em estudo realizado no Mato Grosso do Sul (BASTA et al., 2013) e em Fortaleza (HARLING et al., 2017).

Nessa investigação não foi possível identificar, por meio de modelos epidemiológicos multivariados, que a raça/cor possuía influência no abandono do tratamento da TB no estado do Ceará. Entretanto, estudo de Mendes et al. (2016) encontrou que maior porcentagem de casos de abandono (18,3%) da doença estava em pessoas autodeclaradas negras.

A idade mostrou-se como fator de proteção mínimo ao abandono do tratamento, pois, pelos três modelos multivariados, ela mostrou-se com padrão borderline. Mesmo assim, estudos trazem sus a importância dentro da linha de cuidado à pessoa com TB e prevenção do abandono do tratamento. Por exemplo, o mesmo padrão pode ser identificado por Snyder et al. (2016). Já Cardoso et al. (2017) também identificaram achados semelhantes no sentido de que a idade possa diminuir o abandono ao tratamento.

Além disso, observou-se que, a partir do nível do ensino médio, a taxa de prevalência de abandono tende a diminuir significativamente. Estudos realizados em outras regiões Brasileiras identificaram que estudo até ensino fundamental contribui significativamente para o abandono o tratamento (SILVA, MOURA, CALDAS, 2014; CARDOSO et al., 2017). Diferente desses achados, pesquisas conduzidas por Silva, Anjos e Nogueira (2014) e por Silva, Andrade e Cardoso (2013) identificaram que essas faixas de escolaridade estão mais associadas com o abandono do tratamento.

A zona de moradia rural mostrou-se como fator de proteção ao abandono da doença. Isso pode explicar que a maioria dos casos se dá na zona urbana, assim como seus desfechos negativos. Entretanto, estudo de Basta et al. (2013) não identificou diferenças significativas em relação ao abandono do tratamento e à zona de moradia.

Em relação ao tipo de entrada, foi possível identificar que, quando comparado a caso novo, a prevalência de abandono tende a aumentar para casos

de recidiva (RP = 1,31) e principalmente para reingresso pós abandono (RP = 2,27). Ademais, efeito contrário pode ser observado quando a entrada se dá como transferência (RP = 0,75). A literatura endossa esses achados, evidenciando que o reingresso pós abandono é uma das causas que aumentam as chances de um novo abandono do tratamento (SILVA, ANJOS, NOGUEIRA, 2014; GOMES et al., 2015; SILVA, ANDRADE, CARDOSO, 2013).

A forma extrapulmonar foi identificada como fator de decréscimo do abandono do tratamento (RP = 0,77). Da mesma forma, Gomes et al. (2015) identificaram OR = 0,31 (p=0,10) para formas extrapulmonares da doença e o abandono do tratamento. Semelhante a esses achados, Harling et al. (2017) identificaram que as chances de abandono decrescem 26% quando a doença se manifesta de forma extrapulmonar, já Snyder et al. (2016) identificaram 56% de decréscimo do abandono do tratamento quando associado a formas não pulmonares.

Outro fator importante no abandono do tratamento da TB é a sua coinfeção com o HIV. Esta investigação identificou aumento na prevalência de tal desfecho quando há tal coinfeção. Além disso, quando associado ao HIV, as chances de abandono do tratamento, em escala nacional, apresentaram significativo crescimento (de 9,52% para 13,60%) no período de 2002 a 2012 (GASPAR et al., 2016). Além disso, pesquisa de Harling et al. (2017) identificou que essa associação aumenta em 2,10 vezes as chances do referido desfecho e pesquisa de Snyder et al. (2016) identificou aumento de 1,97 nas chances de abandono.

Além disso, identificou-se que a não realização de alguns exames diagnósticos aumentam as chances de abandono do tratamento, como a radiografia de tórax (RP = 1,13), o teste tuberculínico (RP = 1,34), o exame de HIV (RP = 1,46) e o histopatológico (RP = 1,21). Entende-se que o sucesso de programas de manejo da doença está associado a baixas taxas de abandono do tratamento, mesmo em locais com menor carga de TB (ARAKAWA et al., 2017).

Palha et al. (2012) apontam que a não contratualização entre usuário e sistema de saúde em relação ao tratamento e como ele deve ser procedido promovem sentimentos de não cuidado e respeito por parte dos usuários, colaborando para o abandono do tratamento. Além disso, afirma-se que o momento

de comunicação na revelação do diagnóstico é crucial para a adesão do usuário ao tratamento da TB. Estudo destaca que, quando a notícia é dada para o indivíduo sem devidas explicações e abordagem grosseira colaboram para a não adesão a ele (OLIVEIRA, LEFVRE, 2017).

Em relação ao acompanhamento do tratamento, destaca-se que a partir do 4º mês os exames baciloscópicos de rotina apresentando resultado positivo ou a sua não realização aumentam significativamente a prevalência de abandono da doença. Dessa forma, o acompanhamento dessas pessoas se faz essencial, principalmente na realização das baciloscopias bimestrais. Assim, percebe-se que maiores níveis de cobertura do tratamento à pessoa com TB promovem diminuição do abandono do tratamento (PALHA et al., 2012).

Assim como a cura da doença também se dá por meio de uma rede de apoio à pessoa com TB, o abandono do tratamento se dá justamente pela sua falta. Destaca-se que, ao descobrir a doença do familiar, membros da família afastam-se por medo do contágio, levando à desunião do elo familiar. Como alternativa a esse problema, destaca-se o papel do profissional de saúde, em especial do enfermeiro, como agente de mudanças, esclarecendo dúvidas e fornecendo apoio inicial ao usuário (CHIRINOS, MEILERELLES, BOUSFIELD, 2015).

Nessa perspectiva, ajudar o usuário a entender sua condição de saúde e apoiá-lo durante essa fase faz-se como uma estratégia de estímulo a continuar o percurso do tratamento. Enfatiza-se nessa ocasião o aspecto tratável e curável da TB e do seguimento do tratamento sem interrompê-lo (COUTO et al., 2014).

Além desses, embora não considerado nessa investigação, um ponto que merece destaque é o abandono do tratamento associado a efeito adversos das medicações que fazem parte do regime terapêutico. Diversos estudos trazem a importância desse fator dentro da linha de cuidado ao paciente, como fator determinante no abandono no tratamento da TB (COUTO et al., 2014).

Portanto, como foi possível observar, o abandono do tratamento está relacionado a diversas variáveis do próprio ser humano (como o sexo e a idade), do seu estado de saúde (HIV), do serviço de saúde (exames realizados e seu acompanhamento). Entretanto, é necessário entender que esse desfecho vai além desses dados, pois é complexo e multifatorial. Dessa forma, pesquisas que

investiguem esses outros aspectos do abandono do tratamento se fazem necessárias

7.4 FATORES ASSOCIADOS AO ÓBITO POR TUBERCULOSE

Dos fatores que influenciaram o óbito de paciente com TB, esse estudo identificou que a idade aumenta em 3% a prevalência desse desfecho. No Rio de Janeiro, estudo realizado por Escada et al. (2017) encontrou resultado igual ao dessa investigação, porém sem significância estatística ($p=0,09$). Já Liew et al. (2015) encontraram que, na Malásia, a idade influencia em 4% na mortalidade ($p<0,001$). Ressalta-se que, enquanto a primeira pesquisa enfocava apenas pacientes coinfectados, a segunda investigação considerou todos os pacientes cadastrados na base.

Ademais, pesquisa realizada nos Estados Unidos identificou que pessoas a partir de 65 anos possuem risco 5,5 vezes maior para o óbito por TB (KATTAN, SOSA, LOBATO, 2012). Snyder et al. (2016) observaram aumento de 4% nas chances de óbito para cada ano completo. Outros estudos também identificaram a mesma relação (GOMES et al., 2015; CHAVES et al., 2017; GAO et al., 2015; GAIFER, 2017; ABDELBARAY et al., 2017). Entretanto, deve-se atentar para essa variável, pois o passar da idade influencia no curso natural de vida do ser humano. Deve-se atentar para que esta variável que deve continuar considerada em estudos futuros.

A raça/cor parda se apresentou como importante fator para aumento da prevalência do óbito por TB (RP = 1,63). Corroborando com essa pesquisa, estudo identificou que o óbito da TB foi mais comum (2,9%) em pessoas autodeclaradas pardas (MENDES et al., 2016).

Em relação a exames diagnósticos, estudo identificou que a radiografia suspeita aumenta em 3,07 as chances para o óbito por TB (SANTOS-REIS et al., 2013), corroborando com esta pesquisa, que também identificou que a radiografia suspeita influencia em 3,05 vezes mais a prevalência do óbito por TB. Resultado semelhante foi evidenciado no estudo de Liew et al.(2015), porém de forma mais branda (OR = 1,26).

Além disso, faz-se necessário ressaltar que o HIV possui influência importante no óbito por TB. Resultados desta pesquisa indicam que a prevalência de tal desfecho aumenta em 2,95 quando a pessoa se mostra coinfetada. Evidenciou-se que o óbito da doença tende a crescer em 94% quando a TB se apresenta em conjunto à infecção por HIV (GASPAR et al., 2016). Maimaiti et al. (2017) demonstraram que dos 234 pacientes com coinfecção TB/HIV, 31 (13,2%) evoluíram para óbito e o tempo de sobrevivência dos pacientes soropositivos decresceu significativamente quando a TB aparece como infecção oportunista. Além disso, estudo de Reis-Santos et al. (2013) não observou associação do HIV com o óbito por TB em pessoas com doença renal crônica, mas identificou essa associação em relação ao óbito por outras causas (OR = 3,32)

Gomes et al. (2015) identificaram aumento de 6,01 vezes de chance de óbito, quando o HIV esteve associado à TB, Snyder et al. (2016) observaram aumento de 8,79 vezes nas chances de óbito para a coinfecção. Já Liew et al. (2015) evidenciaram aumento de 9,3 vezes para o óbito quando o paciente se apresentou com coinfecção. Outros estudos também evidenciaram resultados semelhantes e a importância de se entender o importante papel que a coinfecção TB/HIV pode gerar no processo adoecer-morrer (GAIFER, 2017; DJOUMA et al., 2015; GAO et al., 2015, ABDELBARY et al., 2017).

A sinergia com a TB é um importante fator, quando pessoas são diagnosticadas HIV-positivas, pois se identifica que, durante o tratamento antirretroviral, há 5,4 mais chances de ocorrência do óbito por TB (SALVADORI et al., 2017). Esse fator também foi evidenciado em pesquisa que identificou que a demora do tratamento para TB durante o acompanhamento em pessoas com HIV aumenta o risco de mortalidade (NAGU et al., 2017a).

O último fator abordado nessa parte do estudo foi o acompanhamento do tratamento por meio da realização do Tratamento Diretamente Observado (DOTS) e das baciloscopias bimestrais. Embora o tratamento supervisionado não tenha aparecido como preditor do óbito, a baciloscopia de 6º mês (que integra o programa) demonstrou que, quando positiva, aumenta a prevalência de óbito em 17,4 vezes. Estudos revelam que a não realização do tratamento supervisionado aumenta em 9,3 vezes as chances para o óbito (LIEW et al., 2015), resultado endossado por

Reis-Santos et al. (2013), que trazem diminuição de 56% do óbito por TB quando há a realização do tratamento supervisionado.

Assim, percebe-se que o óbito por tuberculose, além de um problema de saúde pública, possui fatores que devem receber olhar especial durante o curso da doença, dentre eles: a idade, o estado imunológico e possível infecção por HIV e o acompanhamento do tratamento por meio das baciloscopias bimestrais e DOTS. Dessa forma, deve-se enfatizar políticas que garantam que todo paciente com TB tenha sua testagem para o HIV, assim como a realização de baciloscopias e PPD por pessoas que vivem com HIV, além de garantir acompanhamento adequado para pacientes coinfectados.

7.5 FATORES ASSOCIADOS À RESISTÊNCIA AO TRATAMENTO DA TUBERCULOSE

O caráter de resistência da doença se dá de forma cíclica, pois o indivíduo faz uso dos medicamentos e mata boa parte dos bacilos e, quando o tratamento é suspenso, há crescimento de novos bacilos. Assim, esse ciclo acaba favorecendo mutação na bactéria, gerando novas cepas resistente (GOMES et al., 2015).

No modelo logístico, identificou-se que ter ensino superior aumenta em 4,09 vezes a prevalência de resistência medicamentosa. Entretanto, os modelos de Poisson e binomial negativo não consideraram essa variável significativa no modelo e, então, ela foi retirada. Deve-se frisar que estudos apontam baixos níveis de escolaridade como aqueles que contribuem para formas resistentes da doença (VIANA, REDNER, RAMOS, 2018).

As formas de entrada diferentes de caso novo significativamente contribuem para a tuberculose apresentar forma resistente ao tratamento original. Identificou-se que o reingresso pós-abandono aumenta em 4,09 vezes a prevalência de resistência, a recidiva em 1,68 vezes e a transferência em 1,68 vezes. Estudo realizado no Espírito Santo identificou que a recidiva aumenta as chances de resistência em 7,72 vezes e o reingresso pós-abandono em 3,91 vezes (FREGONA et al., 2017).

Ademais, inquérito realizado em Porto Alegre identificou que o retratamento aumenta em 4,1 vezes a prevalência de formas resistentes (MICHELETTI et al., 2014). Também foi possível identificar que essas formas de entrada estão relacionadas ao óbito em pacientes multirresistentes (GAYOSO et al., 2018). Destaca-se, assim, a necessidade de fazer com que os pacientes tenham adesão ao tratamento inicial como forma de prevenção da doença resistente a medicamentos de primeira linha.

Em relação aos exames diagnósticos, identificou-se que a não realização do HIV como agente que diminui a prevalência de resistência medicamentosa (RP = 0,68). Não foi possível explicar o motivo de tal associação, por exemplo, estudo de Gaspar et al. (2016) não encontrou associação da resistência ao tratamento com a coinfeção TB-HIV. Pode-se imaginar que a testagem possa ser realizada somente em pessoas que entrem como caso novo ou que as pessoas que evoluam para TB-DR já tenham seu estado sorológico conhecido. Dentre todas as explicações, acredita-se que essa variável esteja que esteja causando confusão ou interação entre variáveis. De toda forma, enfatiza-se a importância que a coinfeção TB-HIV possui para a saúde pública e que a realização do exame é relevante.

Outro exame importante identificado como fundamental na epidemiologia da resistência foi a cultura de escarro, considerada padrão ouro na detecção do bacilo causador da doença. Evidenciou-se que o resultado positivo da cultura de escarro aumenta em 3,77 vezes a prevalência da TB-DR. Corroborando com esse achado, estudo de Fregona et al. (2017) identificou resultado 3,22 de chances de aumento da resistência.

Deve-se atentar para a importância da isoniazida no esquema RHZE de tratamento da TB, pois neste estudo identificou-se que seu uso reduz em 46% a prevalência de TB-DR no Ceará. Corroborando com esse achado, Nagu et al. (2017b) encontraram que pessoas com resistência à isoniazida possuem 18,7 vezes mais risco para desfechos de tratamento não favoráveis. Em relação ao esquema terapêutico, Rabahi, Silva Júnior e Conde (2017, p. 443) apontam que o etambutol, inserido no esquema RHZ, diminui as chances de “resistência adicional em pacientes com resistência inicial à isoniazida, isoladamente ou com resistência simultânea à isoniazida e pirazinamida ou com resistência isolada à rifampicina” mas

não age como agente protetor em ambos os casos (o que caracterizaria TB multirresistente -TBMR).

Também foi possível identificar nesse estudo que as baciloscopias bimestrais de acompanhamento são importantes preditores da prevalência de TB resistente, pois, se o paciente apresentar resultado positivo no seu 6º mês de tratamento há aumento de 38,24 vezes na prevalência de resistência medicamentosa. Ressalta-se, mais uma vez o caráter do acompanhamento do tratamento da doença, principalmente quanto à administração de medicação, para evitar o aparecimento de formas resistentes ao tratamento convencional.

Portanto, as formas resistentes da TB estão mais fortemente associadas ao retratamento da doença e ao resultado positivo da baciloscopia durante o acompanhamento do usuário. Dessa forma, medidas que façam a pessoa concluir seu tratamento de forma completa, fazendo uso diário de suas medicações, são medidas que devem ser fortalecidas para que haja quebra da linha de resistência bacteriana.

7.6 APLICAÇÃO DO MÉTODO CRISP-DM NA PREDIÇÃO DOS DESFECHOS DA TB

Por meio do método CRISP-DM foi possível desenvolver um sistema de classificação do desfecho da TB baseado em um perceptron multicamadas, ele mostrou acurácia de 74,21%, e apresentou área sob a curva de 0,83. Estudo de Laureano, Caetano e Cortez (2014) aplicou o método para prever a permanência dentro de um hospital por meio de algoritmo de florestas aleatórias, que conseguiu explicar 81% do desfecho.

A aplicação da técnica permitiu que fosse criado um modelo de regras, explicadas pela literatura, para a doença renal crônica, identificando-se quais eram as variáveis de interesse (TAHMASEBIAN et al., 2017). Outro conjunto de regras criadas foi visto no estudo de Kebebe, Zegeye e Zeleke (2017) por meio de algoritmo de florestas aleatórias, que se mostrou superior à rede neural artificial e à árvore J48. A árvore apresentou uma acurácia de 99,84% e mostrou-se superior a acurácia de 87,11% da rede neural e de 83,54% da J48.

Pérez et al. (2015) conseguiram usar a metodologia CRISP-DM para automatizar a preparação de dados para cálculos e predição dos coeficientes de mortalidade e letalidade por meio de base dados populacionais. Os autores apenas ressaltam que tal aplicação deve ser usada apenas em bases com população igual ou superior a 100 mil habitantes.

Estudo conseguiu identificar padrões de acidentes por causas externas com 82,34% de dados corretamente classificados e índice kappa de 0,67. Utilizou-se uma árvore de classificação e decisão J48 para tal feito (TIMARÁN-PEREIRA, CALDERÓN-ROMERO, HIDALGO-TROYA, 2017). Além desse estudo, pesquisa de Hernández e Colares (2012) também utilizou o algoritmo de árvores de classificação e decisão para identificar padrões relacionados a hipertensão arterial e conseguiu identificar três grupos distintos para a enfermidade. Os resultados apresentaram potencialidades de serem incluídos nos sistemas de saúde de Cuba.

Além disso, investigação realizada para prever mortalidade perioperatória por cirurgia de pulmão utilizou diferentes tipos de regressões logísticas para tal fim. Evidenciou-se que o melhor modelo preditivo foi a regressão que apresentava área sob a curva ROC de 0,817, sensibilidade de 9,68%, especificidade de 100%, valor preditivo positivo de 100% e valor preditivo negativo de 93,98% (RIVO et al., 2012).

Existem diversos estudos na área de saúde que já fazem uso de algoritmos de inteligência artificial a fim de prever eventos de interesse de suas respectivas áreas. Entretanto, poucos são aqueles que mostram a aplicação de um método tão detalhado de aquisição de conhecimento como o CRISP-DM. Todos os estudos que o utilizaram concordam que a aplicação de tal método favorece na organização dos dados, identificação de problemas durante a execução dos processos e como fazer para superá-los, além de detalhamento na exibição dos resultados obtidos.

8 CONCLUSÃO

Muitas foram as análises realizadas nesta pesquisa e diversas conclusões podem ser evidenciadas, pois o método CRISP-DM foi utilizado para investigar aspectos do padrão temporal, análise epidemiológica e aplicação de algoritmos de *machine learning* para predição dos desfechos da TB. Assim, as conclusões estão nos parágrafos subsequentes.

Em relação ao padrão temporal dos indicadores epidemiológicos da TB, pode-se identificar que sua incidência e prevalência apresentam-se decrescentes ao longo do período 2001-2017; entretanto, sua taxa de cura está continuamente caindo, ao passo que a mortalidade e resistência (até então pouco investigada) crescem significativamente ano após ano. Somente a taxa de abandono apresentou flutuações importantes ao longo do período, fazendo difícil a tarefa de se encontrar um padrão, mas destaca-se que suas taxas se encontraram elevadas.

Pela análise epidemiológica foi possível identificar que a cura do tratamento tem sua prevalência aumentada por fatores, como a escolaridade, zona de moradia, forma da doença, exame histopatológico e o uso de medicações, como a rifampicina, isoniazida e estreptomicina. Sua prevalência é diminuída pela idade, sexo masculino, raça/cor parda, entrada diferente de caso novo, fazer uso de etambutol e o acompanhamento por meio de resultados positivos das baciloscopias bimestrais.

Evidenciou-se que a prevalência do abandono é aumentada pelo sexo masculino, entradas como a recidiva e o reingresso pós-abandono, exames diagnósticos, como a não realização da radiografia e teste tuberculínico, HIV e histopatológico e resultados positivos das baciloscopias bimestrais de acompanhamento. A sua prevalência pode ser diminuída pela idade da pessoa, escolaridade a partir de ensino médio, moradia em zona rural, entrada como transferência e forma extrapulmonar.

Em relação ao óbito por TB foi possível constatar que somente o resultado não sugestivo de TB e fazer uso de isoniazida diminuem sua prevalência. Fatores como a idade, raça/cor parda, exame de radiografia, teste tuberculínico não realizado e HIV positivo, além da não realização das baciloscopias bimestrais podem aumentar sua prevalência.

Para formas resistentes ao tratamento, foi encontrado que formas de entrada diferente de caso novo, cultura de escarro positiva e baciloscopia de 6º mês positiva influenciam significativamente no aumento da prevalência. Seu decréscimo pode ser observado pelo uso de isoniazida e, controversamente, exame de HIV não realizado.

O modelo preditivo criado por meio do algoritmo de uma rede neural do tipo perceptron multicamadas pode corretamente prever 72,41% dos dados, apresentando melhor desempenho para a cura do tratamento. Esse comportamento se justifica, pois, tal desfecho é que correspondia a maior parte da base de dados.

Como limitação, aponta-se que a pesquisa foi realizada utilizando base de dados secundária, com grande quantidade de casos com incompletude de dados. O pesquisador identificou que houve mudança na ficha de notificação da TB no período 2014-2015 e, assim, muitos dados passaram a não ser mais preenchidos, ao passo que outros iniciaram seu preenchimento nesse período. Ao utilizar a base como um todo, pode-se identificar os dados ausentes e, por meio da aplicação do CRISP-DM, muitas dessas variáveis não foram consideradas na pesquisa.

Além disso, o uso de critérios diferentes dos epidemiológicos para a escolha de atributos pode ser uma alternativa para a criação de modelos preditivos mais precisos. De todo modo, por mais que se apresentem tais limitações, esse estudo não se torna inviável, pois a aplicação método permitiu a modelagem de dados de forma precisa e com resultados fidedignos.

Como desafios a validação do modelo preditivo, faz-se necessário destacar que, computacionalmente falando, a classe cura ser representada por mais de 70% dos casos pode desbalancear o modelo, levando-o a tendências de previsões por esse desfecho. Além disso, por mais que tenham sido selecionadas variáveis com menos de 25% de *missing values*, o modelo pode ter sido prejudicado por atributo com muitos dados ausentes.

Recomenda-se, enfim, que investigações futuras possam abordar tais limitações e melhorar o desempenho de seus modelos. Além disso, recomenda-se a aplicação desses algoritmos para pacientes com TB, para testar o comportamento do modelo em casos reais. Diante dos resultados, é recomendado o fortalecimento

da linha de cuidado a pessoas com TB, promovendo adesão ao tratamento e encaminhando os usuários à cura da doença.

Portanto, esta pesquisa se mostra relevante no contexto de cuidados clínicos em enfermagem e saúde, pois demonstra como a TB se comporta no Ceará, os fatores associados e o que pode ser feito para identificar os encerramentos do tratamento previamente. Assim, dentro do campo gerencial, políticas públicas para o enfrentamento da doença podem ser melhoradas, enfatizando os pontos aqui evidenciados e, dentro da linha de cuidado à pessoa com TB, o enfermeiro, além de outros profissionais de saúde, pode previamente intervir no cuidado ao paciente e mudar possíveis desfechos não favoráveis.

REFERÊNCIAS

- ABDELBARY, B E *et al.* Predicting treatment failure, death and drug resistance using a computed risk score among newly diagnosed TB patients in Tamaulipas, Mexico. **Epidemiol Infect.**, p.1-15, 2017.
- AGUIAR, Fábio S *et al.* Development of two artificial neural network models to support the diagnosis of pulmonary tuberculosis in hospitalized patients in Rio de Janeiro, Brazil. **Medical & Biological Engineering & Computing**, v. 54, n. 11, p. 1751-59, 2016.
- ALONSO, Susel Góngora *et al.* Data Mining Algorithms and Techniques in Mental Health: A Systematic Review. **J Med Syst.**, v. 42, ID 161, 2018.
- ANDRADE, Kaio Vinicius Freitas de *et al.* Effects of social protection on tuberculosis treatment outcomes in low or middle-income and in high-burden countries: systematic review and meta-analysis. **Cad Saúde Pública**, v. 34, n. 1, e00153116, 2018.
- ARAKAWA, Tiemi *et al.* Programa de controle da tuberculose no contexto municipal: avaliação de desempenho. **Rev Saúde Pública**, v. 51, ID 23, p. 1-9, 2017.
- ASHA, T.; NATARAJAN, S.; MURTHY, K. N. B. Effective classification algorithms to predict the accuracy of tuberculosis – a machine learning approach. **International Journal of Computer Sciences and Information Security**, v. 9, n. 7, p. 89-94, 2011.
- ASSOCIAÇÃO BRASILEIRA DE INTERNET DAS COISAS. **A diferença entre inteligência artificial, aprendizado de máquinas e aprendizagem profunda**. 2017. Disponível em <<http://abinc.org.br/a-diferenca-entre-inteligencia-artificial-aprendizado-de-maquinas-e-aprendizagem-profunda/>>. Acessado em 30 de novembro de 2017.
- AVENIR, A. J. E.; VILLOTA, J. D.; ROSAS, M. F. Machine learning: a means of diagnosing and prescribing treatments for tuberculosis patients. **International Journal of Computer Theory and Engineering**, v. 7, n. 3, p. 201-06, 2015.
- BAKKEN, Suzanne. *et al.* Integrating evidence into clinical information systems for nursing decision support. **International Journal of Medical Informatics**, v. 77, p. 413-420, 2008.
- BARREIRA, D. Os desafios para a eliminação da tuberculose no Brasil. **Epidemiol Serv Saúde**, v. 27, n. 1, e00100009, 2018.
- BASTA, Paulo Cesar *et al.* Desigualdades sociais e tuberculose: análise segundo raça/cor, Mato Grosso do Sul. **Rev Saúde Pública**, v. 47, n. 5, p. 854-864, 2013.

BRASIL. Ministério da Saúde. **Tratamento Diretamente Observado (TDO) da Tuberculose na Atenção Básica**: protocolo de enfermagem. Brasília: Ministério da Saúde, 2011.

_____. Ministério da Saúde. Conselho Nacional de Saúde. Departamento de Bioética. **Resolução Nº 466**. Diretrizes e normas regulamentadoras de pesquisas em seres vivos. Brasília: Ministério da Saúde, 2013.

_____. Ministério da Saúde. Boletim Epidemiológico. **Indicadores prioritários para o monitoramento do Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública no Brasil**. Brasília: Ministério da Saúde. 2017a.

_____. Ministério da Saúde. **Brasil livre da tuberculose**: plano nacional pelo fim da tuberculose como problema de saúde pública. Brasília: Ministério da Saúde. 2017b.

_____. Ministério da Saúde. **Tuberculose**: ficha de notificação/investigação, 2017c. Disponível em <http://portalsinan.saude.gov.br/images/documentos/Agravos/Tuberculose/Tuberculose_v5.pdf>. Acessado em 02 de dezembro de 2017.

_____. Ministério da Saúde. **Tuberculose**: ficha de acompanhamento, 2017d. Disponível em <http://portalsinan.saude.gov.br/images/documentos/Agravos/Tuberculose/Tuberculose_Acomp_v5.pdf>. Acessado em 02 de dezembro de 2017.

_____. Ministério da Saúde. Boletim Epidemiológico. **Experiências de programas de controle da tuberculose**: 'porque juntos iremos detectar, tratar e acabar com a tuberculose como problema de saúde pública no Brasil'. Brasília: Ministério da Saúde, 2018.

CAMPOS, Hisbello S. Etiopatogenia da tuberculose e formas clínicas. **Pulmão RJ**, v. 15, n. 1, p. 29-35, 2006.

CARDOSO, Mayara A *et al.* Tuberculosis treatment outcomes and factors associated with each of them in a cohort followed up between 2010 and 2014. **BioMed Research International**, v. 2017, ID 3974651, 2017.

CEARÁ. Secretaria Estadual de Saúde. **Boletim Epidemiológico – Tuberculose**. Fortaleza: Secretaria Estadual de Saúde, 2018.

CECCON, Roger Flores *et al.* Mortalidade por tuberculose nas capitais brasileiras, 2008-2010. **Epidemiol Serv Saúde**, v. 26, n. 3, p. 349-358. 2017.

CECILIO, Hellen Pollyana Mantelo *et al.* Tendência da mortalidade por tuberculose no estado do Paraná, Brasil – 1998 a 2012. **Ciência & Saúde Coletiva**, v. 23, n. 1, p. 241-248, 2018.

CHAVES, Emanuele Cordeiro *et al.* Aspectos epidemiológicos, clínicos e evolutivos da tuberculose em idosos de um hospital universitário em Belém, Pará. **Rev Bras Gerontol.**, v. 20, n. 1, p. 47-58, 2017.

CHIAVEGATTO FILHO, Alexandre Dias Porto. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, v. 24, n. 02, p. 235-332, 2015.

CHIRICO, Cristina *et al.* Tendencia de la tuberculosis en la región sanitaria V de la provincia de Buenos Aires, años 2000–2011. **Medicina (B Aires)**, v. 75, n. 3, p. 147-154, 2015.

CHIRINOS, N. E. C.; MEIRELLES, B. H. S.; BOUSFIELD, A. B. S. Representações sociais das pessoas com tuberculose sobre o abandono do tratamento. **Rev Gaucha Enferm.**, v. 36, esp, p. 207-2014, 2015.

CHURCHYARD, Gavin *et al.* What we know about tuberculosis transmission: an overview. **The Journal of Infectious Diseases**, sup. 6, p. 629-635, 2017.

COELI, Cláudia Medina. Sistemas de Informação em Saúde e o uso de dados secundários na pesquisa e avaliação em saúde. **Caderno de Saúde Coletiva**, v. 18, n. 3, p. 335-336, 2010.

COUTO, Davi Sarmiento *et al.* Fatores determinantes para o abandono do tratamento da tuberculose: representações dos usuários de um hospital público. **Saúde Debate**, v. 38, n. 102, p. 572-581, 2014.

DALAL, Alpa *et al.* Resistance Patterns among Multidrug-Resistant Tuberculosis patients in greater metropolitan Mumbai: trends over time. **Plos One**, v. 10, n. 1, e0116798, 2015.

DEMEMEWE, Z G *et al.* Trends in tuberculosis case notification and treatment outcomes after interventions in 10 zones of Ethiopia. **Int J Tuberc Lung Dis.**, v. 20, n. 9, p. 1192-1198, 2016.

DJOUMA, Fabrice Nembot *et al.* Determinants of death among tuberculosis patients in a semi urban diagnostic and treatment centre of Bafoussam, West Cameroon: a retrospective case-control study. **PanAfrican Medical Journal**, v. 22, ID 253, 2015.

DOSHI, Riddhi. *et al.* Tuberculosis control, and the where and why of artificial intelligence. **European Respiratory Journal**, v. 3, p. 1-5, 2017.

ER, O.; YUMUSAK, N.; TEMURTAS, F. Chest diseases diagnosis using artificial neural networks. **Expert systems with applications**, v. 37, p. 7648-55, 2010.

ESCADA, Rodrigo Otavio da Silva *et al.* Mortality in patients with HIV-1 and tuberculosis co-infection in Rio de Janeiro, Brazil - associated factors and causes of death. **BMC Infectious Diseases**, v. 17, ID 373, 2017.

FALZON, Dennis. *et al.* Digital health for the End TB Strategy: developing priority products and making them work. **European Respiratory Journal**, v. 48, p. 29-45, 2016.

FERREIRA, Anna Carolina Galvão *et al.* Desfechos clínicos do tratamento de tuberculose utilizando o esquema básico recomendado pelo Ministério da Saúde do Brasil com comprimidos em dose fixa combinada na região metropolitana de Goiânia. **J Bras Pneumol.**, v. 39, n. 1, p. 76-83, 2013.

FREGONA, Geisa *et al.* Fatores associados à tuberculose resistente no Espírito Santo, Brasil. **Rev Saúde Pública**, v. 51, ID 41, 2017.

GAIFER, Zield A. Risk factors for tuberculosis mortality in a tertiary care center in Oman, 2006–2016. **International Journal of Mycobacteriology**, v. 6, p. 356-359, 2017.

Gao, Z *et al.* A population-based study of tuberculosis case fatality in Canada: do aboriginal peoples fare less well? **Int J Tuberc Lung D.**, v. 19, n. 7, p. 772-779, 2015.

GASPAR, Rento Simões *et al.* Análise temporal dos casos notificados de tuberculose e de coinfeção tuberculose-HIV na população brasileira no período entre 2002 e 2012. **J Bras Pneumol.**, v. 42, n. 6, p. 416-422, 2016.

GAYOSO, Regina *et al.* Predictors of mortality in multidrug-resistant tuberculosis patients from Brazilian reference centers, 2005 to 2012. **Braz J Infect Dis.**, v. 22, n. 4, p. 305-310, 2018.

GOMES, Nathália Mota de Faria *et al.* Differences between Risk Factors Associated with Tuberculosis Treatment Abandonment and Mortality. **Pulmonary Medicine**, v. 2015, ID 546106, p. 1-8, 2015.

HARLING, Guy *et al.* Determinants of tuberculosis transmission and treatment abandonment in Fortaleza, Brazil. **BMC Public Health**, v. 17, ID 508, 2017.

HAYKIN, S. **Neural networks and learning machines**. 3. ed. New Jersey: Pearson, 2009.

HECK, M. A.; COSTA, J. S. D.; NUNES, M. F. Avaliação do Programa de Tuberculose em Sapucaia do Sul (RS): indicadores, 2000-2008. **Ciência & Saúde Coletiva**, v. 18, n. 2, p. 481-488, 2013.

HERNANDÉZ, F. D.; CORALES, Y. S. Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas. **Revista Cubana de Informática Médica**, v. 4, n. 2, p. 174-183, 2012.

KALHORI, S. R. N.; NASESHI, M.; ZENG, X. J. A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion. **International Journal of Applied Mathematics**, v. 40, n. 2, p. 1-6, 2010.

- KALHORI, S. R. N.; ZENG, X. J. Evaluation and comparison of different machine learning methods to predict outcome of tuberculosis treatment course. **Journal of Intelligent Learning Systems and Applications**, v. 5, p. 184-193, 2013.
- KATTAN, J. A.; SOSA, L. E.; LOBATO, M. N. Tuberculosis mortality: death from a curable disease, Connecticut, 2007–2009. **Int J Tuberc Lung D.**, v. 16, n. 12, p. 1657-1662, 2012.
- KEBEDE, M; ZEGEYE, D. T.; ZELEKE, B. M. Predicting CD4 count changes among patients on antiretroviral treatment: Application of data mining techniques. **Computer Methods and Programs in Biomedicine**, p. 149-157, 2017.
- LAKHANI, P., SUNDARAM, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. **Radiology**, v. 284, n. 2, p. 574-582, 2017.
- LAUREANO, R. M. S.; CAETANO, N.; CORTEZ, P. Previsão de tempos de internamento num hospital português: aplicação da metodologia CRISP-DM. **RISTI**, n. 3, p. 83-98, 2014.
- LIEW, S M *et al.* Tuberculosis in Malaysia: predictors of treatment outcomes in a national registry. **Int J Tuberc Lung D.**, v. 19, n. 7, p. 764-771, 2015.
- LIMA, Lílian Moura *et al.* Avaliação do acompanhamento e desfecho de casos de tuberculose em município do sul do Brasil. **Rev Gaúcha Enferm.**, v. 37, n. 1, e51467, 2016.
- LOPES, M. H. B. M.; HIGA, R. Desenvolvimento de um sistema especialista para identificação de diagnósticos de enfermagem relacionados com a eliminação urinária. **Revista Brasileira de Enfermagem**, v. 58, n. 1, p. 27-32, 2005.
- MAIMAITI, Rena *et al.* High prevalence and low cure rate of tuberculosis among patients with HIV in Xinjiang, China. **BMC Infectious Diseases**, v. 17. 2017.
- MARQUES, Isaac R *et al.* Guia de apoio a decisão de enfermagem obstétrica: aplicação da técnica de lógica fuzzy. **Revista Brasileira de Enfermagem**, v. 58, n. 3, p. 349-54, 2005.
- MENDES, Anapaula Martins *et al.* Situação epidemiológica da tuberculose no Rio Grande do Sul: uma análise com base nos dados do Sinan entre 2003 e 2012 com foco nos povos indígenas. **Rev Bras Epidemiol.**, v. 19, n. 3, p. 658-669, 2016.
- MICHELETTI, Vania Celina Dezoti *et al.* Tuberculose resistente em pacientes incluídos no II Inquérito Nacional de Resistência aos Fármacos Antituberculose realizado em Porto Alegre, Brasil. **J Bras Pneumol.**, v. 40, n. 2, p. 155-163, 2014.
- MODONGO, Chawangwa. *et al.* Amikacin concentrations predictive of ototoxicity and multidrug-resistant tuberculosis patients. **Antimicrobial Agents and Chemotherapy**, v. 59, n. 10, p. 6337-43, 2015.

MODONGO, Chawangwa. *et al.* Artificial intelligence and amikacin exposures predictive of outcomes in multidrug-resistant tuberculosis patients. **Antimicrobial Agents and Chemotherapy**, v. 60, n. 10, p. 5928-32, 2016.

NAGU, Tumaini J *et al.* Tuberculosis associated mortality in a prospective cohort in Sub Saharan Africa: Association with HIV and antiretroviral therapy. **International Journal of Infectious Diseases**, v. 56, p. 39-44, 2017a.

NAGU, Tumaini J *et al.* Effects of isoniazid resistance on TB treatment outcomes under programmatic conditions in a high-TB and -HIV setting: a prospective multicentre study. **J Antimicrob Chemother.**, v. 72, p. 876-881, 2017b.

OLIVEIRA FILHO, P. F. **Epidemiologia e Bioestatística: fundamentos para a leitura crítica.** Rio de Janeiro: Rubio, 2015.

Oliveira, R. A.; LEFVRE, F. Comunicação na revelação do diagnóstico e adesão ao tratamento da tuberculose: representações sociais de profissionais e de pacientes. **Texto Contexto Enferm.**, v. 26, n. 2, e06790015, 2017.

PALHA, Pedro Fredemir *et al.* Acesso aos serviços de atenção à tuberculose: análise da satisfação dos doentes. **Rev Esc Enferm USP**, v. 48, n. 2, p. 342-348, 2012.

PASIPANODYA, Jotam G *et al.* Serum drug concentrations predictive of pulmonary tuberculosis outcomes. **Journal of Infectious Diseases**, v. 208, n. 9, p. 1464-1473, 2013.

PAZ, Letícia Nazareth Fernandes da *et al.* Efetividade do tratamento da tuberculose. **J Bras Pneumol.**, v. 38, n. 4, p. 503-510, 2012.

PÉREZ, Joaquim *et al.* A data preparation methodology in data mining applied to mortality population databases. **J Med Syst.**, v. 39, p. 1-6, 2015.

PINTO, Mayrla Lima *et al.* Ocorrência de casos de tuberculose em Crato, Ceará, no período de 2002 a 2011: uma análise espacial de padrões pontuais. **Rev Bras Epidemiol.**, v. 18, n. 2, p. 313-325, 2015.

PINTO, Priscila Fernanda Porto Scaff *et al.* Perfil epidemiológico da tuberculose no município de São Paulo de 2006 a 2013. **Rev Bras Epidemiol.**, v. 20, n. 3, p. 549-557, 2017.

PIRES NETO, Roberto da Justa *et al.* Características clínico-epidemiológicas de pacientes com coinfeção HIV/tuberculose acompanhados nos serviços de referência para HIV/AIDS em Fortaleza, Ceará, entre 2004 e 2008. **Cad Saúde Colet.**, v. 20, n. 2, p. 244-249, 2012.

PRADO JUNIOR, J. C.; VIRGILIO, T. C.; MEDRONHO, R. A. comparação da proporção de cura por tuberculose segundo cobertura e tempo de implantação de Saúde da Família e fatores socioeconômicos e demográficos no município do rio de

Janeiro, Brasil, em 2012. **Ciência & Saúde Coletiva**, v. 21, n. 5, p. 1491- 1498, 2016.

PUSCH, Tobias. *et al.* Therapy duration and long-term outcomes in extra-pulmonary tuberculosis. **BMC Infectious Diseases**, v. 14, n. 115, p. 1-10, 2014.

RABAHI, M. F.; SILVA JUNIOR, J. L. R.; CONDE, M. B. Avaliação do impacto das mudanças do tratamento da tuberculose implantadas em 2009 no controle da tuberculose pulmonar no Brasil. **J Bras Pneumol.**, v. 43, n. 5, p. 437-444, 2017.

RANDELL, Rebeca. *et al.* Effects of computerized decision support systems on nursing performance and patient outcomes: a systematic review. **Journal of Health Services & Policy**, v. 12, n. 4, p. 242-49, 2007.

RASCHKA, S. **Python Machine learning**. Mumbai: Packt Publishing, 2015.

REIS, Dener Carlos dos *et al.* Perfil epidemiológico da tuberculose no Município de Belo Horizonte (MG), no período de 2002 a 2008. **Rev Bras Epidemiol.**, v. 16, n. 3, p. 592-602, 2013.

RIVO, Eduardo *et al.* Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management. **Clin Transl Oncol.**, v. 14, p. 73-79, 2012.

RODRIGUES, Débora César de Souza *et al.* O discurso de pessoas acometidas por tuberculose sobre a adesão ao tratamento. **Ciência y Enfermería**, v. 23, n. 1, p. 67-76, 2017.

RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. 3. ed. New Jersey: Pearson, 2010.

SALVADORI, Nicolas *et al.* Incidence of tuberculosis and associated mortality in a cohort of human immunodeficiency virus-infected children initiating antiretroviral therapy. **Journal of the Pediatric Infectious Diseases Society**, v. 6, p. 161-167, 2017.

SANTOS, Alcione Miranda *et al.* Árvore de classificação e redes neurais artificiais: uma aplicação à predição de tuberculose pulmonar. In: PROCEEDINGS OF THE VI BRAZILIAN CONFERENCE ON NEURAL NETWORKS; VI CONGRESSO BRASILEIRO DE REDES NEURAIAS, 12., 2003. **Anais...**, 2003. p. 427-432.

SANTOS, Janine Nascimento *et al.* Fatores associados à cura no tratamento da tuberculose no estado do Rio de Janeiro, 2011-2014. **Epidemiol Serv Saúde**, v. 27, n. 3, e2017464, 2018.

SANTOS, T. A.; MARTINS, M. M. F. Perfil dos casos de reingresso após abandono do tratamento da tuberculose em Salvador, Bahia, Brasil. **Cad Saúde Colet.**, v. 26, n. 3, p. 233-240, 2018.

SANTOS-REIS, Barbara *et al.* Desfecho do tratamento da tuberculose em indivíduos com doença renal crônica no Brasil: uma análise multinomial. **J Bras Pneumol.**, v. 39, n. 5 p. 585-594, 2013.

SAYBANI, Mahmoud Reza *et al.* Diagnosing tuberculosis with a novel support vector machine-based artificial immune recognition system. **Iranian Red Crescent Medical Journal**, v. 17, n. 4, p. 1-8, 2015.

SEIXAS, J M *et al.* Artificial neural network models to support the diagnosis of pleural tuberculosis in adult patients. **International Journal of Lung Diseases**, v. 17, n. 3, p. 682-686, 2013.

SELLMER, Danielle. *et al.* Sistema especialista para apoiar a decisão na terapia tópica de úlceras venosas. **Revista Gaúcha de Enfermagem**, v. 34, n. 2, p. 154-162, 2013.

SILVA, C. C. A. V.; ANDRADE, M. S., CARDOSO, M. D. Fatores associados ao abandono do tratamento de tuberculose em indivíduos acompanhados em unidades de saúde de referência na cidade do Recife, Estado de Pernambuco, Brasil, entre 2005 e 2010. **Epidemiol Serv Saúde**, v. 22, n. 1, p. 77-85, 2013.

SILVA, E. A.; ANJOS, U. U.; NOGUEIRA, J. A. Modelo preditivo do abandono do tratamento da tuberculose. **Saúde Debate**, v. 38, n.10, p. 200-209, 2014.

SILVA, P. F.; MOURA, G. S.; GALDAS, A. J. M. Fatores associados ao abandono do tratamento da tuberculose pulmonar no Maranhão, Brasil, no período de 2001 a 2010. **Cad Saúde Pública**, v. 30, n. 8, p. 1745-1754, 2014.

SILVA, Rebeca N *et al.* Non-invasive method to analyse the risk of developing diabetic foot. **Healthcare Technology Letters**, v. 1, n. 4, p. 109-113, 2014.

SILVA, V. D.; MELLO, F. C. Q.; FIGUEREDO, S. C. A. Estimativas das taxas de recidiva, cura e abandono de tratamento em pacientes com tuberculose pulmonar tratados com esquema de dose fixa combinada de quatro drogas em uma unidade de saúde terciária na cidade do Rio de Janeiro. **J Bras Pneumol.**, v. 43, n. 2, p. 113-120, 2017.

SNYDER, Robert E. *et al.* Risk factors for differential outcome following directly observed treatment (DOT) of slum and non-slum tuberculosis patients: a retrospective cohort study. **BMC Infectious Diseases**, v. 16, 494, p; 1-10, 2016.

SOARES, Marcelo Luiz Medeiros *et al.* Aspectos sociodemográficos e clínico-epidemiológicos do abandono do tratamento de tuberculose em Pernambuco, Brasil, 2001-2014. **Epidemiol Serv Saude**, v. 26, n. 3, p.369-378, 2017.

SOUZA FILHO, João B de O *et al.* A screening system for smear-negative pulmonary tuberculosis using artificial neural networks. **International Journal of Infectious Diseases**, v. 49, p. 33-39, 2016.

TAHMASEBIAN, Shahram *et al.* Applying data mining techniques to determine important parameters in chronic kidney disease and the relations of these parameters to each other. **J Renal Inj Prev.**, v. 6, n. 2, p. 83-87, 2017.

TEIXEIRA, Maria João Cardoso *et al.* Os contributos dos sistemas de apoio à tomada de decisão para a prática de enfermagem. **Journal of Health Informatics**, v. 4, n. 2, pp. 59-63, 2012.

TELAROLLI JUNIOR, R.; LOFFREDO, L. C. M.; GASPARETTO, R. M. Clinical and epidemiological profile of tuberculosis in an urban area with high human development index in southeastern Brazil. Time series study. **São Paulo Med J.**, v. 135, n. 5, p. 413-419, 2017.

TIMARÁN-PEREIRA, R; CALDERÓN-ROMERO, A; HIDALGO-TROYA, A. Aplicación de los árboles de decisión en la identificación de patrones de lesiones fatales por causa externa en el municipio de Pasto, Colombia. **Univ Salud**, v. 19, n. 3, p. 359-365, 2017.

VÉLEZ, L. V; MONTOYA, M. P. A. Gender differences in the interpretation of experiences of patients with tuberculosis in Medellín, Colombia. **Invest Educ Enferm.**, v. 33, n. 2, p. 217-226, 2015.

VIANA, P. V. S.; REDNER, P.; RAMOS, J. P. Fatores associados ao abandono e ao óbito de casos de tuberculose drogarresistente (TBDR) atendidos em um centro de referência no Rio de Janeiro, Brasil. **Cad Saúde Pública**, v. 34, n. 5, e00048217, 2018.

VILELA, Pamela Nogueira Silva *et al.* Burden of tuberculosis trends in a Brazilian southern state. **Rev Bras Epidemiol.**, v. 21, e180013, 2018.

World Health Organization. **Global Tuberculosis Report 2014: supplement Drug-Resistant TB Surveillance & Response.** Genebra: OMS, 2014.

WHO. **Global Tuberculosis Report 2018.** Genebra: OMS. 2018.

ANEXOS

ANEXO A – FICHA DE NOTIFICAÇÃO DA TB

SINAN
SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO

República Federativa do Brasil
Ministério da Saúde

Nº

FICHA DE NOTIFICAÇÃO / INVESTIGAÇÃO TUBERCULOSE

CRITÉRIO LABORATORIAL - é todo caso que, independentemente da forma clínica, apresenta pelo menos uma amostra positiva de baciloscopia, ou de cultura, ou de teste rápido molecular para tuberculose.
CRITÉRIO CLÍNICO-EPIDEMIOLÓGICO - é todo caso que não preenche o critério de confirmação laboratorial acima descrito, mas que recebeu o diagnóstico de tuberculose ativa. Essa definição leva em consideração dados clínico-epidemiológicos associados à avaliação de outros exames complementares (como os de imagem, histológicos, entre outros).

Dados Gerais	1 Tipo de Notificação 2 - Individual		2 Agravo/doença TUBERCULOSE		Código (CID10) A16.9	3 Data da Notificação
	4 UF	5 Município de Notificação			Código (IBGE)	
	6 Unidade de Saúde (ou outra fonte notificadora)			Código	7 Data do Diagnóstico	
	8 Nome do Paciente					9 Data de Nascimento
Notificação Individual	10 (ou) Idade 1 - Hora 2 - Dia 3 - Mês 4 - Ano		11 Sexo M - Masculino F - Feminino I - Ignorado	12 Gestante 1-1º Trimestre 2-2º Trimestre 3-3º Trimestre 4 - Idade gestacional Ignorada 5-Não 6- Não se aplica 9- Ignorado		13 Raça/Cor 1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9- Ignorado
	14 Escolaridade 0-Analfabeto 1-1ª a 4ª série incompleta do EF (antigo primário ou 1º grau) 2-4ª série completa do EF (antigo primário ou 1º grau) 3-5ª à 8ª série incompleta do EF (antigo ginásio ou 1º grau) 4-Ensino fundamental completo (antigo ginásio ou 1º grau) 5-Ensino médio incompleto (antigo colegial ou 2º grau) 6-Ensino médio completo (antigo colegial ou 2º grau) 7-Educação superior incompleta 8-Educação superior completa 9-Ignorado 10- Não se aplica					
	15 Número do Cartão SUS			16 Nome da mãe		
	17 UF		18 Município de Residência		Código (IBGE)	19 Distrito
Dados de Residência	20 Bairro		21 Logradouro (rua, avenida,...)		Código	
	22 Número	23 Complemento (apto., casa, ...)			24 Geo campo 1	
	25 Geo campo 2		26 Ponto de Referência			27 CEP
	28 (DDD) Telefone		29 Zona 1 - Urbana 2 - Rural 3 - Periurbana 9 - Ignorado		30 País (se residente fora do Brasil)	
	Dados Complementares do Caso					
	31 Nº do Prontuário		32 Tipo de Entrada 1 - Caso Novo 2 - Recidiva 3 - Reingresso Após Abandono 4 - Não Sabe 5 - Transferência 6 - Pós-óbito			
33 Populações Especiais 1 - Sim 2 - Não 9 - Ignorado			População Privada de Liberdade População em Situação de Rua		34 Beneficiário de programa de transferência de renda do governo 1 - Sim 2 - Não 9 - Ignorado	
35 Forma 1 - Pulmonar 2 - Extrapulmonar 3 - Pulmonar + Extrapulmonar		36 Se Extrapulmonar 1 - Pleural 2 - Gang. Perif. 3 - Geniturinária 4 - Óssea 5 - Ocular 6 - Miliar 7 - Meningoencefálico 8 - Cutânea 9 - Laringea 10 - Outra				
37 Doenças e Agravos Associados 1 - Sim 2 - Não 9 - Ignorado						
38 Baciloscopia de Escarro (diagnóstico) 1 - Positiva 2 - Negativa 3 - Não Realizada 4 - Não se aplica		39 Radiografia do Tórax 1 - Suspeito 2 - Normal 3 - Outra Patologia 4 - - Não Realizado		40 HIV 1 - Positivo 3 - Em Andamento 2 - Negativo 4 - Não Realizado		
41 Terapia Antirretroviral Durante o Tratamento para a TB 1 - Sim 2 - Não 9 - Ignorado			42 Histopatologia 1 - Baar Positivo 2 - Sugestivo de TB 3 - Não Sugestivo de TB 4 - Em Andamento 5 - Não Realizado			
43 Cultura 1 - Positivo 2 - Negativo 3 - Em Andamento 4 - Não Realizado		44 Teste Molecular Rápido TB (TMR-TB) 1 - Detectável sensível à Rifampicina 2 - Detectável Resistente à Rifampicina 3 - Não Detectável 4 - Inconclusivo 5 - Não Realizado		45 Teste de Sensibilidade 1 - Resistente somente à Isoniazida 2 - Resistente somente à Rifampicina 3 - Resistente à Isoniazida e Rifampicina 4 - Resistente a outras drogas de 1ª linha 5 - Sensível 6 - Em andamento 7 - Não realizado		
46 Data de Início do Tratamento Atual			47 Total de Contatos Identificados			
Município/Unidade de Saúde					Cód. da Unid. de Saúde	
Nome		Função		Assinatura		
Tuberculose		Sinan NET		SVS 02/10/2014		

ANEXO B – TELA DE ACOMPANHAMENTO TB

TELA DE ACOMPANHAMENTO DE TUBERCULOSE

48 UF	49 Município de Notificação Atual	Código (IBGE)	50 N° Notificação Atual
51 Data da Notificação Atual	52 Unidade de Saúde Atual	Código	
53 UF	54 Município de Residência Atual	Código (IBGE)	55 CEP
56 Distrito de Residência Atual	57 Bairro de Residência Atual		
58 Baciloscopias de acompanhamento (escarro) 1 - Positivo 2 - Negativo 3 - Não Realizado 4 - Não se aplica			
<input type="checkbox"/> 1º mês <input type="checkbox"/> 2º mês <input type="checkbox"/> 3º mês <input type="checkbox"/> 4º mês <input type="checkbox"/> 5º mês <input type="checkbox"/> 6º mês <input type="checkbox"/> Após 6º mês			
59 Número do prontuário atual	60 Tratamento Diretamente Observado (TDO) realizado 1 - Sim 2 - Não 9 - Ignorado	61 Total de contatos examinados	
62 Situação de Encerramento 1 - Cura 2 - Abandono 3 - Óbito por TB 4 - Óbito por outras causas 5 - Transferência 6 - Mudança de Diagnóstico 7 - TB-DR 8 - Mudança de esquema 9 - Falência 10 - Abandono Primário			
63 Se transferência 1 - Mesmo município 2 - Município diferente (mesma UF) 3 - UF diferente 4 - País diferente 9 - Ignorado			
64 UF de transferência	65 Município de transferência	66 Data de Encerramento	

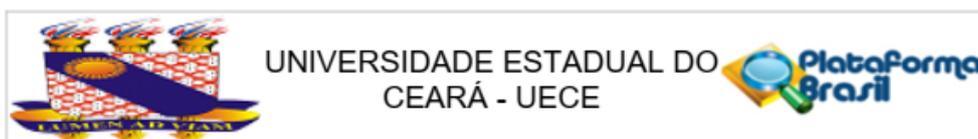
Tuberculose

Sinan NET

SVS 09/01/2014

Fonte: SINAN-net – Ministério da Saúde (BRASIL, 2017d)

ANEXO C – PARECER CONSUBSTANCIADO DO CEP



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Aplicação de Algoritmos de Inteligência Artificial na Predição do Desfecho do Tratamento da Tuberculose

Pesquisador: GEORGE JO BEZERRA SOUSA

Área Temática:

Versão: 1

CAAE: 88396718.1.0000.5534

Instituição Proponente: PROGRAMA DE PÓS-GRADUAÇÃO EM ENFERMAGEM E SAÚDE

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.687.046

Apresentação do Projeto:

Aplicação de Algoritmos de Inteligência Artificial na Predição do Desfecho do Tratamento da Tuberculose

Objetivo da Pesquisa:

aplicar algoritmos de inteligência artificial para prever os desfechos da tuberculose

Avaliação dos Riscos e Benefícios:

O estudo traz riscos mínimos aos participantes, uma vez que os dados já foram coletados e o pesquisador utilizará somente o banco de dados que contém essas informações.

Comentários e Considerações sobre a Pesquisa:

Pesquisa relevante.

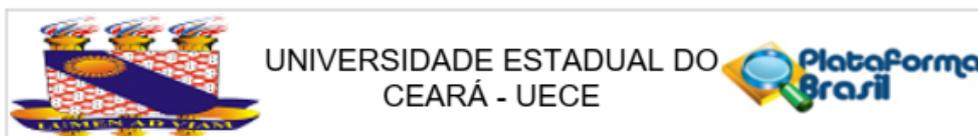
Considerações sobre os Termos de apresentação obrigatória:

Possui termo de autorização do responsável pelo setor onde a pesquisa será realizada. Orçamento Presente. Folha de rosto presente e assinada pelo chefe do departamento. Possui termo de fiel depositário. Cronograma presente e de acordo com as normativas. Solicitação de isenção de TCLE uma vez que se trata de um estudo retrospectivo com dados secundários.

Conclusões ou Pendências e Lista de Inadequações:

Projeto bem escrito. Atende todas as normativas éticas.

Endereço: Av. Silas Munguba, 1700
 Bairro: Itaperi CEP: 60.714-903
 UF: CE Município: FORTALEZA
 Telefone: (85)3101-9890 Fax: (85)3101-9906 E-mail: cep@uece.br



Continuação do Parecer: 2.687.046

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1090489.pdf	24/04/2018 15:11:39		Aceito
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1090489.pdf	24/04/2018 15:06:13		Aceito
Outros	anuencia.pdf	24/04/2018 14:53:02	GEORGE JO BEZERRA SOUSA	Aceito
Projeto Detalhado / Brochura Investigador	Projeto.docx	16/04/2018 10:57:57	GEORGE JO BEZERRA SOUSA	Aceito
Cronograma	CRONOGRAMA.pdf	16/04/2018 10:55:23	GEORGE JO BEZERRA SOUSA	Aceito
Outros	fiel_depositario.pdf	16/04/2018 09:53:44	GEORGE JO BEZERRA SOUSA	Aceito
Folha de Rosto	folha_rosto.pdf	16/04/2018 09:51:27	GEORGE JO BEZERRA SOUSA	Aceito
Declaração de Instituição e Infraestrutura	termo_do_setor.pdf	16/04/2018 09:51:14	GEORGE JO BEZERRA SOUSA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Ausencia_TCLE.doc	12/03/2018 13:46:27	GEORGE JO BEZERRA SOUSA	Aceito
Orçamento	orcamento.pdf	12/03/2018 13:19:35	GEORGE JO BEZERRA SOUSA	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

FORTALEZA, 01 de Junho de 2018

Assinado por:
ISAAC NETO GOES DA SILVA
(Coordenador)

Endereço: Av. Silas Munguba, 1700
Bairro: Itaperi CEP: 60.714-903
UF: CE Município: FORTALEZA
Telefone: (85)3101-9890 Fax: (85)3101-9906 E-mail: cep@uece.br