

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Escola de Veterinária

Programa de Pós-Graduação em Ciência Animal

Isabela de Lourdes Araújo

**Comparação da performance de algoritmos de aprendizado de máquina para análise
preditiva de febre amarela no estado de Minas Gerais**

Belo Horizonte

2023

Isabela de Lourdes Araújo

Comparação da performance de algoritmos de aprendizado de máquina para análise preditiva de febre amarela no estado de Minas Gerais

Tese apresentada à Universidade Federal de Minas Gerais, Escola de Veterinária, como requisito parcial para obtenção de grau de Doutor em Ciência Animal

Área de concentração: Epidemiologia

Orientador: Prof. Dr. Marcos Xavier Silva

Co-orientadores: Prof. Dr. Frederico Gualberto Ferreira Coelho e Profa. Dra. Soraia de Araújo Diniz

Belo Horizonte

2023

A663c Araújo, Isabela de Lourdes, 1982 -
Comparação da performance de algoritmos de aprendizado de máquina para análise preditiva de febre amarela no estado de Minas Gerais / Isabela de Lourdes Araújo . – 2023.
139 f. il

Orientador: Marcos Xavier Silva
Coorientadores: Frederico Gualberto Ferreira Coelho
Soraia de Araújo Diniz

Tese (Doutorado) apresentada à Escola de Veterinária da Universidade Federal de Minas Gerais para obtenção do título de Doutora.
Área de concentração: Epidemiologia
Bibliografias: f. 109 a 128
Anexos: f. 129 a 139.

1. Animais - Doenças - Teses - 2. Epidemiologia - Teses - 3. Saúde pública - Teses -
I. Silva, Marcos Xavier - II. Coelho, Frederico Gualberto Ferreira - III. Diniz, Soraia de Araújo
IV. Universidade Federal de Minas Gerais, Escola de Veterinária - V. Título.

CDD – 636.089

Bibliotecária responsável Cristiane Patrícia Gomes – CRB2569
Biblioteca da Escola de Veterinária, Universidade Federal de Minas Gerais.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE VETERINÁRIA
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

FOLHA DE APROVAÇÃO

ISABELA DE LOURDES ARAUJO

Tese submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA ANIMAL, como requisito para obtenção do grau de DOUTOR em CIÊNCIA ANIMAL, área de concentração EPIDEMIOLOGIA.

Aprovada em 16 de fevereiro de 2023, pela banca constituída pelos membros:

Prof. Dr. Marcos Xavier Silva - Orientador

Prof. Dr. Nelson Rodrigo da Silva Martins

Dr. Tiago Mendonça de Oliveira

Prof. Dr. David Soeiro Barbosa

Prof. Dr. Luiz Carlos Bambirra Torres



Documento assinado eletronicamente por Marcos Xavier Silva, Professor do Magistério Superior, em 23/02/2023, às 16:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Nelson Rodrigo da Silva Martins, Professor do Magistério Superior, em 26/02/2023, às 09:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por David Soeiro Barbosa, Professor do Magistério Superior, em 27/02/2023, às 12:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Luiz Carlos Bambirra Torres, Usuário Externo, em 27/02/2023, às 16:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Tiago Mendonça de Oliveira, Usuário Externo, em 03/03/2023, às 18:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 2053031 e o código CRC 9F94C948.

Dedico esse trabalho à minha família: minhas filhas Luiza e Júlia, meu esposo Paulo, minha mãe Arlete, meu pai Araújo e meus irmãos Diogo e Marcelo, por serem luz e inspiração nessa caminhada.

AGRADECIMENTOS

Agradeço inicialmente à minha família por estarem ao meu lado me apoiando nos momentos difíceis dessa jornada. Passar por uma pandemia durante o curso de doutorado foi assustador e nos fez repensar a vida. Pesar por tantas vidas perdidas e gratidão por ter sido permitido continuar a caminhada nessa jornada.

Agradeço ao meu esposo Paulo por acreditar tanto em mim e me apoiar, ouvir e me incentivar a me supurar um pouquinho a cada dia. À Juju e Luiza, minhas maiores preciosidades por me incentivarem a ir mais longe, sempre, e ser forte.

Agradeço ao novo amigo que a UFMG me deu: Professor Frederico Gualberto F. Coelho, que sempre disposto a dividir o saber, me ajudou em todos os momentos a superar os desafios com alegria e determinação.

Agradeço aos amigos da pós-graduação que dividiram comigo esse momento tão difícil que vivemos: Tiago e Mariana Frankó.

Agradeço ao professor Marcos Xavier por todo apoio e incentivo nessa segunda jornada em busca de novos aprendizados.

Agradeço à professora Soraia Diniz por ser sempre tão assertiva e disposta a ajudar.

Agradeço aos meus filhos pets por serem sempre consolo e amor nas horas mais importantes: Amora, Janis, Miliante, Chiquinho e Lelê.

Agradeço ao universo por ser tão cuidadoso, gentil e acolhedor comigo. Sou muito grata pelas dificuldades, pelas provas e por todos que me cercam.

“ O que a vida quer da gente é coragem”

Guimarães Rosa

Resumo

A febre amarela é uma arbovirose de caráter zoonótico, de extrema importância para a saúde pública no Brasil. A ocorrência de casos de febre amarela tem incidido entre os meses de dezembro e maio, período sazonal da doença, ocorrendo em regiões de mata, quando o vírus encontra condições favoráveis para a transmissão. O desenvolvimento de modelos preditivos desempenha um papel importante na interpretação de elementos complexos do mundo real. Com o aumento acentuado na quantidade, variedade e velocidade de produção de dados, novos desafios como extrair recursos e reconhecer padrões latentes, surgiram, exigindo novas abordagens e soluções eficazes. Otimizar modelos de aprendizado de máquina que possam prever com boa performance a ocorrência de febre amarela na população humana em Minas Gerais/Brasil pode apoiar decisões e ações em saúde pública de forma a reduzir a morbidade e a mortalidade da doença. Objetivou-se utilizar técnicas supervisionadas de aprendizado de máquina e comparar as performances em problema de regressão para prever resposta de interesse para a saúde pública, no caso, a ocorrência de febre amarela humana em Minas Gerais. O projeto está inserido na linha de pesquisa: avaliação e implementação de políticas e estratégias em saúde pública e ambiental. As variáveis meteorológicas utilizadas no presente trabalho foram extraídas do banco de dados do Instituto Nacional de Meteorologia (INMET), dados de casos confirmados de febre amarela de epizootias em primatas não humanos e seres humanos oriundos do Sistema de Informação de Agravos de Notificação (SINAN) e cobertura vacinal de febre amarela oriundos do Sistema de Informação do Programa Nacional de imunizações (SI-PNI) foram obtidos junto à Secretaria de Estado de Saúde de Minas Gerais. Os resultados e discussão estão organizados em três artigos científicos. O primeiro trata-se de uma revisão integrativa da literatura científica, realizada para nortear a seleção inicial das variáveis utilizadas na pesquisa e envolveu temas relacionados à eco-epidemiologia da febre amarela na região Sudeste do Brasil. O segundo artigo trata-se de um estudo epidemiológico, retrospectivo e descritivo dos bancos de dados do Sistema de Informação de Agravos de Notificação (SINAN) contendo informações sobre casos de febre amarela humana e epizootias em primatas não humanos entre 2007 e 2020 registrados em Minas Gerais. O terceiro artigo objetivou prever casos de febre amarela humana em Minas Gerais e comparar a performance dos modelos. Quatro modelos de aprendizado de máquina foram utilizados: Random forest, Xgboost, SVM e rede neural MLP. Os resultados mostraram que random forest obteve o melhor desempenho comparativo apresentando MSE= 0,22; MAE= 0,07; RMSE= 0,47; e R²= 77,68% com pouca diferença de performance para a rede MLP e Xgboost. Os modelos de previsão da febre amarela utilizando algoritmos de aprendizado de máquina com aprendizado

supervisionado apresentaram ser boas ferramentas para previsão de série temporal multivariada. Técnicas de modelagem algorítmicas estão sendo constantemente aprimoradas e quase não há justificativas para se utilizar um ou outro algoritmo, sendo indicado a comparação entre performances em problemas de previsão.

Palavras-chave: febre amarela, epizootias em primatas não humanos, aprendizado de máquina.

Abstract

Yellow fever is a zoonotic arbovirus of extreme importance to public health in Brazil. Seasonal period of the disease occurring in forest regions, when the virus encounters favorable conditions for transmission. The development of predictive models plays an important role in the interpretation of complex elements of the real world. With the sharp increase in the amount, variety and speed of data production, new challenges such as extracting features and recognizing latent patterns have emerged, requiring new approaches and effective solutions. Optimizing machine learning models that can predict the occurrence of yellow fever in the human population in Minas Gerais/Brazil with good performance can support decisions and actions in public health in order to reduce the morbidity and mortality of the disease. The objective was to use supervised machine learning techniques and compare performances in a regression problem to predict a response of interest to public health, in this case, the occurrence of human yellow fever in Minas Gerais. The project is part of the research line evaluation and implementation of policies and strategies in public and environmental health with coverage in machine learning models. The meteorological variables used in the present work were extracted from the National Institute of Meteorology database. Data from confirmed cases of yellow fever from epizootics in non-human primates and humans were obtained from The Notifiable Diseases Information System and vaccination coverage from Information System National vaccination program through the Secretary of Health State of Minas Gerais. The results and discussion were organized in three scientific articles. The first is an integrative review of the scientific literature, carried out to guide the initial selection of variables used in the research and involved topics related to the eco-epidemiology of yellow fever in the Southeast region of Brazil. The second article is an epidemiological, retrospective and descriptive work of the databases of the Notifiable Diseases Information System (SINAN) containing information on cases of human yellow fever and epizootics in non-human primates between 2007 and 2020 registered in Minas Gerais. The third article aimed to predict cases of human yellow fever in Minas Gerais and compare the performance of the models. Four machine learning models were used: Random forest, Xgboost, SVM and MLP neural network. The results showed that the random forest had the best comparative performance, presenting $MSE= 0.22$, $MAE= 0.07$, $RMSE= 0.47$ and $R^2= 77.68\%$ with little difference in performance for the MLP and Xgboost. Supervised models showed to be good tools for forecasting multivariate time series. Algorithmic modeling techniques are being constantly improved and there are almost no justifications for using one or another algorithm, being indicated the comparison between performances in prediction problems.

Key-words: Yellow fever, epizootics non-human primate, machine learning.

LISTA DE FIGURAS

Figura 1. Provável rota da disseminação da onda epizootico-epidêmica de febre amarela no Brasil – 1932-1942 -----	22
Figura 2. Casos de febre amarela silvestre, confirmados laboratorialmente, de 1932 a 1967 no Brasil-----	24
Figura 3. Os ciclos da febre Amarela no Brasil -----	26
Figura 4. Dispersão do <i>Aedes aegypti</i> no continente americano ao longo do século XX-----	27
Figura 5. Série histórica do número de casos humanos confirmados para febre amarela e a taxa de letalidade, segundo o ano de início dos sintomas no Brasil, de 1980 a 26ª semana epidemiológica de 2018-----	30
Figura 6. Evolução geográfica do risco de transmissão de febre amarela no Brasil, de 2001 a 2010-----	31
Figura 7. Reemergência da febre amarela no Brasil. Distribuição dos municípios com casos humanos e/ou epizootias em PNH confirmados durante reemergência extra-Amazônica da febre amarela, de acordo com o período de monitoramento, julho de 2014 a junho de 2020, Brasil -- -----	32
Figura 8: Esquema representativo do aprendizado supervisionado -----	42
Figura 9: Esquema ilustrativo do aprendizado não supervisionado-----	43
Figura 10: Esquema representativo das fases que compreendem o uso de algoritmos de aprendizado de máquina. -----	45
Figura 11: Representação gráfica de <i>Underfitting</i> , <i>Good Fit</i> e <i>Overfitting</i> -----	50
Figura 12: Diagrama que ilustra o conceito geral de perceptron-----	53
Figura 13: Arquitetura de uma rede neural MLP-----	54
Figura 14: Descritores de busca utilizados na base de dados Pubmed-----	60
Figura 15: Mapa do estado de Minas Gerais na América do Sul. -----	63
Figura 16: Principais biomas do estado de Minas Gerais -----	64

Figura 17: Divisão das Macrorregiões de saúde do estado de Minas Gerais, segundo Plano Diretor de Regionalização-----	65
---	----

ARTIGO 1

Figura 1: Fluxograma de inclusão e exclusão de artigos publicados entre 2016 e 2022 na base de dados Pubmed-----	71
--	----

ARTIGO 2

Figura 1: Distribuição das notificações de febre amarela em Minas Gerais entre os anos de 2007 e 2020 segundo a classificação final dos casos-----	83
--	----

ARTIGO 3

Figura 1: Matriz de correlação (r) entre as covariáveis utilizadas no estudo de previsão da FA em Minas Gerais entre 2007 e 2020-----	95
---	----

Figura 2: Resultado da análise do seletor de variáveis Boruta utilizado na pesquisa-----	96
--	----

Figura 3: Comportamento mensal das variáveis independentes e dependente dos dados de treino (macrorregião de saúde centro) utilizados no sistema de previsão para a série histórica entre 2007 e 2020 em Minas Gerais-----	97
--	----

Figura 4: Comportamento das variáveis independentes dos grupos de teste macrorregião nordeste utilizados no sistema de previsão para a série histórica entre 2007 e 2020 em Minas Gerais-----	99
---	----

Figura 5: Valores observados e previstos pelas modelagens de RF (a), MLP (b), Xgboost (c) e SVM (d) para previsão de casos de febre amarela em Minas Gerais com zoom na época de ocorrência de casos de FA-----	102
---	-----

Figura 6: Valores observados e previstos pelas modelagens: RF (a), Xgb (b), RNA (c) e SVM (d) para previsão de casos de febre amarela diários em Minas Gerais com zoom na época de ocorrência-----	103
--	-----

Figura 7: Valores de shap para as variáveis de entrada utilizadas nos algoritmos de previsão da febre amarela entre 2007 e 2020 em Minas Gerais-----	104
--	-----

LISTA DE TABELAS

Tabela 1: Estações meteorológicas convencionais e automáticas presentes nas macrorregiões de saúde Centro e Nordeste segundo município de localização, tipo de estação, código da estação, latitude, longitude e altitude-----67

ARTIGO 1

Tabela 1: Artigos incluídos na revisão integrativa segundo fator de impacto do ano 2020 ----72

ARTIGO 2

Tabela 1: Distribuição dos atendimentos de febre amarela humana segundo classificação final do caso e critério de classificação, em Minas Gerais entre 2007 e 2020-----88

Tabela 2: Distribuição do número de animais envolvidos nas epizootias registradas no SINAN segundo a área de ocorrência das epizootias em Minas Gerais entre 2007 e 2020-----93

ARTIGO 3

Tabela 1: Estatística dos valores das variáveis observadas utilizadas nos modelos de previsão no período de 2007 a 2020 em Minas Gerais para os dados de treino (Centro) dos modelos--98

Tabela 2: Estatística dos valores das variáveis independentes utilizadas nos modelos de previsão no período de 2007 a 2020 em Minas Gerais para os dados de teste dos modelos -----100

Tabela 3: Métricas de avaliação de desempenho dos modelos de aprendizado de máquina utilizados na previsão da febre amarela em Minas Gerais entre 2007 e 2020 para região nordeste-----103

LISTA DE ABREVIATURAS

FA- Febre amarela

IM- Imputação múltipla

MLP- Multi layer perceptron

PDRS- Plano diretor da regionalização da saúde

PNH- Primata não humanos

RF- Random forest

RNA- Rede neural artificial

SINAN- Sistema de Informação de Agravos de Notificação

SVM- Support vector machine

VFA- Vírus da febre amarela

XGboost- Extreme gradient boosting

Sumário

1. INTRODUÇÃO	11
2. OBJETIVOS	13
2.2.1. Artigo 1.....	13
2.2.2. Artigo 2.....	13
2.2.3. Artigo 3.....	13
3. CONTRIBUIÇÃO DA PESQUISA	14
4. REVISÃO DE LITERATURA - FEBRE AMARELA.....	14
4.1. Aspectos históricos	14
4.2. Aspectos gerais	25
4.3. Epidemiologia da febre amarela do Brasil.....	29
4.4. Epidemiologia da febre amarela em Minas Gerais	34
4.5. Vigilância da febre amarela no Brasil.....	37
5. REVISÃO DE LITERATURA - MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE DADOS	39
5.1. Aspectos gerais	39
5.2 Séries temporais	40
5.3. Paradigmas de aprendizado	41
5.3.1. Aprendizado supervisionado	42
5.3.2. Aprendizado não supervisionado	43
5.3.3. Aprendizado por reforço.....	44
5.4. Pré processamento dos dados	44
5.4.1. Padronização dos dados.....	45
5.4.2. Análise de correlação de Pearson.....	46
5.4.3. Imputação de dados ausentes em séries meteorológicas	46
5.4.4. Seleção de variáveis, atributos ou dados de entrada.....	47
5.5. Etapa de treinamento.....	48
5.6. Validação e teste do modelo.....	49
5.7. Dilema viés variância.....	49
5.8. Hiperparâmetros de algoritmos de aprendizagem.....	51
5.9. Redes neurais artificiais.....	52
5.10. Random forest	55
5.11. Algoritmos de boosting	56
5.12. SVM- Support vector machine	57
5.13. Avaliação da performance dos modelos de predição.....	57

5.14. Interpretação dos modelos de predição	59
6. MATERIAL E MÉTODOS.....	60
6.1. Artigo 1.....	60
6.1.1 Delineamento do estudo	60
6.2 Artigo 2.....	61
6.2.1. Delineamento do estudo	61
6.2.2. Variáveis do estudo	61
6.2.3. Análise de dados	61
6.2.4. Aspectos éticos	62
6.3. Artigo 3.....	62
6.3.1. Delineamento do estudo	62
6.3.2. Local do estudo.....	62
6.3.3. Fonte de dados	66
6.3.4 Análise de dados	69
6.3.5. Aspectos éticos	70
8. RESULTADOS E DISCUSSÃO	70
8.1 Artigo 1.....	70
8.1.2. Eco-epidemiologia da febre amarela na região Sudeste do Brasil entre 2016 e 2018: uma revisão integrativa da literatura científica	70
8.2 Artigo 2.....	82
8.2.1. Perfil epidemiológico dos atendimentos de febre amarela e epizootias notificados e registrados no SINAN, em Minas Gerais, entre 2007 e 2020	82
8.3. Artigo 3.....	94
8.3.1. Comparação de modelos de aprendizado de máquina para análise preditiva da febre amarela em Minas Gerais/Brasil.	94
9. CONCLUSÕES E CONSIDERAÇÕES FINAIS	106
10. REFERÊNCIAS BIBLIOGRÁFICAS	109
10. ANEXOS.....	129
10.1 Lista de municípios das macrorregiões de saúde centro e nordeste de acordo com Plano Diretor de regionalização da saúde de Minas Gerais 2014.....	129
10.2 Ficha de notificação de investigação de Epizootia.....	137
10.3 Ficha de investigação de febre amarela.....	138

1. INTRODUÇÃO

A febre amarela (FA) no Brasil é uma arbovirose de caráter zoonótico, não contagiosa, transmitida, principalmente, por culicídeos de dois gêneros: *Haemagogus* e *Sabethes* (Abreu *et al.*, 2019), tendo como principal fonte de infecção, primatas não humanos (PNH), sendo o homem hospedeiro acidental da doença (Roig *et al.*, 2009; Bicca-Marques *et al.*, 2010; Tauil *et al.*, 2010; Brasil, 2019a, Hill *et al.*, 2020).

No ser humano a doença se manifesta sob as formas assintomáticas, leves, moderadas e graves, sendo que na última, a taxa de letalidade pode chegar a 60% (Mascheretti *et al.*, 2013; Monath e Vasconcelos, 2015).

A ocorrência de casos de febre amarela têm incidido entre os meses de dezembro e maio, período sazonal da doença, ocorrendo em regiões de mata, quando o vírus encontra condições favoráveis para a transmissão como: elevadas temperaturas, umidade e pluviosidade, tendo como consequência o aumento na densidade de vetores; presença de hospedeiros primários susceptíveis como primatas não humanos; presença de indivíduos suscetíveis seja por motivo laboral ou lazer, associados às baixas coberturas vacinais (Brasil, 2017a).

O último relato de febre amarela urbana, transmitida por *Aedes aegypti*, no Brasil, ocorreu no ano de 1942 (Franco, 1969). Diante disso, houve historicamente entre pesquisadores e formuladores de políticas em saúde pública, um consenso de que a transmissão do vírus da febre amarela (VFA) permaneceria restrita ao ciclo silvestre, em áreas de mata, com baixa densidade populacional humana e, portanto, apresentando baixa incidência anual no país (Possas *et al.*, 2018). Entretanto, as arboviroses, tem representado um grande desafio à saúde pública, uma vez que ações antrópicas como desmatamento para diversos fins, defaunação, processo de urbanização desordenado, alterações climáticas e ambientais, aproximaram nas últimas décadas, o homem, animais silvestres e vetores, fortalecendo cada vez mais a troca de agentes entre espécies em consequência de um grande desequilíbrio ecológico (Volpato *et al.*, 2020).

Após décadas de silêncio, o vírus da febre amarela que se restringia à região Amazônica, considerada endêmica no país, chegou ao Sudeste e Sul do Brasil se espalhando pela região de domínio de Mata Atlântica, alcançando em menos de doze meses, quatro populosos estados

brasileiros (Minas Gerais, São Paulo, Rio de Janeiro e Espírito Santo). No período sazonal de 2016/2017, a doença se espalhou por uma área na qual estima-se que residiam 8 milhões de pessoas e, em 2017/2018, na área afetada os valores ultrapassaram os 32 milhões de indivíduos (Brasil, 2018a), nos quais, parte da população não estava incluída no programa de vacinação contra a doença.

A baixa cobertura vacinal da população nestas áreas resultou em um aumento acentuado no número de casos humanos da doença (Vasconcelos, 2010a; Possas *et al*, 2018) e reacendeu o debate sobre o risco de reurbanização da doença em regiões com alta infestação por *Aedes aegypti* (Vasconcelos, 2010b; Cavalcante e Tauil, 2016; Possas *et al.*, 2018; Lobão e Rodrigues, 2019, Silva *et al*, 2020a).

O recente deslocamento do VFA pelo território brasileiro não foi linear. Como em grandes surtos silvestres anteriores, ocorreu em vias complexas e múltiplas, em ondas epizoóticas combinada com a mobilidade intensa sem precedentes de humanos infectados assintomáticos envolvidos no trabalho rural, ecoturismo, extrativismo e tráfico de animais (Possas *et al.*, 2018).

Casos humanos e epizootias em primatas não humanos (PNH) foram registrados em uma ampla área do território nacional. Inicialmente, entre 2014 e 2015, a transmissão se deu na região norte, com posterior expansão no sentido leste e sul do país, onde afetou prioritariamente a região Centro Oeste entre 2015 e 2016 (Brasil, 2019b).

Mais recentemente, entre 2016 e 2018, foram registrados os surtos mais expressivos de febre amarela no Brasil. Os estados da região Sudeste foram os mais acometidos, onde 1.376 casos humanos e 483 óbitos foram notificados pelos serviços de saúde, além de 864 epizootias em primatas não humanos confirmadas para febre amarela (Brasil, 2019b).

A vigilância da febre amarela é indispensável para prevenção de novos casos, através da identificação de áreas de risco antes que a doença se torne uma ameaça à saúde humana e animal. Ressalta-se a importância de reconhecer os fatores antropogênicos e suas conexões ecológicas com a dinâmica da doença, seus reservatórios e ambientes específicos (Nava *et al.*, 2017).

Diante do exposto, se faz necessário encontrar modelos matemáticos que tenham alta eficiência na previsão de séries temporais, para que as tomadas de decisão e gestão em saúde pública sejam mais assertivas e eficientes a fim de se evitar que surtos de febre amarela de tamanha

magnitude, como os ocorridos entre os anos de 2016 a 2018 no Sudeste brasileiro venham a se repetir.

2. OBJETIVOS

2.1. Objetivo geral

O presente trabalho objetivou conhecer a eco-epidemiologia da febre amarela, descrever os bancos de dados utilizados na pesquisa e comparar a performance de algoritmos de aprendizado de máquina para análise preditiva de problema de regressão envolvendo série temporal, frequentemente presente na área de saúde pública. Utilizamos quatro algoritmos de aprendizado de máquina para prever casos de febre amarela no estado de Minas Gerais e comparamos as performances preditivas.

2.2. Objetivos específicos

2.2.1. Artigo 1

Descrever os aspectos eco-epidemiológicos da febre amarela no estado de Minas Gerais, relacionadas ao ambiente, ao vetor e aos principais hospedeiros através de uma revisão integrativa da literatura científica pertinente.

2.2.2. Artigo 2

Descrever a ocorrência de febre amarela no estado de Minas Gerais entre 2007 e 2020 envolvendo casos humanos e epizootias em primatas não humanos através da análise de dados oriundos das fichas de notificação de febre amarela e epizootias em primatas não humanos registrados pelos órgãos competentes no Sistema de Informação de Agravos de Notificação-SINAN.

2.2.3. Artigo 3

Modelar e treinar quatro algoritmos de aprendizado de máquina para prever casos de febre amarela em uma macrorregião de saúde do estado de Minas Gerais utilizando variáveis relacionadas ao ambiente, aos vetores e aos hospedeiros humanos e primatas não humanos.

Avaliar a performance preditiva de cada algoritmo de aprendizado de máquina para a série temporal da febre amarela no estado de Minas Gerais.

3. CONTRIBUIÇÃO DA PESQUISA

Encontrar modelos de aprendizado de máquina que possam prever com alta performance casos humanos da febre amarela utilizando como dados de entrada variáveis relacionadas ao ambiente (meteorológicas), aos vetores, aos hospedeiros naturais (PNH) e aos hospedeiros acidentais (seres humanos) é importante para o desenvolvimento de uma ferramenta útil para a saúde pública pois poderá impactar de forma positiva as tomadas de decisão e reduzir assim os impactos da doença nas populações expostas ao risco.

4. REVISÃO DE LITERATURA - FEBRE AMARELA

4.1. Aspectos históricos

Febre amarela nas Américas e origem geográfica

Por muitos séculos, debates sobre a origem geográfica da febre amarela ocorreram no meio científico.

Bérenger- Féraud comenta em seu livro: *Traité Théorique et clinique de la fièvre jaune* (Tratado teórico da clínica da febre amarela), publicado na França em 1890, que primeira aparição da doença ocorreu cerca de 400 anos antes da publicação de seu livro sobre a febre amarela, entre os anos de 1494 e 1495 em uma viagem de Cristóvão Colombo às Américas. Bérenger-Féraud ainda afirma em seu livro: “*En réalité la fièvre jaune n'a pas été observée dans l'antiquité; cest en vain qu'on voudrait prétendre qu'hippocrate l'a décrite dans son livre des épidémies*” (Na realidade a febre amarela não foi observada na antiguidade, seria em vão acreditar que Hipócrates a descreveu em seu livro das epidemias) (Bérenger- Féraud, 1890).

O primeiro relato da febre amarela ocorreu em 1495, quando espanhóis travaram um combate contra indígenas, na ilha Espanhola, que hoje é o Haiti, numa batalha conhecida como Vega Real ou Santo Serro. Apesar de estarem em maior número, os nativos refugiaram-se nas florestas diante do poderio europeu. Colombo se viu obrigado a fazer incursões pelas florestas no interior da ilha em busca dos indígenas (Franco, 1969).

Aproximadamente dois meses após a batalha de Vega Real, surgiu uma epidemia entre indígenas e europeus na qual grande número de indivíduos ficaram doentes, destes, muitos foram à óbito. Mesmo com a descrição incompleta dos sintomas sugeriu-se que se tratava de febre amarela (Franco, 1969). Nota-se que houve um lapso de tempo relativamente grande entre as incursões nas matas pelos europeus na busca dos índios na batalha de Vega Real até a ocorrência da primeira epidemia. Apesar disso, muitos médicos acreditaram que a doença já existia na América (Chippaux e Chippaux, 2018).

De acordo com Bérenger-Féraud (1890), já se debatia sobre a ocorrência da febre amarela no continente africano. A ocorrência espontânea da doença em navios vindo da da África também foi sugerida baseadas nas condições desumanas em que os africanos eram levados nos navios europeus, e apesar de existirem descrições sobre epidemias de escorbuto, beribéri e disenteria, sinais compatíveis com febre amarela não eram descritos até a epidemia de 1495.

Vários foram os defensores da origem africana da doença como Pym em 1815, Audouard em 1824, Faget em 1859 além de Henry Carter e Harold Scott (Franco, 1969).

Ocorreram inúmeros relatos por navegadores europeus em rotas marítimas ao longo da costa africana e nas Ilhas Canárias de casos de doença com nomes diversos e sintomas muito similares aos da febre amarela, mesmo antes do encontro do europeu com o povo americano (Augustin, 1909). Os primeiros casos de febre amarela em solo americano resultaram, portanto, da introdução do vírus amarílico pela tripulação de Colombo vindo das ilhas canárias na África, onde a embarcação fez seu último abastecimento antes de atravessar o atlântico (Chippaux e Chippaux, 2018).

Hoje, através de estudos filogenéticos, a tese aceita com consenso entre pesquisadores é que a origem do vírus da febre amarela é africana e a entrada do vírus nas Américas cursou com a expansão colonial nas Américas e África em que o vírus e o vetor *Aedes aegypti* provavelmente chegaram no continente Americano nos porões de embarcações durante o comércio de escravos oriundos da África, encontrando em solo Americano, condições ecológicas favoráveis à sua manutenção, incluindo vetores silvestres competentes (Bryant *et al.*, 2007).

A primeira epidemia descrita com relativa precisão foi feita pelo jesuíta Raymond Bréton em 1635 na ilha de Guadalupe. Dentre os sintomas descritos, Bréton mencionava presença de dores lombares, icterícia (“as pessoas doentes ficavam mais amarelos do que marmelos”), presença de vômitos negros, sendo que a morte ocorria entre o terceiro e o quinto dia. Bréton estabeleceu

uma relação próxima entre a derrubada de árvores e a doença, ao mencionar que “à medida que cortavam os bosques, a terra arrojava seu veneno” (Franco, 1969).

Cinco anos após o relato de Bréton, Padre Du Tertre recém-chegado a Guadalupe relatou a mesma a epidemia, a qual ainda estava presente naquele local. Du Tertre descreveu que pessoas que se ocupavam do corte de árvores adoeciam e tinham como sintoma fortes dores lombares, esse sintoma era conhecido como golpe de barra e acreditava-se que a febre era oriunda dos vapores exalados no momento da queda das árvores (Franco, 1969).

É importante ressaltar que Bréton e Du Tertre descreviam provavelmente a forma silvestre da febre amarela, já que as incursões na mata e atividades extrativistas antecederam o surgimento de sintomas da doença. Ambos relatos são considerados como as primeiras descrições mais aceitáveis da febre amarela devido à riqueza de detalhes (Franco, 1969).

Outro relato importante sobre a febre amarela é o manuscrito maia de Chumayll que descreve a doença cursando com vômito de sangue levando o povo à morte no ano de 1648 (Franco, 1969).

Entre 1668 e 1893, ocorreram cerca de 135 epidemias nos Estados Unidos da América. Em 1793, uma delas matou um em cada dez residentes da Filadélfia. Em 1878 morreram de febre amarela cerca de 20.000 pessoas no vale do Mississippi (McCarthy, 2001). Em 1905 eclodiu a última epidemia de febre amarela nos Estados Unidos (Franco, 1969).

Febre Amarela no Brasil

O primeiro relato da ocorrência da febre amarela no Brasil, data de 1685, na cidade de Recife, quando o Brasil era colônia de Portugal (Franco, 1969; Teixeira, 2001; Costa *et al.*, 2011). De acordo com Franco (1969) a doença chegou ao país através de uma embarcação vinda de São Tomé, na África, a qual fez parada nas Antilhas, local onde a doença assolava a população. Essa foi a primeira epidemia de febre amarela do país.

Em 1686, uma epidemia de Febre Amarela teve início em Salvador, na Bahia, com duração de aproximadamente seis anos, com aproximadamente 25.000 infectados e cerca de 900 óbitos (Franco, 1969; Teixeira, 2001; Costa *et al.*, 2011).

Em 1691 foram instituídas as primeiras medidas profiláticas de que se tem conhecimento no Brasil. Por relacionarem a doença ao “ar pestilial”, baseados na teoria miasmática que imperava

nessa época, utilizou-se a "ditadura sanitária" que incluía a segregação dos doentes, purificação do ar, das casas, de portos, ruas, cemitérios e outros, alcançando assim, o resultado esperado. Essa campanha lançou bases do modelo das estratégias de vigilância e controle que se seguiram ao longo dos anos (Costa *et al.*, 2011).

Não houveram mais relatos de epidemias de febre amarela no Brasil colônia (Teixeira, 2001).

Mais de um século se passou até que uma nova epidemia foi deflagrada em Salvador, Bahia, em 1849. Nesse momento de Brasil Império, a doença encontrava-se quase esquecida, entretanto, a partir daí, a febre amarela se espalhou de norte ao sul do país (Franco, 1969) com altíssimas taxas de morbidade e alta taxa de letalidade tornando-se um dos maiores problemas de saúde pública e econômico, já que representavam um entrave à continuidade do processo de importação de mão de obra e dificultavam o desenvolvimento do comércio com os países europeus, elementos vitais à manutenção do modelo econômico naquele período (Teixeira, 2001).

A capital do Império, Rio de Janeiro, em 1850 registrou 4.160 óbitos pela febre amarela, tornando-se um gravíssimo problema nacional. Diante desse quadro, nesse mesmo ano, foi, então, instituído pela Secretaria de Estado de Negócios do Império o "Regulamento Sanitário", o qual estabelecia normas para a execução da segunda campanha contra a Febre amarela organizada no Brasil (Franco, 1969; Benchimol, 2001; Lowy, 2006).

As práticas instituídas para enfrentar a epidemia incluíam a desinfecção de navios, quarentena dos ocupantes, cuidados especiais com enterros e velórios, medidas sanitárias coletivas que incluíam aterramento de valas e limpeza de esgotos, dentre outras, as quais eram muito parecidas com as medidas profiláticas tomadas 159 anos antes (Franco, 1969; Costa *et al.*, 2011).

Após os resultados satisfatórios da campanha orientada pelo regulamento sanitário, uma Junta de Higiene Pública e uma Comissão de Engenheiros foram instituídas para organizar a defesa sanitária no país com papel de propor orientações e medidas a fim de promover benefícios à saúde pública no país, regulamentadas pela lei 578 de setembro de 1850. Esse é o primeiro relato sobre organização governamental direcionada para as ações de controle da febre amarela no Brasil. (Costa *et al.*, 2011)

Ao longo dos anos, a doença foi perdendo o caráter epidêmico e tornando-se de caráter endêmico e assim permaneceu, com maior ou menor intensidade, até o verão de 1861, quando praticamente foi extinta (Franco, 1969; Costa *et al.*, 2011).

Naquela época, a doença era considerada como contagiosa e a classe médica encontrava-se até então dividida entre médicos de renome que eram classificados em contagionistas e não-contagionistas. A controvérsia entre estes dois grupos durou longos anos e somente terminou com os trabalhos de Pasteur e Koch no fim do século XIX, que lançaram luz a um novo campo para explicar a etiologia das enfermidades. As hipóteses sobre a natureza da febre-amarela passaram a orientar-se, então, para os microrganismos (Franco, 1969, Teixeira, 2001).

Controvérsias da febre amarela. Do modo de transmissão ao agente etiológico

Em meio às epidemias que se espalharam ao longo do continente Americano, cientistas de várias partes do mundo tentavam descobrir duas controvérsias: a forma de transmissão e o agente causador da febre amarela.

Nesse contexto, Finlay, um médico e pesquisador de Cuba, levantou a hipótese sobre a transmissão da febre amarela ser feita por mosquitos (*Stegomyia fasciata*), atualmente conhecido como *Aedes aegypti*. Em seu artigo Finlay discorre sobre a transmissão do agente, até então desconhecido, no corpo de um ser humano são, via repasto sanguíneo realizado por mosquito fêmea hematógenas. Finlay ainda discorre sobre fatores ambientais como: calor, umidade, presença de água parada, os quais favorecem o desenvolvimento dos mosquitos (Finlay, 1881).

No ano de 1900, uma comissão especial foi enviada, pelo governo Norte Americano, à Cuba para investigar as mortes de soldados por febre amarela. A comissão composta por médicos, dentre eles, Walter Reed, inicialmente foram tentar confirmar a alegação de Sanarelli de que *B icteroides* era o agente responsável pela doença, mas determinaram, por fim, que era um contaminante na amostra estudada (McCarthy, 2001; Teixeira, 2001).

A hipótese de Finlay sobre a transmissão da febre amarela por mosquitos foi então testada e comprovada pela equipe de Reed, que na ausência de animais, fizeram seus experimentos em pessoas, sendo os resultados publicados no *Journal of the American Medical Association* (McCarthy, 2001).

A origem microbiana da febre amarela foi discutida e defendida por diversos pesquisadores do mundo. No Brasil, o pioneiro foi o professor de química orgânica e biologia da escola de medicina do Rio de Janeiro, em 1880, Domingos Freire. Segundo ele, a febre amarela era causada pelo *Cryptococcus xanthogenicus*, “um micro germe, um parasita, um micróbio da espécie das algas”, que aparecia inicialmente sob a forma de pequenos pontos quase imperceptíveis e aumentava gradualmente de diâmetro. Domingos Freire passou 20 anos de sua vida defendendo sua teoria sobre o micróbio causador da febre amarela e sendo o primeiro apologista da vacina quando afirmava que o desenvolvimento da mesma seria o futuro da profilaxia da doença (Franco, 1969).

O agente etiológico da febre amarela foi finalmente isolado em 1927 por duas equipes independentes: a Fundação Rockefeller e o Instituto Pasteur. Os médicos da Fundação Rockefeller H. Bauer e A. F. Mahaffi comprovaram a suscetibilidade do macaco rhesus ao agente da febre amarela mediante a inoculação de sangue de um paciente de nome Asibi (daí o nome da cepa) na África Ocidental, o qual apresentara a forma moderada da doença, e em seguida, Stokes, Bauer e Hudson confirmaram que o agente era um pequeno vírus filtrável (Costa *et al.*, 2011).

Febre amarela no Brasil no século XX

No Brasil, a entrada do século XX foi marcada por graves epidemias de febre amarela que voltaram a assolar a população brasileira. Emílio Ribas, o então diretor do Serviço Sanitário do estado de São Paulo (1898-1917), publicou um opúsculo sobre o mosquito ser o propagador da febre amarela. Correlacionou em seu documento sete providências a serem tomadas contra o vetor da febre amarela. A primeira cidade a colocar em prática as medidas desenvolvidas por Emílio Ribas foi Sorocaba com sucesso (Franco, 1969; Teixeira, 2001; Costa *et al.*, 2011). Ribas também foi responsável pela criação do Instituto Butantan em São Paulo (Franco, 1969).

No Rio de Janeiro, Oswaldo Cruz foi responsável pelas ações de controle à febre amarela, as quais foram marcadas por ousadia e pela “força sanitária”. Com bases técnicas concretas sobre a transmissão da doença, e no fato da doença não ser contagiosa, Oswaldo Cruz munuiu-se de plenos poderes para adotar as medidas de controle da doença e direcionou as ações de vigilância e controle para o ponto focal: o mosquito transmissor. O trabalho foi organizado sob forma de campanhas inspiradas na disciplina militar, com os exércitos de “mata mosquitos” entrando nas casas à procura de focos do vetor. A febre amarela foi a primeira doença de notificação obrigatória no país. (Franco, 1969; Costa *et al.*, 2011).

Após a Era Oswaldina, com a redução de recursos para combate ao vetor e relaxamento das medidas de controle, as epidemias que arrefeceram passados 20 anos, voltaram a vitimar grande número de pessoas em diversos municípios do país inclusive no Rio de Janeiro que em 1928 e 1929 foi palco das últimas epidemias em meio urbano em que 738 casos foram notificados, com uma taxa de letalidade de 64,8%. Nesse caso observa-se que houve, provavelmente, uma subestimação dos casos de febre amarela na época. Em 1942 ocorreu o último caso de febre amarela urbana, transmitida por vetor urbano, o *Aedes aegypti*, no estado do Acre (Franco, 1969; Costa *et al.*, 2011).

É importante ressaltar que as consequências das epidemias de febre amarela no Brasil, seja ele colônia, Império ou República até o início do século XX, não se resumiram a problemas relacionados à saúde pública, mas envolviam problemas de cunho econômico, político e social (Costa *et al.*, 2011).

Economicamente os prejuízos relacionados ao turismo e exportações eram drásticos, uma vez que navios vindos do Brasil precisavam passar por quarentena nos portos até que a carga fosse liberada e o risco relacionado à doença afastava possíveis turistas que pretendiam visitar o país (Vasconcelos, 2003). A população brasileira, em sua maioria, vivia em condições sanitárias inadequadas e precárias nos centros urbanos, onde era assolada não só pela febre amarela, mas também pela varíola e peste bubônica. O embate político, que se instaurava no país em relação às tomadas de decisão em saúde, era complexo e muitas vezes realizados por meio da força. Assim, a febre amarela consolidou-se como um dos mais graves problemas ocorridos em solo brasileiro até o início do século XX (Costa *et al.*, 2011).

Era Rockefeller no Brasil

Na passagem do século XIX para o século XX, os pesquisadores verificaram que muitas doenças tropicais possuíam hospedeiros intermediários, em geral insetos hematófagos. Os estudos epidemiológicos foram, então, ampliados, de modo a abarcar a ecologia e o ciclo de vida desses vetores invertebrados de doenças, o ambiente natural em que viviam e suas interações com os humanos. Uma abordagem holística da epidemiologia das doenças tropicais procurou compreender todo o ecossistema em que a doença ocorria, e se mostrou interessada nas multidimensionais interações entre humanos, patógenos e insetos (Löwy, 1999).

Em 1920 iniciaram-se os estudos sobre febre amarela pela Fundação Rockefeller no Brasil. À frente das ações relacionadas à febre amarela, encontrava-se o Dr. Soper, o qual reconheceu

que a epidemiologia da doença não mais obedecia às regras clássicas, pois se mantinha no Brasil apesar da aplicação de métodos que antes foram bem-sucedidos (Franco, 1969).

Mediante estudos epidemiológicos a existência do ciclo silvestre da febre amarela foi comprovada no ano de 1932, no Vale do Canaã, Estado do Espírito Santo. De acordo com as pesquisas, não se encontrou o vetor *Aedes aegypti* na região de ocorrência da epidemia de febre amarela apesar das seis semanas de buscas. Outros culicídeos foram encontrados em número adulto e fase larval em quantidade suficiente para ser responsabilizado pela epidemia, dentre eles destacou-se o *Aedes (Ochlerotatus) scapularis* (Soper *et al.*, 1933).

A descrição da febre amarela silvestre em 1932 e a nova maneira de ver a doença, agora associada a animais silvestres, e apenas acidentalmente transmitida ao ser humano, pôs fim ao objetivo de eliminá-la totalmente no Brasil. Em contrapartida, o aumento da eficácia das campanhas contra os mosquitos levou à erradicação não planejada, realmente imprevista, do *Aedes aegypti* em áreas selecionadas. O alvo das campanhas contra a febre amarela então passou a ser a erradicação plena do vetor das formas epidêmicas, o mosquito *Aedes aegypti* e a identificação de áreas onde a febre amarela silvestre persistia endemicamente, acrescentando-se, a partir de 1936, o controle do vírus através da vacinação (Löwy, 1999).

Após a descoberta do ciclo silvestre da doença, a busca pelos possíveis vetores e hospedeiros da febre amarela e a tentativa de se conhecer sua área de abrangência geográfica tornaram-se os objetivos principais da Fundação Rockefeller no Brasil (Costa *et al.*, 2011).

A Fundação Rockefeller produziu um grande acervo de imagens contendo cerca de 15 mil documentos, associadas a uma rica descrição epidemiológica, ambiental e social das viagens pelo interior do país, que incluía trabalho de campo, moradias dos "amareletos", mapas aéreos, desenhos, representações cartográficas (Fig. 1), atividades de laboratório, e outros itens, gerando, assim, ferramentas valiosas para caracterizar os cenários de manifestação da doença e estabelecer a política de intervenção técnica (Costa *et al.*, 2011).



Figura 1. Provável rota da disseminação da onda epizootico-epidêmica de febre amarela no Brasil – 1932-1942. Fotógrafo: A. Fialho.

Fonte: Acervo da Casa de Oswaldo Cruz, Departamento de Arquivo e Documentação - Foto FR (SFA-EC) 12-5 de A. Fialho In: Costa *et al*, 2011.

A vacina da febre amarela

No ano de 1937, foi criada e registrada a primeira vacina eficaz contra febre amarela, chamada de cepa 17D. Em seguida, Theiler e Smith realizaram subculturas do vírus 17D em embrião de galinha, até obterem a atenuação do viscerotropismo e neurotropismo por meio de testes em primatas (Costa *et al.*, 2011).

A febre amarela tornou-se um problema de saúde pública tão grave que se acelerou as etapas do ensaio clínico com a nova vacina e, assim que foi confirmada sua capacidade imunogênica, ela já foi testada em 100 voluntários humanos da Fundação Rockefeller, em Nova York (Franco, 1969).

No Brasil, a vacina contra a febre amarela começou a ser produzida no Rio de Janeiro com uma amostra do vírus 17D trazida de Nova York por Hugh H. Smith, em 1937. Essa amostra já vinha sendo modificada por culturas sucessivas em meios diversos como macacos, mosquitos, tecido embrionário de camundongo com soros normais e heterólogos até se chegar às passagens seriadas *in vitro*, em meio de cultura contendo tecido de embrião de galinha despojado do sistema nervoso central. A partir da subcultura 214, contada desde o Asibi original, foram iniciadas muitas séries paralelas, algumas por passagem direta em ovos embrionados. Essas séries conservavam, mais ou menos inalteradas, as características benignas para macacos e

homens da amostra 17D, ainda que se verificassem pequenas variações na velocidade e intensidade de sua replicação na cultura em tecido ou em ovos embrionados (Benchimol, 2001).

Desde março daquele ano o Instituto Oswaldo Cruz passou a preparar a vacina 17D (Franco, 1969). Atualmente, no Brasil, está sendo utilizada a cepa 17DD, que difere da anterior somente pelo número de passagens em um hospedeiro animal, mas é igualmente produzida a partir de lotes sementes derivados da cepa original 17D (Costa *et al.*, 2011).

A vacina antiamarílica foi utilizada por muitos anos como a medida profilática de escolha em regiões de florestas e matas, onde o vetor *Aedes aegypti* não tem participação no ciclo de transmissão da doença, sendo assim protegeria a população residente e os trabalhadores expostos à infecção nessas regiões, além de prevenir a entrada do vírus em áreas urbanas por meio de indivíduos em fase de viremia (Costa *et al.*, 2011).

O foco da vigilância da febre amarela no país passou a ser a erradicação do vetor, *Aedes aegypti*, sendo este um grande desafio, já que o Brasil é um país de dimensões continentais (Franco, 1969; Costa *et al.*, 2011).

A Campanha contra a Febre Amarela se alicerçava em três pilares: na vigilância anti-aegypti, que consistia na inspeção vetorial de 100% dos prédios, com tratamento à base de inseticidas naqueles positivos para o mosquito; na viscerotomia, com o controle rígido do número de amostras de fígado coletadas; e na vacinação em larga escala em área endêmica (Costa *et al.*, 2011).

O País foi dividido pelo Serviço Nacional de Febre-Amarela em 6 regiões administrativas para o combate ao vetor e em 1958, o Departamento Nacional de Endemias Rurais, que já havia absorvido o Serviço Nacional de Febre-Amarela, declarou estar erradicado o *Aedes aegypti* do país (Franco, 1969).

Foi então organizado um Serviço de Vigilância, que tem por objetivo a fiscalização das portas de entrada (aeroportos de intercâmbio internacional, localidades situadas nas fronteiras, portos marítimos e fluviais etc.), a fim de evitar a penetração do mosquito procedente de países ainda infestados (Franco, 1969).

Enquanto o governo empenhava forças em acabar com o vetor urbano da doença, casos de febre amarela silvestre ocorriam em diversos estados do Brasil, com número maior de casos

notificados naqueles que tinham como vegetação predominante o cerrado, mata Atlântica e Floresta Amazônica (Fig. 2).

Unidade da Federação	Nº de casos humanos de febre amarela
Acre	14
Amazonas	17
Bahia	15
Espirito Santo	281
Goias	210
Maranhão	3
Mato Grosso	115
Minas Gerais	417
Pará	63
Paraná	91
Piauí	1
Rio Grande do Sul	7
Rio de Janeiro	65
Rondônia	2
Roraima	5
Santa Catarina	40
São Paulo	329
Total	1.675

Figura 2. Casos de Febre amarela silvestre, confirmados laboratorialmente, de 1932 a 1967 no Brasil.

Fonte: Adaptado de Odair Franco, 1969.

É certo que países das Américas que mantinham ligações marítimas e aéreas com a Região Norte, ou já haviam se reinfestado ou ainda não tinham erradicado o vetor de seus territórios, constituindo dessa forma uma ameaça permanente de reinfestação de nosso país, até que em março de 1967, na cidade de Belém, foi confirmado um foco de *Aedes aegypti* (Franco, 1969).

O país encontrava-se despreparado para enfrentar o problema. No ano seguinte, focos de *Aedes aegypti* foram reencontrados também no Maranhão; em 1976, na Bahia e assim o mosquito foi reconquistando seu território primitivo, chegando às cidades do Rio de Janeiro e Natal em 1977;

a Santos em fevereiro de 1980; estaria presente em 226 municípios seis anos depois, ao irromper nas manchetes dos jornais como protagonista de um "novo" tipo de epidemia urbana, a epidemia de dengue (Fiocruz, 2008).

Atualmente, com a inexequibilidade da eliminação completa do *Aedes aegypti*, transmissor não só a febre amarela, mas de outras importantes arboviroses, a vacina é o único meio eficaz para prevenir e controlar a febre amarela, uma vez que a mesma, interrompe o ciclo de transmissão da doença (Costa *et al.*, 2011).

4.2. Aspectos gerais

A febre amarela caracteriza-se como uma doença infecciosa febril aguda, imunoprevenível, não contagiosa, de gravidade variável e elevada letalidade nas suas formas graves. O vírus amarílico, agente etiológico da doença, é transmitido por artrópodes, da família Culicidae, popularmente conhecidos como mosquitos e pernilongos. A importância epidemiológica da doença decorre da gravidade clínica, da elevada letalidade e do potencial de disseminação e impacto, sobretudo caso ocorra a transmissão urbana, por *Aedes aegypti* (Vasconcelos, 2003; Brasil, 2019a).

O vírus amarílico é endêmico em países da América Central, América do Sul e nas regiões tropicais da África. Ao todo existem 47 países no grupo de risco para febre amarela, sendo 34 na África e 13 nas Américas (OPAS, 2017).

A Febre amarela é uma arbovirose, e como tal, é causada pelos chamados arbovírus (*ARthropod BORne VIRUS*), que incluem o vírus da dengue, Zika vírus, Chikungunya e febre amarela, os quais têm sido motivo de grande preocupação em saúde pública em diversas regiões do mundo (De Paiva *et al.*, 2019; Brasil, 2019a).

O vírus da Febre Amarela (VFA) é membro protótipo do gênero *Flavivirus* (família Flaviviridae), com um genoma de RNA de fita simples (Lindenbach *et al.*, 2013). Desde seu isolamento e caracterização molecular até os dias atuais, um sorotipo e sete genótipos foram descritos na África e na América, sendo eles: África Ocidental I, África Ocidental II, África Oriental, Angola e Centro-leste Africano. (Monath e Vasconcelos, 2015) e América do Sul I e II, sendo os últimos derivados do genótipo da África Ocidental; e o América do Sul I, o genótipo predominante no Brasil (Vasconcelos *et al.*, 2003).

O genótipo América do Sul I possui cinco linhagens distintas. Até meados dos anos 90, as antigas linhagens (1A, 1B e 1C) co-circularam na América do Sul, mas foram substituídas pelas modernas, 1D e 1E (Mir *et al.*, 2017), sendo a linhagem 1E, a responsável pelos recentes surtos de febre amarela no Brasil entre os anos de 2016 e 2019, a qual foi, provavelmente, originada de áreas endêmicas no Norte do Brasil (Faria *et al.*, 2018).

O VFA, no Brasil, é transmitido por artrópodes silvestres (vetores), da família Culicidae, de dois gêneros, *Haemagogus* (*H. janthinomys* e *H. albomaculatus*) e *Sabethes*, tendo como principal fonte de infecção os PNH (De Paiva *et al.*, 2019; Brasil, 2019a).

Os gêneros, *Haemagogus* são considerados os vetores primários FA e devido à sua grande distribuição espacial combinada com abundância populacional e boas taxas de infecção natural, atuaram para uma rápida disseminação e severidade do surto no Sudeste do Brasil iniciado em 2016 (Abreu *et al.*, 2019).

O clima apresenta forte influência na ecologia dos vetores artrópodes transmissores de doenças. Por serem ectotérmicos, seu metabolismo varia com as flutuações diárias de temperatura, o que pode afetar características fisiológicas relacionadas à competência do vetor (Paaijmans *et al.*, 2013) como: atividade muscular e taxa de alimentação; além de aumento da densidade populacional, sazonalidade, fecundidade e capacidade de dispersão (Carvalho *et al.*, 2017).

Existem dois ciclos da febre amarela no Brasil, definidos epidemiologicamente: o ciclo silvestre e o ciclo urbano (Fig. 3).

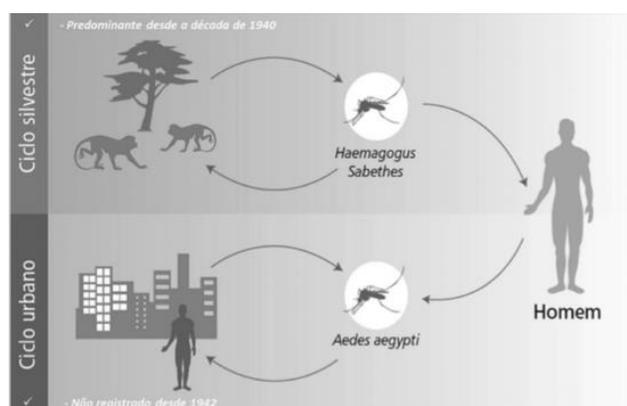


Figura 3. Os ciclos da Febre Amarela no Brasil

Fonte: Guia de Vigilância de Epizootias em Primatas Não Humanos e Entomologia Aplicada à Vigilância da Febre Amarela, 2014.

No ciclo silvestre da febre amarela, os PNH são considerados os principais hospedeiros, vítimas da doença, assim como o homem, que neste ciclo se apresenta como hospedeiro acidental. As principais espécies de culicídeos (mosquitos silvestres) implicadas na transmissão são *Haemagogus janthinomys* e *Haemagogus leucocelaenus*, além de diversas espécies do gênero *Sabethes*, os quais são classificados como acrodendrófilos, ou seja, seus hábitos alimentares e reprodutivos se dão nas copas das árvores (Vasconcelos, 2003).

Durante as epizootias nas florestas, os humanos podem se infectar acidentalmente, nas margens de florestas ou entrando na mata, ao serem picados por mosquitos infectados. Os mosquitos são considerados os verdadeiros reservatórios do vírus da febre amarela, pois, uma vez infectados, permanecem assim durante toda a vida (Vasconcelos, 2003).

No ciclo urbano, não registrado no Brasil desde 1942, o homem é o principal hospedeiro com importância epidemiológica, e o vetor, o *Aedes aegypti*, mantendo-se assim, o ciclo homem-mosquito (Fig. 4) (Brasil, 2019a).

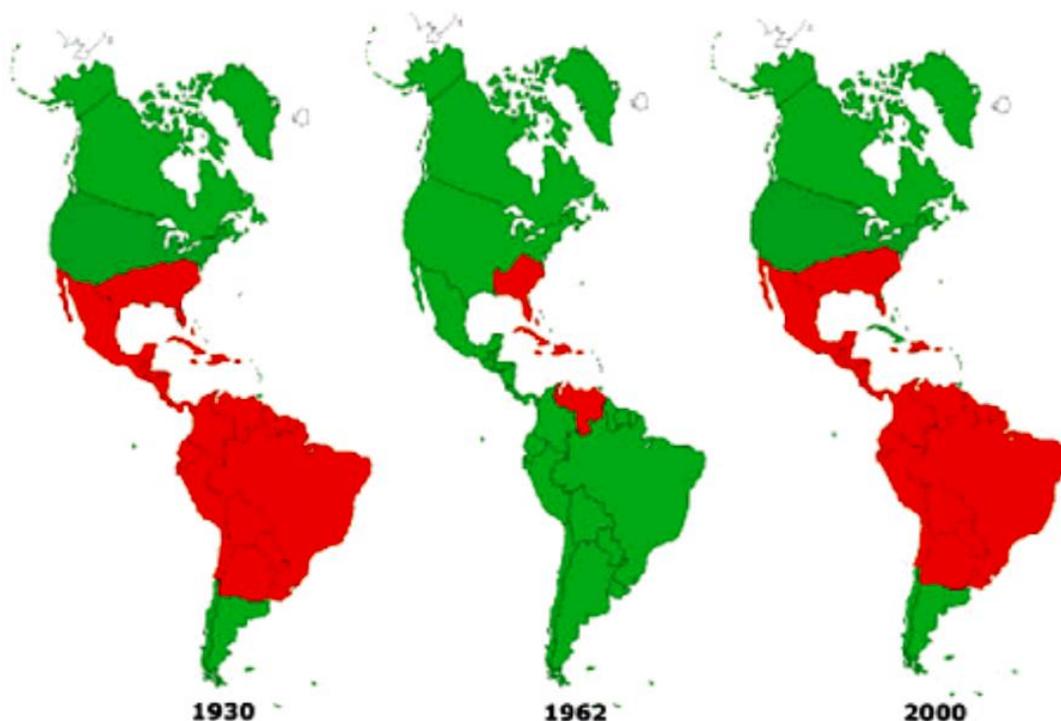


Figura 4. Dispersão do *Aedes aegypti* no continente americano ao longo do século XX.

Fonte: Vasconcelos, 2003.

Nas regiões tropicais do continente africano, onde a doença é endêmica, além do ciclo silvestre e urbano, existe também o ciclo intermediário, no qual vetores adaptados a se reproduzirem

tanto em ambientes urbanos quanto em ambientes silvestres se encarregam de transmitir a doença a primatas não humanos e aos humanos. É dessas áreas de emergência africanas que surgem surtos que chegam ao ambiente urbano naquele continente. No ciclo urbano, o *Aedes aegypti* assume a transmissão independente da presença de primatas amplificadores. Estimou-se através de modelamento matemático que o surto, ocorrido no ano de 2013, de febre amarela na África teve entre 84.000-170.000 casos graves e 29.000-60.000 mortes pela doença (WHO, 2021).

A febre amarela apresenta um padrão temporal de ocorrência sazonal, com a maior incidência de casos entre dezembro e maio. Os surtos apresentam periodicidade irregular, quando o vírus encontra condições favoráveis para a transmissão (elevadas temperatura e pluviosidade; alta densidade de vetores e hospedeiros primários; presença de indivíduos suscetíveis; baixas coberturas vacinais; eventualmente, novas linhagens de vírus) (Brasil, 2021b).

Na América Latina, de forma geral, as espécies de PNH presentes são susceptíveis ao VFA, sendo seus hospedeiros naturais, pois são basicamente arborícolas e habitam o mesmo estrato arbóreo que as principais espécies de mosquitos vetores (Souza, 2013). Dentre os PNH neotropicais os gêneros *Alouatta* e *callithrix* conhecido popularmente como macaco Bugio e Saguis, respectivamente, apresentam-se potencialmente afetados pelo VFA, nos quais a doença apresenta elevada mortalidade, sinalizando um desequilíbrio importante para essas espécies (Moreno *et al.*, 2011). Além desses, o gênero *Ateles sp* (macaco aranha) também apresenta sensibilidade significativa ao VFA, entretanto são mais resistentes à doença, podendo curar-se adquirindo imunidade ativa.

Vale ressaltar a importância das epizootias de PNH como eventos sentinela para a vigilância da FA no Brasil, uma vez que a ocorrência da doença em PNH tem precedido a doença em seres humanos, sinalizando a atividade viral e, portanto, o risco do VFA infectar seres humanos (Silva *et al.*, 2020a).

Kaul e colaboradores (2018) e Childs e colaboradores (2019) relatam que a riqueza de primatas tanto em tamanho de populações como em diversidade, atua como impulsionador de transbordamento (*spillover*) de febre amarela.

A Mata Atlântica dentre os biomas presentes no estado de Minas Gerais é dos que mais enfrenta graves ameaças, onde 83% dos fragmentos possuem menos de 50 ha e são altamente isolados

(Ribeiro *et al.*, 2009).

Sabe-se que quanto menor o fragmento de floresta, maior é a perda de espécies, sendo que, aquelas que conseguem sobreviver, desenvolvem grande capacidade de se adaptar a esse ambiente em desequilíbrio. Essa alta fragmentação territorial atua nos processos ecológicos e demográficos das populações de PNH tanto no fluxo gênico como na dispersão espacial das espécies (Moraes *et al.*, 2018), resultando na diminuição da diversidade genética e no número de indivíduos de uma população e, conseqüentemente, no aumento do risco de endogamia (Wilcove *et al.*, 1986; Fahrig, 2003, Grativol *et al.*, 2008). Com isso pode-se ter tanto um cenário de resistência às doenças, dentre elas a FA, por seleção natural em espécies menos sensíveis ao VFA, favorecendo a permanência do vírus no ciclo silvestre (Grynszpan, 2017), como pode-se desencadear um processo irreversível de redução populacional, aumentando o risco de extinções locais em espécies mais sensíveis (Bicca-Marques, 2009; Bicca-Marques & Freitas, 2010).

Assim, o surto de FA no estado de Minas Gerais deve ser tratado com muita atenção, considerando-se também as questões ambientais, como a fragmentação das matas nativas, pois as espécies de PNH que habitam essas regiões, parecem se tornar potentes transmissoras e mantenedoras de agentes infecciosos (Cavalcante e Tauil, 2016).

Em relação aos seres humanos, hospedeiros acidentais da FA silvestre no Brasil, destaca-se a baixa cobertura vacinal dos residentes mineiros antes do surto. Segundo dados do Ministério da Saúde, em 2016, dos 533 municípios com recomendação para a vacina do Estado, apenas 25 (4,8%) tinham cobertura vacinal maior que 95%, sendo que 253 (47,4%) tinham menos de 50% de cobertura vacinal e 255 (47,8%) apresentaram cobertura entre 50 e 94,9% (Brasil, 2017b). Quando os primeiros casos de epizootias surgiram no Estado, a população e os serviços de saúde não associaram de imediato esse fato à doença, demonstrando pouca orientação, despreparo e, principalmente, ausência de memória cultural da FA (Valadares *et al.*, 2017).

4.3. Epidemiologia da febre amarela do Brasil

Nas duas últimas décadas, foram registradas transmissões de febre amarela fora dos limites da área considerada endêmica no país, a região Amazônica, caracterizando uma expansão nos sentidos leste e sul do país, que afetou áreas consideradas “índenes” até então, onde o vírus não era registrado há décadas (Ministério..., 2020).

Processos de reemergência da febre amarela geraram impactos relevantes na saúde pública. Importantes surtos humanos e epizootias em PNH das últimas décadas ocorreram entre 1998 e 2003 e entre 2007 e 2009. No primeiro ocorrem surtos na região Norte, Sudeste e Sul. No segundo período ocorreram surtos no Norte/Centro-Oeste, Sudeste e Sul do país (Fig. 5) (Ministério..., 2020).

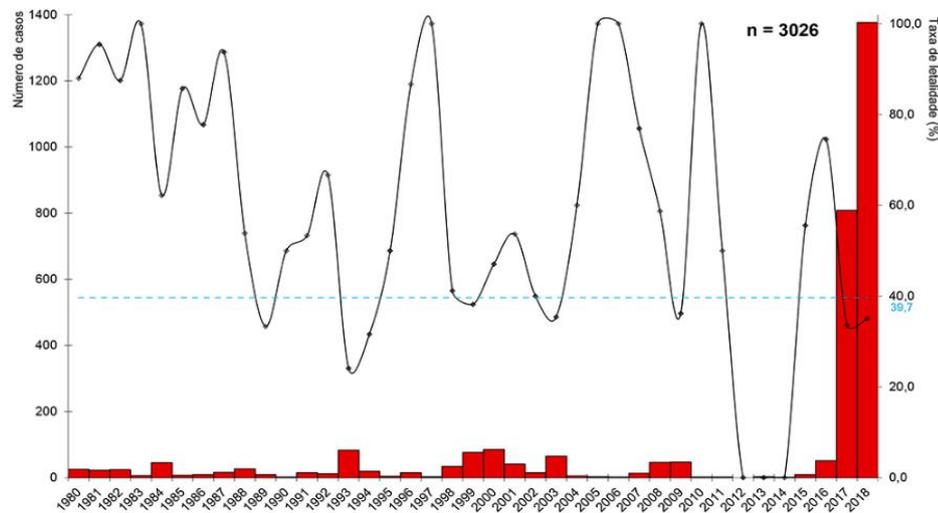


Figura 5. Série histórica do número de casos humanos confirmados para febre amarela e a taxa de letalidade, segundo o ano de início dos sintomas no Brasil, de 1980 a 26ª semana epidemiológica de 2018.

Fonte: Ministério da Saúde, 2021.

Na figura 5 observa-se em coluna vermelha o número de casos de febre amarela humana e a linha representa a taxa de letalidade da doença.

Em 1998, o território brasileiro foi dividido em três grandes áreas relacionadas à transmissão da febre amarela: endêmica, de transição e indene (Brasil, 2004). Entretanto foi sendo necessária uma reavaliação contínua dessas áreas, considerando o processo de circulação, expansão e manutenção do vírus da febre amarela (Fig. 6).

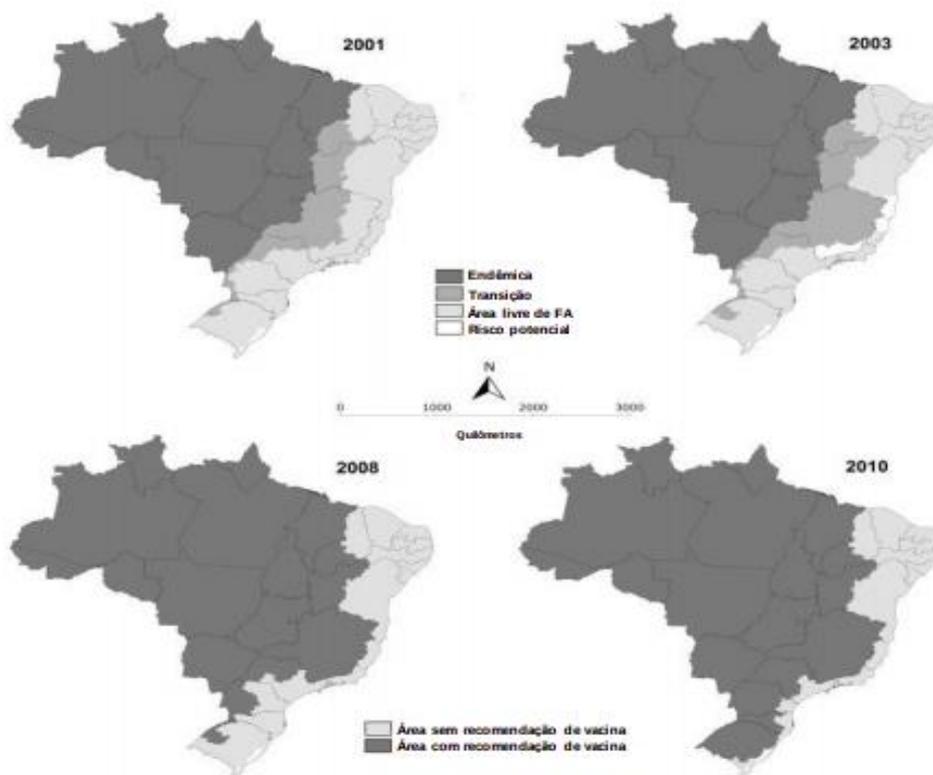


Figura 6. Evolução geográfica do risco de transmissão de febre amarela no Brasil, de 2001 a 2010.

Fonte: Romano, 2014.

Em 2003, foi observado aumento do número de casos e expansão da área de transmissão da doença, com ativação de focos em áreas até então indenes. (Brasil, 2004; Mascheretti *et al.*, 2013).

A partir de 2007, o ciclo silvestre da febre amarela fora da região amazônica, demonstrou maior expansão da circulação viral no Brasil. As áreas atingidas foram as regiões Sudeste e Sul do país sendo preocupante em virtude da proximidade de grandes centros urbanos, densamente ocupados por uma população até então, não vacinada, conseqüentemente sem proteção contra a doença, além disso, caracterizam-se por apresentarem alta infestação por *Aedes aegypti* (Cavalcante e Tauil, 2016). Essa realidade trouxe à tona a discussão a respeito do risco da retomada da transmissão urbana da febre amarela no Brasil (Mascheretti *et al.*, 2013).

Atualmente, o Brasil é dividido pela vigilância da doença em áreas sem recomendação para vacinação (ASRV) e áreas com recomendação para vacinação (ACRV) sendo esta última ampliada para todo território nacional em 2019/2020 (Fig. 7) (Romano *et al.*, 2011; Brasil, 2020).

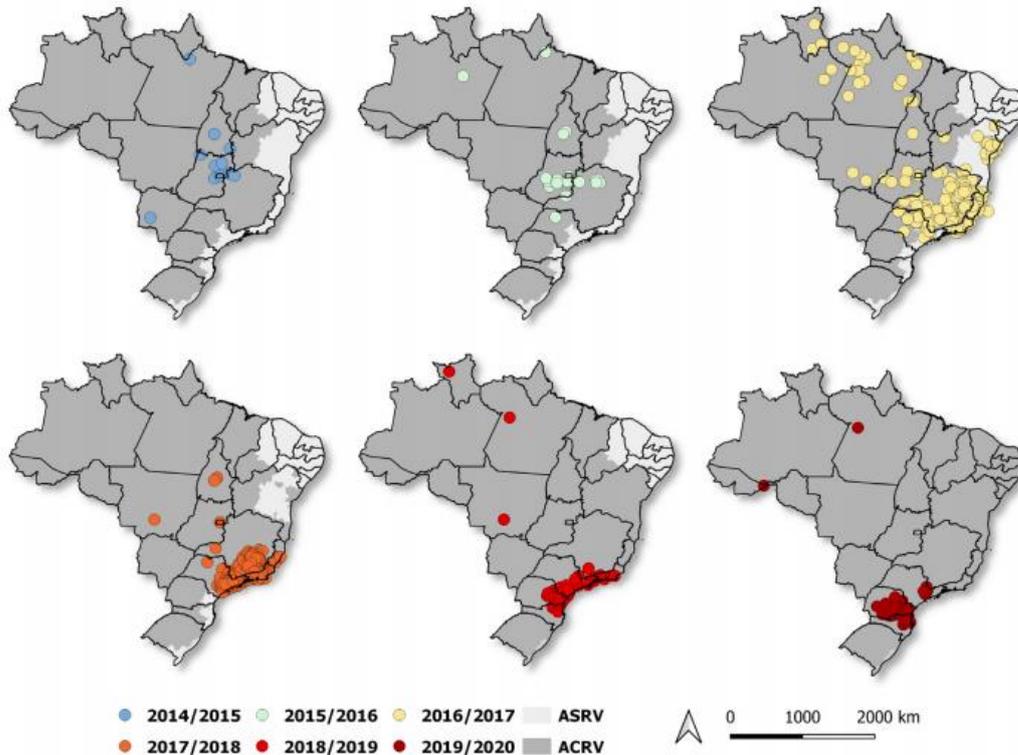


Figura 7. Reemergência da febre amarela no Brasil. Distribuição dos municípios com casos humanos e/ou epizootias em PNH confirmados durante reemergência extra-Amazônica da febre amarela, de acordo com o período de monitoramento, julho de 2014 a junho de 2020, Brasil.

Fonte: Ministério da Saúde, boletim epidemiológico 04, vol. 52, fev. 2021.

A maior parte dos casos de febre amarela, têm incidido entre dezembro e maio, período sazonal da doença, ocorrendo em regiões de mata, quando o vírus encontra condições favoráveis para a transmissão como: elevadas temperaturas, umidade e pluviosidade, tendo como consequência o aumento na densidade de vetores; presença de hospedeiros primários susceptíveis como primatas não humanos; presença de indivíduos susceptíveis seja por motivo laboral ou lazer, associados a baixas coberturas vacinais (Brasil, 2017a).

Casos humanos e epizootias em primatas não humanos (PNH) foram registrados em uma ampla área do território nacional. Inicialmente, entre 2014 e 2015, a transmissão se deu na região norte, com posterior expansão no sentido leste e sul do país, onde afetou prioritariamente a região centro-oeste entre 2015 e 2016 (Brasil, 2019b).

Mais recentemente, no período sazonal de 2016/2017 e 2017/2018, foram registrados os surtos mais expressivos de febre amarela da atualidade no Brasil. Entre 2016/2017 o surto afetou principalmente os estados da região Sudeste, quando foram registrados 779 casos humanos e 262 óbitos, além de 1.659 epizootias por FA no Brasil. Os estados da região Sudeste foram os mais acometidos, quando 1.376 casos humanos e 483 óbitos foram notificados pelos serviços de saúde, além de 864 epizootias em primatas não humanos, confirmadas para febre amarela (Brasil, 2019b). Em 2016/2017, a doença se espalhou por uma área na qual estima-se que residam 8 milhões de pessoas e, em 2017/2018, na área afetada os valores ultrapassaram os 32 milhões de indivíduos (Brasil, 2018a).

De junho de 2019 a janeiro de 2020 foi detectada a circulação do vírus amarílico em primatas não humanos nos estados de São Paulo, Paraná e Santa Catarina. A estratégia da vigilância tem utilizado a análise descritiva de corredores ecológicos pelos quais existe maior probabilidade de dispersão do vírus da febre amarela permitindo uma maior agilidade no bloqueio vacinal, mesmo antes da sinalização de epizootia em PNH (Brasil, 2020).

Entre julho de 2020 e janeiro de 2021 (SE-1), foram notificadas 574 epizootias em PNH suspeitas de FA, das quais 37 foram confirmadas por critério laboratorial. Detecções do vírus amarílico em PNH foram registradas no Distrito Federal (1), em Goiás (16), em São Paulo (1), no Paraná (11) e em Santa Catarina (8), sinalizando a circulação ativa do vírus nesses estados e o aumento do risco de transmissão às populações humanas durante o período sazonal (Brasil, 2021a).

No período, foram notificados 125 casos humanos suspeitos de FA, dos quais 107 foram descartados e 18 estão em investigação. Nenhum caso foi confirmado até então (Brasil, 2021a).

Detecções do vírus amarílico em PNH no Paraná e em Santa Catarina confirmam a expectativa de retomada da transmissão na região Sul, com possibilidade de dispersão para o Paraguai e a Argentina, já que as últimas detecções nesses estados ocorreram a cerca de 70 quilômetros da fronteira, requerendo ações de mobilização regional internacional. Registros de epizootias por FA em Goiás e no Distrito Federal alertam para a possibilidade de um novo processo de

reemergência do vírus a partir da região Amazônica (endêmica), com risco de dispersão para novas áreas, sobretudo da região Nordeste. Estudos genômicos estão em andamento para a identificação das linhagens presentes em cada um dos focos de transmissão, no sentido de compreender a dinâmica de dispersão e manutenção do vírus nessas áreas (Brasil, 2021a).

4.4. Epidemiologia da febre amarela em Minas Gerais

Em Minas Gerais, local de estudo do presente do trabalho, no ano 2000, ocorreram dois casos fatais de febre amarela em regiões distintas do estado. Posteriormente, dois importantes surtos foram registrados. O primeiro ocorreu em 12 municípios com 32 casos confirmados e 16 óbitos. O segundo atingiu 6 municípios com 64 casos confirmados e 23 óbitos. Após os dois surtos, foram registrados dois casos isolados: um no ano de 2008 na região noroeste do estado e um no ano de 2009 na Zona da Mata. No período de 2010 a 2015 não foram registrados casos humanos de febre amarela no estado de Minas Gerais (Minas Gerais, 2019).

Após um período de sete anos sem registro de casos humanos, Minas Gerais registrou duas ondas epidêmicas consecutivas. A primeira com início no final do ano de 2016, contabilizando 475 casos confirmados nos Vales do Rio Doce e Mucuri e em parte da Zona da Mata e Jequitinhonha em que 162 casos evoluíram para óbito. E o segundo, com 528 casos confirmados oriundos da região metropolitana de Belo Horizonte, zona da mata e partes das regiões de campos das vertentes, oeste e sul/sudoeste mineiros em que 177 casos evoluíram para óbito (Minas Gerais, 2019).

No período de monitoramento 2017/2018 (julho/2017 a junho/2018), até a semana epidemiológica (SE) 26, foram notificadas ao Ministério da Saúde 9.154 epizootias em PNH, das quais 3.477 foram descartadas, 3.508 foram indeterminadas (s/ coleta de amostras), 1.305 permanecem em investigação e 864 foram confirmadas por FA. Foram registradas epizootias em PNH confirmadas no Tocantins (4); no Mato Grosso (1); no Espírito Santo (2); no Rio de Janeiro (71), em Minas Gerais (119) e em São Paulo (667), com o maior número de epizootias confirmadas na região Sudeste (99,4%; 859/864) (Brasil, 2018a).

No período de monitoramento compreendido entre julho de 2018 e junho de 2019 não foram confirmados casos humanos de febre amarela silvestre no estado de Minas Gerais. Neste mesmo período foram notificadas epizootias em 167 municípios do estado de Minas Gerais, sendo que, apenas em um município a presença do vírus foi confirmada, em 119 municípios houve

epizootia em investigação e em 47 municípios a epizootia foi classificada como indeterminada por não ter sido possível proceder a coleta de amostras (Minas Gerais, 2019).

De acordo com o Boletim epidemiológico n. 12 de 2020, em Minas Gerais, entre os meses de junho de 2019 a junho de 2020, foram notificadas 239 epizootias sendo 20 descartadas, 182 indeterminadas e 37 ainda em investigação (Brasil, 2020).

Desde o ano de 1937 utiliza-se, no Brasil, a vacina que contém vírus atenuado, com a cepa 17-DD, procedente da amostra africana Asibi, desprovida de neuro e viscerotropismo e cultivada em ovos de galinha embrionados. A vacina, produzida pelo Laboratório Bio-Manguinhos, da Fundação Oswaldo Cruz (Ministério da Saúde), é bastante eficaz e segura. Seu efeito protetor inicia-se no décimo dia pós-vacinal e permanece por toda vida (Tauil, 2010; Brasil, 2019a).

No atual Calendário Nacional de Vacinação, a vacina contra febre amarela é indicada a partir dos nove meses de idade, tendo como meta mínima a ser atingida, 95% de cobertura vacinal. Em Minas Gerais todos os 853 municípios são classificados como área com recomendação para vacinação (ACRV) contra febre amarela desde o ano de 2008 (Minas Gerais, 2019).

De acordo com o último relatório da Secretaria do Estado de Saúde de Minas Gerais, a cobertura vacinal acumulada de febre amarela em Minas Gerais está em torno de 92,71%. Ainda havia uma estimativa de 2.726.228 pessoas não vacinadas em 2019. Entre os 853 municípios do Estado, 18,28% (156) deles não alcançaram 80% de cobertura vacinal; outros 35,75% (305) dos municípios têm entre 80% e 94,99% de sua população vacinada; com mais de 95%, estão 45,97% (392) das cidades mineiras (Minas Gerais, 2019).

Por meio de análises moleculares e epidemiológicas comprovou-se a origem de ciclo silvestre do VFA, o qual circulou nos surtos do Sudeste do Brasil a partir do ano de 2016, sendo esse, o mesmo que já havia circulado em Minas Gerais por volta de 2005 (Figueiredo *et al.*, 2018) e apesar desse estado da federação, em 2016, já ser considerado em sua completude como área com recomendação para vacina (ACRV), os casos humanos de FA ocorreram devido à grande parte dos municípios apresentarem baixa cobertura vacinal para a doença (Figueiredo *et al.*, 2018).

Análises moleculares realizadas por Rezende e colaboradores (2018) combinadas com registros epidemiológicos dos anos de 2016 a 2018, apoiam a ideia de que o VFA tenha sido introduzido, provavelmente, do Centro-Oeste para a região Sudeste, possivelmente no estado de São Paulo,

seguidos por Minas Gerais (Rezende *et al.*, 2018) e foram expandindo sua área de atuação na região Sudeste do país cobrindo uma faixa de mata Atlântica, com grande potencial de atingir áreas densamente povoadas pelo vetor urbano da doença, o *Aedes aegypti*, sendo esta a provável rota inicial das ondas epidêmicas de FA no Brasil.

Durante a duração das ondas epidêmicas de febre amarela, no Sudeste, entre os anos de 2016 e 2019, o VFA foi reintroduzido de Minas Gerais para estados do Espírito Santo e do Rio de Janeiro várias vezes (Giovanetti *et al.*, 2019).

O estado da Bahia, dentre as regiões afetadas pelos surtos de 2016 a 2019, foi o único que não apresentou casos humanos de FA. Observou-se, através de análise genômica do VFA em amostras de PNH, que houveram nesse estado (BA) duas introduções distintas, uma através do estado de Minas Gerais e outra pelo Espírito Santo, o que se justifica tanto pela distância entre as regiões de introdução, tempo e pela característica da vegetação predominantemente baixa dificultando a circulação dos primatas endêmicos na região, os quais dificilmente descem ao solo para migrar (Goes de Jesus *et al.*, 2020).

Análises filogenéticas do VFA em PNH confirmaram a presença do vírus em diferentes áreas urbanas de Minas Gerais, próximas a populações humanas, os quais apresentavam a mesma linhagem indicando a disseminação do VFA das áreas rurais / silvestres para as áreas urbanas durante os surtos, com alto de risco de reintrodução da FA urbana no país (Sacchetto *et al.*, 2020).

Outro fato importante a respeito do VFA foi a observação de uma assinatura molecular distinta, em que alterações em aminoácidos localizaram-se nas duas proteínas importantes do complexo de replicase viral (NS3 e NS5) e podem estar associadas a alguma vantagem seletiva, relacionada a uma maior capacidade do vírus de infectar hospedeiros vertebrados e vetores invertebrados (Bonaldo *et al.*, 2017). Assim, essas particularidades virais podem ter desempenhado um papel na aceleração da disseminação e gravidade do surto em curso (Possas *et al.*, 2018), entretanto, não se sabe se as alterações encontradas são exclusivas das cepas pertencentes ao surto iniciado em 2016 (Bonaldo *et al.*, 2017).

A persistência do VFA na região Sudeste, causando surtos de 2016 a 2018, sugere que esta região apresenta condições ecológicas e climáticas adequadas para a sua manutenção durante as estações epidêmica e inter-epidêmica. Este fato apresenta riscos para o estabelecimento de ciclos enzoóticos e epidemias de FA fora da Bacia Amazônica no Brasil, em áreas densamente

povoadas e densamente habitadas por *Aedes aegypti*, vetor urbano da doença, principalmente na região de Mata Atlântica.

O VFA encontrou no estado de Minas Gerais um ambiente com fatores eco-epidemiológicos ideais para sua multiplicação e expansão, já que a onda iniciada a partir 2016, já é considerada a de maior magnitude em 100 anos, com 2.050 casos humanos confirmados, 681 casos que terminaram em óbito e 764 casos confirmados de epizootia em PNH, a qual teve início neste estado da federação (Goes de Jesus et al., 2020).

4.5. Vigilância da febre amarela no Brasil

A observação de um padrão sazonal de ocorrência de casos humanos a partir da análise da série histórica da febre amarela no Brasil deu suporte à adoção da estratégia de vigilância baseada na sazonalidade (Ministério..., 2020).

O período anual de monitoramento da FA inicia em julho e encerra em junho do ano seguinte, de modo que os processos de transmissão que irrompem durante os períodos sazonais (dezembro a maio) possam ser analisados à luz das especificidades de cada evento (Ministério..., 2020).

A vigilância da febre amarela encontra-se normatizada pela Portaria nº 104/GM/MS, de 25 de janeiro de 2011, que a classifica como uma doença de notificação compulsória e imediata, devendo esta, ser realizada em no máximo 24 horas a partir da suspeita de caso inicial humano e pela Portaria nº 782 MS/GM de 15 de março de 2017, que define as epizootias de primata não humano como de notificação compulsória e suas diretrizes para notificação no Brasil ocorrem através da ficha de notificação do SINAN (Sistema de Informação de Agravos de Notificação) nas esferas do SUS (Romano, 2011, Brasil, 2017c). Além disso, a febre amarela é objeto de vigilância internacional de acordo com o Regulamento Sanitário Internacional da Organização Mundial de Saúde (OMS), devendo as notificações serem encaminhadas à Organização Pan-Americana de Saúde/OPAS através do Ministério da Saúde (Brasil, 2004).

Diante da informação da morte de um PNH, o qual atua como evento sentinela, além da notificação, inicia-se a investigação sobre a veracidade do ocorrido e a magnitude do evento (número de animais acometidos, área de ocorrência, causas suspeitas). Os achados iniciais da investigação devem ser consolidados e encaminhados para as diferentes esferas do SUS (até 24

horas), visando favorecer a tomada de decisão articulada entre União, Estados e Municípios (Romano, 2011; Brasil, 2017c).

Sendo a febre amarela silvestre uma zoonose, sua transmissão não é passível de eliminação, necessitando de vigilância e manutenção das ações de controle, especialmente por meio de cobertura vacinal humana adequada. Uma pessoa com febre amarela silvestre pode vir a ser fonte de infecção e desencadear um surto da chamada febre amarela urbana, transmitida principalmente pelo *Aedes aegypti* (Brasil, 2017a).

A entomologia aplicada à vigilância de casos humanos e epizootias suspeitas de febre amarela, se dá de forma ativa quando as ações que se baseiam no monitoramento de áreas estratégicas (sentinelas e vulneráveis/receptivas), com o intuito de acompanhar espacial e temporalmente populações de culicídeos potencialmente vetores, detectar alterações ecológicas que possam indicar risco de transmissão, detectar precocemente a circulação viral e definir áreas com potencial de transmissão (receptivas). Já a forma passiva, por sua vez, refere-se às atividades desencadeadas por ocasião de notificações de casos humanos ou epizootias em primatas não humanos suspeitos de febre amarela, a partir das quais são desencadeadas medidas de bloqueio de transmissão. O objetivo é apoiar a investigação no local de ocorrência e contribuir para pesquisa de vírus que permita atribuir causa por vínculo epidemiológico por meio de levantamento de dados (Brasil, 2017c).

Dentre os grandes desafios da vigilância e controle da febre amarela no Brasil podemos destacar a redução da febre amarela silvestre em humanos e manter nula a incidência de casos do ciclo urbano, isto é, prevenir a reurbanização da doença (Tauil, 2010). Para tal, deve-se fortalecer e aprimorar a vigilância epidemiológica a fim de detectar a circulação viral por meio da vigilância de epizootias e entomológica. Identificar e definir casos suspeitos, confirmados e descartados da febre amarela em humanos, trabalhar junto à população com medidas sócio-educativas e divulgação de dados sobre a doença no país (Brasil, 2004).

A incorporação de novas ferramentas de vigilância tem contribuído para a melhoria da qualidade dos dados de vigilância e para a maior oportunidade da notificação e investigação de eventos suspeitos. Nesse sentido, a implementação do SISS-Geo para a notificação de epizootias em PNH tem possibilitado a notificação simultânea e em tempo real de eventos para todas as instâncias administrativas (municipal, regional, estadual e federal), além de reduzir lacunas de informações sobre os eventos notificados. Durante o monitoramento 2019/2020, o

SISS-Geo foi utilizado pelos estados da região Sul (PR, SC e RS), com redução de 82,9% na proporção de eventos sem identificação do gênero/espécie dos animais envolvidos, e de 76,5% na proporção de eventos sem registro das coordenadas geográficas. Atualmente, o uso do SISS-Geo está em processo de implantação em diversos estados (TO, GO, DF, MT, MS, MG e SP), com perspectiva de implantação em todo o país. Do mesmo modo, um formulário eletrônico (FormSUS 2/REDCap) de notificação de casos humanos suspeitos tem sido utilizado em conjunto com os entes estaduais para a notificação e atualização de dados epidemiológicos em tempo real, sobretudo nos estados com registro da presença do vírus, no sentido de promover a tomada de decisão integrada, coordenada e oportuna entre todas as esferas de gestão (Brasil, 2021a).

5. REVISÃO DE LITERATURA - MODELOS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE DADOS

5.1. Aspectos gerais

Aprendizado de máquina é um ramo da inteligência artificial que permite que sistemas de computador aprendam diretamente com exemplos, dados e experiência de forma a realizar tarefas complexas (Vopham *et al*, 2018). Sendo um conceito abrangente, aprendizado de máquina engloba uma grande diversidade de modelos e estratégias que envolvem modelagem algorítmica (Royal society, 2017).

Tanto ferramentas de modelagem na bioestatística e epidemiologia quanto aprendizado de máquina, requerem análises de dados para compreensão de fenômenos, e apesar das abordagens serem diferentes, com a bioestatística trabalhando com modelos frequentistas e de testes de hipóteses, *machine learning* com modelagem algorítmica, a combinação dessas abordagens parecem se complementar de forma a potencializar as análises de dados em saúde (Wiemken e Kelly, 2020).

O fluxo de trabalho da ciência de dados geralmente se assemelha a um processo iterativo de importação e processamento de dados, seguido por limpeza, transformação, visualização, modelagem e, finalmente, comunicação de resultados (Wickham e Golemund, 2016). A crescente popularidade dos modelos de aprendizado de máquina pode ser atribuída às suas habilidades para aproximar quase todo relacionamento funcional não linear complexo.

A capacidade de os modelos de aprendizado de máquina mapearem relações não lineares entre as variáveis que compõe situações do mundo real fazem deles importantes ferramentas tanto para classificação (para variável de saída categórica) quanto para regressão (para variável de saída numérica) (James *et al*, 2013; Hastie *et al*, 2008).

A análise preditiva, objeto de trabalho do presente estudo, consiste na aplicação de algoritmos para compreender a estrutura dos dados existentes e gerar regras de predição. Esses algoritmos podem ser utilizados em problemas supervisionados, quando, além dos preditores, estão disponíveis também uma resposta de interesse, responsável por guiar a análise (Hastie *et al.*, 2008).

5.2 Séries temporais

Uma série temporal é uma sequência de observações de determinada variável de interesse, geralmente, observada em pontos discretos do tempo. Uma análise de séries temporais consiste em descrever o processo ou fenômeno que gera essa sequência com o objetivo de entender o comportamento da série, realizar previsões e, ocasionalmente, controlá-la (Ehlers, 2009).

De acordo com Rouquayrol (2018) as variáveis ligadas à distribuição de casos de uma doença em função do tempo, formam com as ligadas a pessoas e lugares o tripé da epidemiologia descritiva. A organização adequada desses dados pode oferecer um diagnóstico dinâmico de doenças e agravos na população.

O estudo da distribuição das doenças no tempo pode oferecer diversas informações para apontar riscos aos quais populações podem estar expostas, monitoramento da saúde de populações específicas, previsão de ocorrência de eventos, auxiliar o planejamento da saúde e avaliar impactos das intervenções (Rouquayrol, 2018)

Em um conjunto de dados temporais existe uma ordem de dependência explícita entre as observações: a dimensão do tempo. O termo série temporal pode ser definido, portanto, como sequência de observações ordenadas no tempo (Brownlee, 2017).

Seja, então, $\{Z_t, t \in T\}$ uma série temporal em que T representa os tempos t em que a série foi observada. Em dados longitudinais, coletados ao longo de um tempo determinado, uma importante característica é que as observações que possuem proximidade temporal são dependentes e o interesse consiste em analisar e modelar esta dependência, buscando a utilização de ferramentas específicas para este tipo de dado (Ehlers, 2009).

A relação entre uma sequência de marcos cronológicos sucessivos e uma variável de frequência constitui numa distribuição de frequências de casos. Na maioria dos estudos epidemiológicos tomam-se como marco cronológico os anos do calendário, entretanto, meses, semanas epidemiológicas e dias também podem ser utilizados (Rouquayrol, 2018).

Predição de séries temporais é uma importante área de aprendizado de máquina que por vezes prioriza tarefas de classificação e regressão. Sua relevância está relacionada à componente temporal envolver diversos problemas de previsão. A dificuldade de se trabalhar previsão com componente temporal faz desse tipo de problema um desafio na área de aprendizado de máquina (Brownlee, 2017).

Modelos de aprendizado de máquina vem sendo amplamente utilizados na predição de séries temporais com dados de saúde como na previsão de casos de COVID-19 (Gumaei et al., 2021; Goo *et al.*, 2021), ebola (Pandey e Subbian, 2017), dengue (Mussumeci e Coelho, 2020), demanda de sangue (Shih e Rajenran, 2019) e vagas em atendimento hospitalar (Guan, 2019).

Um problema de série temporal pode ser tratado como aprendizado de máquina supervisionado usando passos de tempo anteriores como variáveis de entrada e passos futuros como variáveis de saída. A esse método dá-se o nome de janela deslizante ou método lag (*sliding window*), que busca trazer uma correlação entre os valores das linhas aos modelos de aprendizado de máquina permitindo que os mesmos reconheçam padrões existentes entre os valores da série temporal. O número de casos de febre amarela do dia anterior ($t-1$) é apresentado para o modelo (Brownlee, 2017).

Séries temporais podem ser univariadas quando uma única observação é observada no tempo, e multivariadas quando mais de uma observação é observada em um instante t do tempo (Brownlee, 2017).

5.3. Paradigmas de aprendizado

A aprendizagem é o processo pelo qual os parâmetros de um modelo de aprendizado de máquina são ajustados através de uma forma contínua de estímulos pelo ambiente no qual o modelo está operando.

Um conjunto de procedimentos bem definidos para adaptar os parâmetros de um modelo, para que o mesmo possa aprender uma determinada função é chamado de algoritmo de aprendizado (Braga *et al.*, 2000).

A resolução de uma determinada tarefa passa pela fase de aprendizagem, quando algoritmo extrai informações relevantes de padrões de informações apresentados ao modelo criando assim, uma representação própria do problema. Essa fase consiste em um processo iterativo de ajustes de parâmetros, os quais ao final do processo guardam o conhecimento adquirido do ambiente em que está operando (Braga *et al.*, 2000).

Os três principais tipos de aprendizado são aprendizado supervisionado, não supervisionado e por reforço.

5.3.1. Aprendizado supervisionado

No aprendizado supervisionado os pares de entrada de saída desejada são fornecidos ao modelo por um supervisor externo (professor). O objetivo é ajustar os parâmetros do modelo, de forma a encontrar uma ligação entre os pares de entrada e saída (Braga *et al.*, 2000).

O aprendizado supervisionado tem como objetivo o ajuste os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. A rede tem sua saída corrente calculada comparada com saída desejada (padrão ótimo a ser realizado pelo modelo), sendo então mostrado o erro da resposta atual e ajustando-se os pesos das conexões até que o erro mínimo seja aceitável (Fig. 8) (Braga *et al.*, 2000).

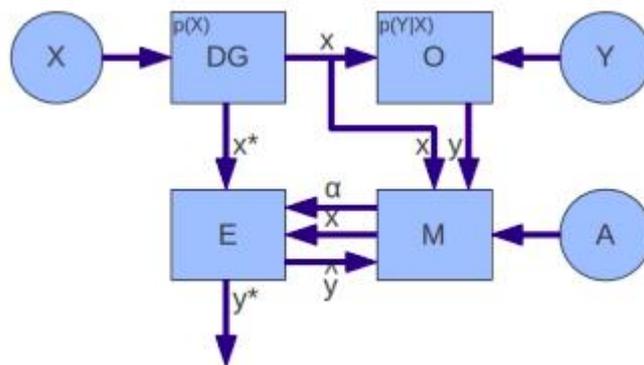


Figura 8: Esquema representativo do aprendizado supervisionado

Fonte: Coelho (2013).

Na Figura 8, há um bloco, denominado DG, representando a figura do Gerador dos Dados, que fornece os dados de acordo com uma distribuição $p(x)$. O conjunto de dados \mathbf{X} é uma amostra de DG e é entregue a O (oráculo) que, para cada observação neste conjunto de dados, conhece a variável de destino y . Em outras palavras, dado x , o oráculo, de alguma forma, conhece o alvo y , portanto, uma aproximação da distribuição condicional $p(y|x)$ é conhecido. Um Modelo M pode ser definido com os dados x e o alvo variável y , a fim de ser usado por um Estimador E para classificar ou prever a variável alvo y^* para um novo dado x^* , dado um conjunto A de parâmetros α . O Modelo M pode ser alimentado com a estimativa de saída \hat{y} , a fim de ajudar no ajuste fino dos parâmetros α .

5.3.2. Aprendizado não supervisionado

Aprendizado não supervisionado são normalmente usados para descobrir padrões desconhecidos em dados. Nesse caso, somente os padrões de entrada estão disponíveis para a rede, ao contrário do aprendizado supervisionado, cujo conjunto de treinamento possui pares de entrada e saída (Fig. 9) (Braga *et al.*, 2000).

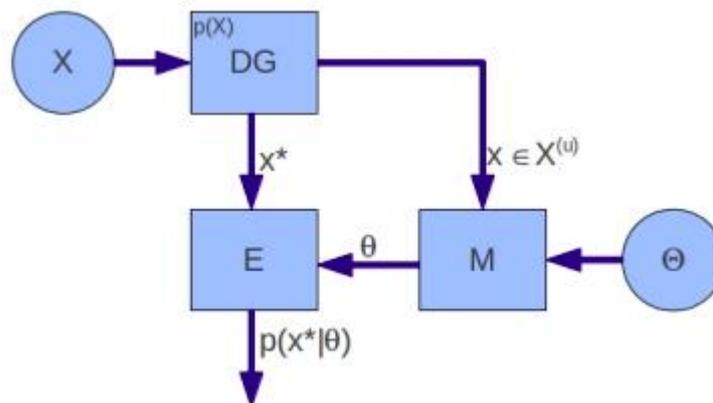


Figura 9: Esquema ilustrativo do aprendizado não supervisionado

Fonte: Coelho (2013)

Na Figura 9, pode-se observar que a probabilidade condicional $p(y|x)$ não é conhecida, ou seja, não há agente de rotulagem (oráculo). Considerando isso, o modelo M tem que ser construído com habilidade suficiente para extrair informações estruturais de \mathbf{X} , a fim de permitir que o estimador E estime a probabilidade $p(x^*|\theta)$ de um novo dado x^* pertencer ou não a um grupo específico (dado um conjunto de parâmetros Θ).

5.3.3. Aprendizado por reforço

O aprendizado por reforço pode ser considerado um caso particular de aprendizado supervisionado. No aprendizado por reforço a única informação de realimentação fornecida à rede é se uma determinada saída está correta ou não, isto é, não é fornecida à rede a resposta correta para o padrão de entrada e sim um sinal de reforço. Quando o sinal de reforço diz que uma saída está errada, ele não dá nenhuma sugestão de qual resposta poderia ser a correta (Braga *et al.*, 2000).

5.4. Pré processamento dos dados

Dados brutos raramente vem na forma necessária para o desempenho ideal do algoritmo de aprendizado de máquina. Sendo assim, o pré-processamento dos dados se torna um passo crucial para aplicação dos algoritmos (Raschka, 2019).

No pré-processamento dos dados são realizadas as seguintes atividades: (1) definição e separação dos conjuntos de treinamento, validação e teste; (2) transformação de variáveis quantitativas (via padronização ou normalização); (3) redução de dimensionalidade do conjunto de dados (exclusão de preditores altamente correlacionados e/ou utilização de algoritmos como Boruta); (4) exclusão de variáveis/observações com dados faltantes ou utilização de técnicas de imputação (média, mediana ou algoritmos de imputação); (5) organização de variáveis qualitativas (decomposição das variáveis categóricas em um conjunto de variáveis indicadoras que serão utilizadas como preditores).

No pré-processamento, a meta principal é treinar os modelos como um banco de dados, mais completo e organizado para que as previsões sejam mais próximas da realidade possível.

É importante destacar que parâmetros estimados durante a etapa de pré-processamento como padronização dos preditores ou redução da dimensionalidade devem ser obtidos no conjunto de treinamento, somente depois aplicados ao conjunto de teste (Fig. 10).

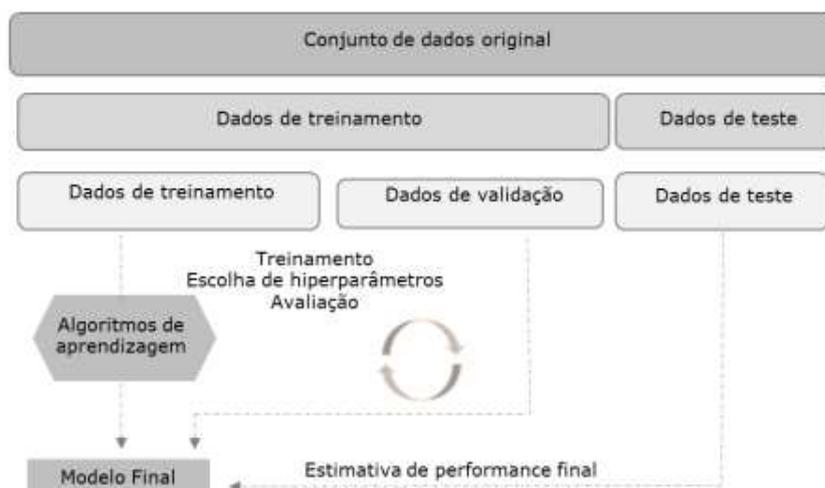


Figura 10: Esquema representativo das fases que compreendem o uso de algoritmos de aprendizado de máquina.

Fonte: Santos, 2018, adaptado de Raschka, 2017.

Inicialmente separa-se o conjunto de teste e treinamento. Quando o conjunto de dados for grande o suficiente, pode-se separar uma parte dos dados de treinamento para validação do modelo. Na fase de treino e validação estima-se os hiper parâmetros do modelo de aprendizado de máquina e na fase de teste estima-se a performance do modelo final frente aos dados trabalhados.

5.4.1. Padronização dos dados

À medida que a variabilidade das categorias de dados aumenta, podemos ter diferentes grandezas nos trabalhos que envolvem diferentes bancos de dados. Busca-se, portanto, transformar todas as variáveis na mesma ordem de grandeza, ou seja, trazer os valores para a mesma escala de valores.

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x} \quad (1)$$

Os dados são padronizados pelo valor observado menos a média/desvio padrão, para evitar a interferência das diferentes unidades de medida nos cálculos (James *et al.*, 2019).

5.4.2. Análise de correlação de Pearson

A permanência de preditores altamente correlacionados, e que, portanto, estão medindo a mesma informação podem contribuir para o mau desempenho do modelo.

A correlação de Pearson para uma variável de um dataset é dada por:

$$r = \frac{\sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}}_i)(\hat{x}_i - \bar{\hat{x}}_i)}{\sqrt{\sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}}_i)^2 \sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}}_i)^2}} \quad (2)$$

5.4.3. Imputação de dados ausentes em séries meteorológicas

Um dos fatores mais importantes para conhecimento e estudo dos fenômenos meteorológicos são os dados observados. À medida que os dados coletados pelas estações meteorológicas fornecem registros contínuos, é possível entender de forma mais real o fenômeno estudado. Entretanto, existem situações em que nos deparamos com dados faltantes nas bases disponíveis, o que pode afetar a qualidade do estudo. Situações como falta de calibração do aparelho, deterioração do aparelho de aferição, pane, deficiências no procedimento de observação e medição podem ser causas de ausência de dados na série (Ruiz-Cárdenas e Krainski, 2011). Deve-se, portanto, encontrar meios de se repor de forma mais precisa possível esses dados faltantes, sejam com dados de estações próximas ou por meio de técnicas estatísticas para minimizar ou até mesmo excluir de análises futuras os possíveis vieses que possam ocorrer em virtude desse problema.

A substituição dos valores faltantes por valores plausíveis, que são os valores imputados, é uma técnica conhecida como imputação de dados, a qual tem se mostrado eficiente (Harrell, 2001).

A imputação consiste na substituição de dados incompletos por valores calculados através dos dados observados. Na imputação única utiliza-se a média, a mediana da variável, métodos de interpolação ou até mesmo regressão linear para substituir os dados faltantes. Com o passar do tempo foram surgindo técnicas de imputação múltipla para corrigir incertezas associadas à imputação única.

Imputação múltipla

Os softwares estatísticos modernos e atuais, se enquadram como recursos com capacidade para preenchimento de lacunas. A Imputação Múltipla (IM), tem se mostrado mais eficaz, gerando

conjuntos completos de dados, preenchendo os valores omissos por meio de modelos de imputação, geralmente mais precisos do que aqueles fornecidos pelos métodos de imputação única (Little e Rubin, 2015).

Segundo Harrell (2001), deve-se utilizar diferentes técnicas para imputar dados em pesquisas de acordo com o percentual de dados faltantes:

Percentual $\leq 0,05$: Aplica-se a imputação única ou analisar somente os dados completos;

Percentual entre 0,05 e 0,15: Imputação única pode ser usada aqui provavelmente sem problemas, entretanto o uso da IM é indicado;

Percentual $\geq 0,15$: A IM é indicada na maior parte dos casos.

De acordo com Van Buuren e Groothuis-Oudshoorn (2011) que implementaram em R o pacote o MICE, a IM baseia-se na Cadeia de Markov Monte Carlo (MCMC- “Markov chain Monte Carlo”), cujo algoritmo é um amostrador de Gibbs, técnica de simulação Bayesiana, que amostra distribuições condicionais com a finalidade de obter amostras da distribuição conjunta (Van Buuren; Groothuis-Oudshoorn, 2011; Espinosa *et al.*, 2019).

5.4.4. Seleção de variáveis, atributos ou dados de entrada

A seleção de variáveis ou recursos geralmente é uma importante etapa na aplicação de métodos de aprendizado de máquina. Normalmente, quando um grande número de variáveis é apresentado ao modelo, uma parte significativa pode se apresentar como irrelevante para previsão e obviamente sua relevância não é conhecida antecipadamente (Kursa e Rudnicki, 2010).

Lidar com grande número de variáveis ou atributos podem tornam os algoritmos mais lentos e muitos algoritmos de aprendizado de máquina apresentam uma diminuição de precisão quando o número de variáveis é significativamente maior do que o ideal (Kursa e Rudnicki, 2010).

Trabalhar com um conjunto de recursos minimamente possível tendo melhores resultados é conhecido como problema mínimo-ideal (Nilsson *et al.*, 2007).

Boruta

No mundo real, a maioria dos fenômenos apresenta características dinâmicas, estocásticas, não lineares e principalmente dependentes de N variáveis, que interagem entre si, provocando diversas consequências no fenômeno de interesse.

A qualidade da previsão de séries temporais por qualquer modelo de aprendizado de máquina está fortemente associada à correta escolha das variáveis explicativas ou independentes ou ainda variáveis de entrada. Caso essas variáveis não sejam suficientemente e verdadeiramente representativas do problema a ser modelado, a capacidade de ajuste e previsão do modelo fica comprometida (Kursa & Rudnicki, 2010).

O algoritmo Boruta (Kursa e Rudnicki, 2010) é um seletor de variáveis do tipo *wrapper*, que utiliza subconjuntos ou pacotes de um banco de dados para o treinamento do modelo. A partir de um resultado preliminar, o algoritmo decide se uma ou mais variáveis devem ser removidas do subconjunto (Kursa & Rudnicki, 2010).

Ao tratar de seleção de variáveis pelo Boruta, a medida de importância de uma variável em relação a outra é obtida pelo cálculo da perda de acurácia do modelo, que ocorre por conta da permutação aleatória dos valores das variáveis entre as observações. A variável dependente, no caso deste estudo a notificação de febre amarela humana é avaliada a partir do cálculo da importância das variáveis independentes (Kursa & Rudnicki, 2010).

O boruta baseia-se na ideia de variáveis sombra, as quais são obtidas por meio da cópia das observações do banco de dados para uma determinada variável e o embaralhamento aleatório dos valores entre as linhas. Com isso, todas as correlações aleatórias entre a variável sombra e a variável de análise são perdidas. A classificação, pela floresta aleatória, no algoritmo Boruta é realizada neste banco de dados estendido com cópias embaralhadas das variáveis (Kursa & Rudnicki, 2010). A importância calculada, pelo parâmetro Z , é comparada entre as colunas originais e as duplicadas.

5.5. Etapa de treinamento

Todos algoritmos de aprendizado de máquina exigem treinamento com dados de entrada. O formato exato pode diferir, mas em sua forma básica os dados de entrada assumem a forma de exemplos (Lantz, 2013).

Ao selecionar a proporção de dados para um conjunto de treinamento, as principais considerações incluem: o número de variáveis, o tamanho da amostra e o próprio grau de dispersão das variáveis. A importância está em quão bem o conjunto de dados de treinamento descreve todos os padrões possíveis de dados. Nessa fase, os modelos aprendem com seus erros e vão se tornando à medida que o treinamento progride, de forma iterativa, para sua versão mais aperfeiçoada e capaz de executar o que lhe foi proposto (Wiemken e Kelly, 2020).

É na fase de treinamento que os parâmetros e hiperparâmetros dos modelos são ajustados e na fase de teste avalia-se o equilíbrio entre viés e variância.

5.6. Validação e teste do modelo

Se houver um conjunto de dados suficientemente grande, é recomendado subdividir o conjunto de treinamento em dois subconjuntos: treinamento e validação (Hastie *et al.*, 2008). A fase de validação consiste em testar o desempenho do modelo de aprendizado de máquina, já ajustado nos seus parâmetros e hiperparâmetros. Em RNA, Braga e colaboradores (2000) citam que o treinamento deve ser interrompido quando o erro do conjunto de validação começa a subir, ou seja, quando a rede começa a incorporar o ruído presente nos dados o que causa degradação na sua capacidade de generalização. Essa alternativa se mostra eficiente quando o conjunto de dados muito grande está disponível, já que os dados do conjunto de validação não podem ser utilizados para treinamento.

A fase de teste consiste em apresentar ao modelo já ajustado e validado o conjunto de dados que ele não conhece. Apresenta-se somente as variáveis de entrada e espera que o modelo possa prever ou classificar a variável resposta o mais próximo possível da realidade, avaliando-se o erro de generalização.

5.7. Dilema viés variância

O dilema entre o viés e a variância (Geman *et al.*, 1992) trata da decomposição da esperança do erro de generalização em três parcelas: o viés, a variância e o erro irreduzível.

Para compreender o impacto de cada parcela é necessário definir a capacidade de um modelo. Esta última está relacionada ao grau de flexibilidade do modelo, ditando o nível de complexidade da função de aproximação que foi aprendida.

A Capacidade de um modelo depende da quantidade de amostras contidas no conjunto de dados de treinamento e da complexidade das funções implementadas (Hastie *et al.*, 2008).

Quando se utiliza uma capacidade maior que a requerida para um determinado problema, pode-se obter modelos com baixo viés e grande variância, acarretando em um sobreajuste do modelo. Tais características estão relacionadas a modelos que se ajustam bem aos dados de treinamento, mas que apresentam desempenho ruim quando aplicados a novas observações, ou seja, apresentam baixa capacidade de generalização (*overfitting*). Já o uso de uma capacidade inferior ao requisitado pelo problema, pode-se resultar em um modelo com alto viés e pequena variância, levando ao efeito de sub ajuste do modelo (Hastie *et al.*, 2008) o que pode ser observado abaixo (Fig. 11).

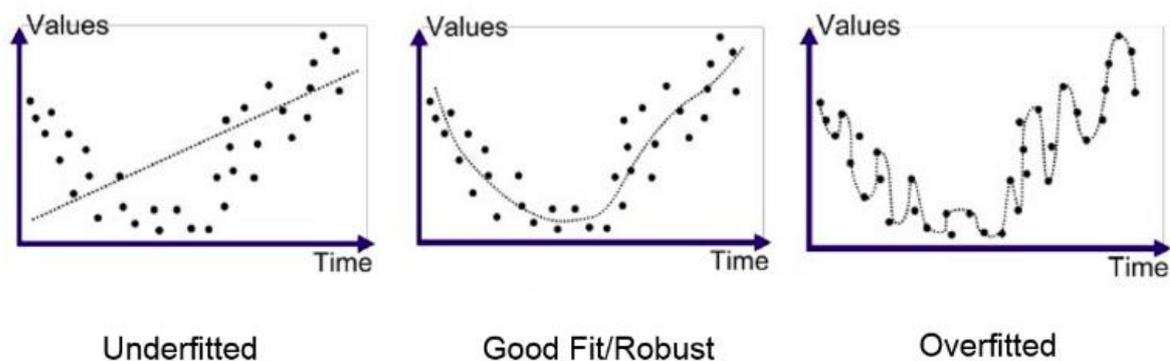


Figura 11: Representação gráfica de *Underfitting*, *Good Fit* e *Overfitting*

Fonte: abraacd.org

O aprendizado de modelos de aprendizado de máquina preditivos é norteado principalmente por dois objetivos: a seleção e a avaliação do modelo.

Na seleção, a performance de diferentes modelos baseados em variações dos valores para os hiperparâmetros, é avaliada para selecionar aquele que resulte em melhor desempenho (equilíbrio entre viés-variância). Não há um único algoritmo capaz de apresentar bom desempenho para todas as aplicações, sendo que a comparação entre modelos diante de um problema para selecionar o mais adequado é um prudente caminho (James *et al.*, 2014).

Na avaliação do modelo, após a seleção, busca-se estimar seu erro de predição (erro de generalização) em novas observações (Hastie *et al.*, 2008).

5.8. Hiperparâmetros de algoritmos de aprendizagem

A análise preditiva consiste na aplicação de algoritmos para compreender a estrutura dos dados existentes e gerar regras de predição. Esses algoritmos podem ser utilizados em um cenário não supervisionado, em que apenas preditores (covariáveis) estão disponíveis no conjunto de dados, ou em problemas supervisionados, quando, além dos preditores, está disponível também uma resposta de interesse, responsável por guiar a análise.

Em aprendizado de máquina, a maioria dos algoritmos apresentam um ou mais parâmetros que controlam a complexidade do modelo ajustado, ou seja, o equilíbrio entre viés e variância. Esses parâmetros chamados de hiperparâmetros podem ser especificados diretamente antes do ajuste do modelo. Escolhas inadequadas podem resultar em modelos sobrejustados (*overfitting*).

O objetivo de se avaliar as configurações de um modelo é descobrir a configuração que apresenta o melhor desempenho esperado com dados não vistos (Sanderson *et al*, 2020). A maioria dos algoritmos de aprendizado de máquina têm vários hiperparâmetros que devem ser ajustados para fornecer um resultado válido e preciso.

Um hiperparâmetro é um parâmetro do algoritmo cujo valor é usado para controlar o processo de aprendizado. Dependendo dos valores dos hiperparâmetros podemos ter resultados melhores ou piores nas tarefas que pretendemos realizar, demandando mais ou menos tempo, sendo, portanto, o ajuste dos hiperparâmetros etapa fundamental na modelagem algorítmica de aprendizado de máquina (Wiemken e Kelly, 2020).

Embora a maioria dos modelos de aprendizado de máquina tenha valores padrão para cada hiperparâmetro, vale a pena otimizar esses parâmetros.

Existem muitas heurísticas para determinar o quanto dos dados devem ser utilizados para ajustar esses parâmetros, mas não há consenso. Recomenda-se que aproximadamente 50% dos dados disponíveis sejam selecionados aleatoriamente para ajuste de hiperparâmetros. Esta é apenas uma heurística genérica e deve ser modificada com base na variação presente nas características e no resultado (Wiemken e Kelly, 2020).

Os modelos de RNA têm vários hiperparâmetros que são definidos pelo modelador, incluindo o número de neurônios em cada camada oculta, o número de camadas ocultas, a taxa de aprendizado, função de ativação e o número de épocas. A taxa de aprendizado é o quanto os parâmetros de peso são ajustados a cada iteração, e o número de épocas é o número de vezes

que todo o conjunto de dados de treinamento é usado para atualizar os parâmetros de peso (Sanderson *et al.*, 2020).

Segundo Braga e colaboradores (2000) para a definição do número de nodos ou neurônios da camada intermediária de uma RNA se dá em função do número de entradas e saídas da rede. Outra forma é utilizar um número de conexões dez vezes menor que o número de exemplos. A solução neural mais eficiente é aquela em que o número de unidades cresce apenas polinomialmente com o aumento do número de unidades de entrada.

As funções de ativação, também consideradas hiperparâmetros da RNA, mais utilizadas são tangente hiperbólica, sigmoideal e função logística.

Os hiperparâmetros do algoritmo de RF e algoritmo de Boosting incluem o número de árvores (estimadores) e a profundidade máxima da árvore, número máximo de folhas. A taxa de aprendizado também é um hiperparâmetro (Sanderson *et al.*, 2020).

Hiperparâmetros do SVM são conhecidos como C e Γ .

De acordo com Kou *et al.* (2010), o parâmetro C é usado para proporcionar equilíbrio entre erros no treinamento e complexidade do modelo, mas não tem um propósito específico, portanto é difícil definir seu valor”. A presença do parâmetro C encontra-se tanto no SVM linear quanto no não-linear, e possui a responsabilidade de controlar a tolerância a erros em um modelo de treinamento.

Γ é utilizado somente no SVM não-linear, por meio das funções de kernel. Ele controla a flexibilidade do classificador, de modo que a fronteira entre as classes tem a sua forma mais ajustada ao conjunto (Ben-Hur; Weston, 2008).

5.9. Redes neurais artificiais

As RNA são uma classe flexível de modelos de aprendizado de máquina que foram inspirados pela neurociência, os quais são compostos de camadas de neurônios, em que os neurônios de uma camada são conectados aos neurônios da próxima camada. Cada neurônio é uma unidade computacional que multiplica seus valores de entrada por um conjunto correspondente de parâmetros de peso, soma os valores multiplicados, transforma o valor somado usando uma função de ativação e produz o valor transformado (Fig. 12) (Sanderson *et al.*, 2020).

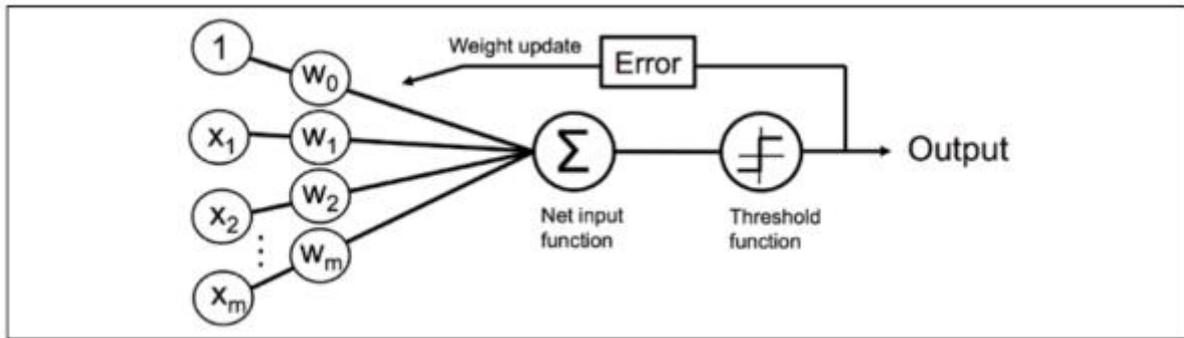


Figura 12: Diagrama que ilustra o conceito geral de perceptron

Fonte: Raschka, 2017.

A primeira camada em um modelo de RNA é a camada de entrada e cada unidade na camada de entrada contém o valor de um dos preditores para uma observação específica. A camada de entrada passa todos os valores preditores de uma observação específica para cada neurônio na primeira camada oculta. Cada neurônio na primeira camada oculta calcula uma função diferente com os valores do preditor. A primeira camada oculta então passa seus valores de saída para cada neurônio na segunda camada oculta, onde cada neurônio calcula uma função diferente com seus valores de entrada e assim por diante para a camada de saída final que faz uma previsão (Sanderson *et al.*, 2020).

Um modelo de RNA aprende comparando iterativamente suas previsões com os resultados observados e, em seguida, atualizando seus parâmetros de peso para melhorar suas previsões (Sanderson *et al.*, 2020).

Redes Neurais *MultiLayer Perceptron* (MLP)

Redes *MultiLayer Perceptron* (MLP) são uma classe de RNA que apresentam pelo menos uma camada intermediária ou escondida (*hidden*) fato que as torna capazes de resolver problemas não linearmente separáveis e implementar qualquer função contínua. A utilização de duas ou mais camadas intermediárias permite a aproximação de qualquer função matemática (Braga *et al.*, 2000)

Em Redes *MLP* o processamento realizado por cada nodo é definido pela combinação dos processamentos realizados pelos nodos da camada anterior que estão conectados a ele. Pode-se dizer que as unidades intermediárias de uma rede MLP funcionam como detectores de

características. Elas geram uma codificação interna dos padrões de entrada, que é então utilizada para definição da saída da rede (Fig.13) (Braga *et al.*, 2000).

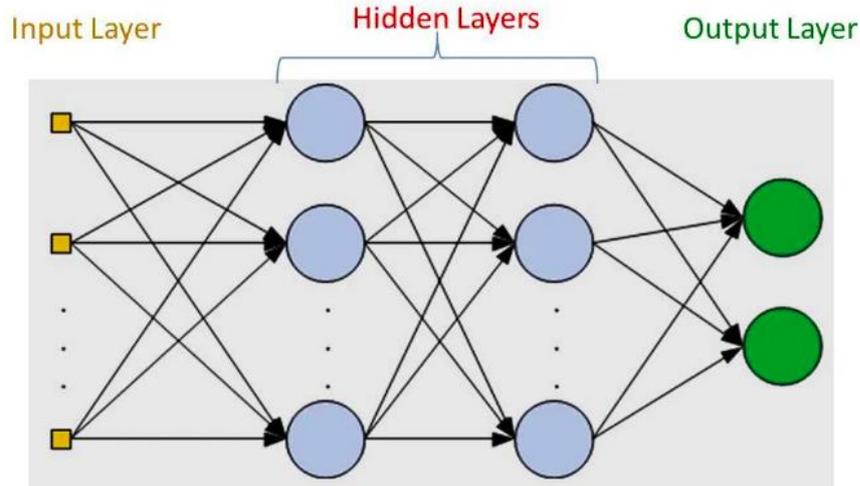


Figura 13: Arquitetura de uma rede neural *MLP*

Fonte: deeplearningbook.com.br

Na Figura 3, temos a camada de entrada (do inglês- input layer), as camadas escondidas (do inglês- hidden layers) e a camada de saída (output layer).

Um dos maiores desafios de se trabalhar com RNA é o dimensionamento de sua topologia. Tanto o número de camadas intermediárias como o número de nodos da camada são definidos em função de uma inspeção prévia da complexidade do problema e dos dados disponíveis. Uma vez definida a topologia inicial um processo de refinamento sucessivo é realizado até que a estrutura final para o modelamento do problema é obtida e que não seja muito rígida a ponto de não modelar fielmente os dados e não seja muito flexível a ponto de modelar também os ruídos (Braga *et al.*, 2000).

Dentre os algoritmos de treinamento de RNA, o algoritmo de backpropagation é um dos mais utilizados, sendo ele um algoritmo supervisionado que utiliza pares (entrada e saída desejadas) para, por meio de um mecanismo de correção de erros, ajustar os pesos da rede. O treinamento ocorre em duas fases que têm sentidos opostos. A fase *forward* (para frente) é usada para definir a saída da rede para um dado padrão de entrada. A fase *backward* (para trás) utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões (Braga *et al.*, 2000).

5.10. Random forest

Random forest (floresta aleatória) é um algoritmo de aprendizagem supervisionado que pode ser utilizado tanto para classificação como para regressão (Sanderson *et al.*, 2020). Inicialmente a floresta aleatória foi proposta para melhorar a precisão da árvore de decisão única que em muitos casos sofriam *overfitting* e conseqüentemente má generalização dos dados (Hastie *et al.*, 2008).

RF é um exemplo de comitês de algoritmos (*ensembles*) de aprendizado de máquina no qual um conjunto de árvores de decisão é utilizado para obter uma única resposta como saída, que tende a apresentar melhor desempenho que as obtidas com cada árvore do modelo em separado, devido a redução de variância (Breiman, 2001, Raschka *et al.*, 2017). A equação a seguir descreve a saída para o modelo RF:

$$\hat{f}_{fa}^B(\mathbf{x}_i) = \frac{1}{B} \sum_{b=1}^B T(\mathbf{x}_i, \Theta_b), \quad (3)$$

Onde, B é o número total de árvores, T() representa a resposta de uma árvore *b* para um vetor de entrada \mathbf{x}_i , e Θ_b representa os parâmetros desta árvore.

O algoritmo RF extrai diferentes subconjuntos de dados a partir de uma base de dados de entrada. As árvores de decisão são componentes classificadores não enviesados e independentes, construídas pela formação dos subconjuntos, utilizando uma amostragem por *bootstrap*, em que cada árvore é treinada com um conjunto de observações de treinamento obtidas por meio de uma seleção aleatória com repetição (Kursa & Rudnicki, 2010).

Na construção de preditores através do método *bagging* (ensacamento), em média dois terços das amostras do conjunto de dados são utilizados no treinamento de cada preditor. O restante das amostras recebe o nome de *Out-of-bag* (Breiman, 1996).

Uma característica importante de uma floresta aleatória é a possibilidade do uso destas amostras para estimar o desempenho do modelo, e também como critério de parada no treinamento deste último, por meio da interrupção do acréscimo de novas árvores ao RF floresta aleatória, com base no valor da estimativa *Out-of-bag* calculada. Esta estimativa é feita tomando a média do desempenho da classificação para cada amostra apresentada a floresta aleatória, com a

peculiaridade de considerar somente as árvores de decisão que não utilizaram a amostra em questão em seus treinamentos (Parmezan, 2016)

O resultado da floresta, formada pelas árvores de decisão, é obtido pela média dos resultados individuais, no caso de regressões, ou pela decisão obtida mais vezes (maioria de votos), no caso de classificações (Kursa & Rudnicki, 2010). O que diferencia esta técnica de outras é que os componentes das árvores de decisão são formados aleatoriamente.

5.11. Algoritmos de boosting

Algoritmos de *boosting* fazem parte de preditores tipo *ensemble* que utilizam uma combinação de resultados de preditores fracos. A ideia principal do *boosting* é adicionar novos modelos ao *ensemble* sequencialmente. A cada iteração particular, um novo modelo aprendiz fraco e básico é treinado em relação ao erro de todo o conjunto aprendido até agora. A abordagem por *boosting* foi proposta por Schapire (1990), a partir daí vários algoritmos dentro do contexto de *boosting* foram desenvolvidos.

Modelos XGBoost constituem *ensembles* de árvores de regressão semelhantes que formulam hipóteses sobre os exemplos, agregando as respostas de uma assembleia de preditores simples. O conjunto de árvores de regressão que compõem esta assembleia é elaborado em duas etapas: (1) uma árvore T_0 é construída adicionando à sua estrutura a ramificação do atributo F_i que mais aperfeiçoa o preditor a cada etapa e, definida sua estrutura, são calculados os valores ótimos para as folhas l ; (2) a inclusão de novas árvores $T_1...N$ ao *ensemble* é orientada pelo erro residual do modelo, de maneira que novos preditores complementem as imperfeições dos anteriores (Friedman, 2002).

A principal diferença entre *Bagging* presente em RF e *boosting*, é que no último, os conjuntos de dados re-amostrados são construídos especificamente para gerar aprendizados complementares e a importância do voto é ponderado com base no desempenho de cada modelo. Em *bagging* ocorre atribuição de mesmo peso para todos os votos. Essencialmente, esse procedimento permite aumentar o desempenho de um limiar arbitrário simplesmente adicionando *learners* mais fracos. Dada a utilidade desse achado, *Boosting* é considerado uma das descobertas mais significativas em aprendizado de máquina (Lantz, 2013).

5.12. SVM- Support vector machine

O conceito de Máquinas de Vetores de Suporte (SVM) foi introduzido nos anos 60, mas a implementação do algoritmo, como classificador não linear através da aplicação de funções Kernel, só surgiu no final do século XX (Boser *et al.*, 1992).

Máquinas de vetores de suporte (SVM do inglês Support Vector Machines) são modelos de aprendizagem supervisionados aplicados tanto em tarefas de classificação quanto de regressão e predição (Tapak *et al.*, 2019).

O SVM é usado para encontrar uma função que separe os dados e preveja a partição da variável resposta dados as variáveis explicativas. No reconhecimento de padrões, esta técnica tem sido considerada o estado-da-arte, pois têm a capacidade de operar em problemas de classificação de elevado grau de complexidade.

SVM tenta encontrar o hiperplano ótimo que permita maximizar a separação de dados de classes distintas (Boser *et al.*, 1992). A margem de separação dos dados define a fronteira de decisão (hiperplano ótimo) e os padrões que estão mais próximos dela são os vetores de suporte. Além do ótimo desempenho em problemas de classificação, o SVM também pode ser utilizado em problemas de regressão em que cada variável a ser predita é definida como *target*, enquanto as restantes são consideradas para treino. Este método apresenta uma fase de treino, onde são utilizados os dados completos com o intuito de determinar os parâmetros ótimos do modelo.

Em problemas de regressão, os SVMs já foram aplicados por exemplo: à previsão da qualidade do ar (Liu *et al.*, 2017), demanda de água e previsão da qualidade da água (Ghalekhondabi *et al.*, 2017; Zhang *et al.*, 2017), e Previsão de surtos de Influenza (Tapak *et al.*, 2019). Prieto e colaboradores (2020) utilizaram SVM em problemas de classificação de linguagem natural para detecção de casos de uso indevido de opióides em relatórios de paramédicos.

5.13. Avaliação da performance dos modelos de predição

A avaliação da performance de um algoritmo de aprendizado de máquina em um *dataset* é realizada por meio da avaliação do quão bem as predições realizadas pelo modelo ajustado reproduzem o valor observado para a resposta de interesse.

É necessário comparar a saída do modelo, que é o valor predito (\hat{y}_i), com o esperado se aproxima do valor observado (y_i).

Para análises de regressão em aprendizado de máquina o erro quadrático médio (EQM) é uma medida que pode ser utilizada.

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

Onde n é o número de amostras dos dados \hat{y}_i e y_i . A raiz quadrada resulta em valor na mesma unidade dos dados originais. O resultado descreve a distância média entre valores observados e preditos pelo modelo. Neste caso, quanto mais próximo de 0 for o valor de EQM, melhor é o desempenho do modelo e menor é a diferença entre valor previsto e observado (James *et al.*, 2014)

RMSE (*Root Means Square Error*) ou raiz do erro quadrático médio, é uma medida de desvio da diferença entre o valor real e o valor previsto, ou seja, da distribuição de resíduos e reflete como os dados se concentram em torno do modelo previsto. Essa métrica também é usada para escolher a rede de melhor ajuste (González-Bandala *et al.*, 2020). O RMSE tem a mesma unidade de medida do atributo alvo, tornando mais fácil mensurar o desempenho do modelo avaliado (Faceli *et al.*, 2011).

MAPE e RMSPE compartilham algumas características úteis, visto que o denominador é o valor real esperado, o resultado não é afetado pela unidade de medida da série. Isso torna o MAPE e o RMSPE uma boa métrica para comparar o desempenho de um método de previsão em séries diferentes ou o desempenho de muitos métodos em uma série (González-Bandala *et al.*, 2020).

MAE, parecido com MSE, em vez de elevar a diferença entre a previsão do modelo, e o valor real, ao quadrado, ele toma o valor absoluto. Neste caso, em vez de atribuir um peso de acordo com a magnitude da diferença, ele atribui o mesmo peso a todas as diferenças, de maneira linear (González-Bandala *et al.*, 2020).

MSE calcula a média dos erros do modelo ao quadrado. Ou seja, diferenças menores têm menos importância, enquanto diferenças maiores recebem mais peso (González-Bandala *et al.*, 2020).

O RMSPE mede a diferença entre os valores previstos e reais como uma porcentagem (González-Bandala *et al.*, 2020).

Quando o problema em questão for de classificação, a métrica utilizada mais comumente é a área abaixo da curva (AUC) ROC (*receiver operating characteristic*) (James *et al.*, 2014).

5.14. Interpretação dos modelos de predição

Modelos de aprendizado de máquina oferecem muitas oportunidades para melhorar a vida privada e pública das pessoas. Com o advento do *big data*, descobrir padrões e estruturas em grandes volumes de dados de maneira automatizada tornou-se um componente central da ciência de dados e atualmente impulsiona aplicações em diversas áreas como biologia, epidemiologia, medicina, direito e finanças. No entanto, um impacto tão positivo traz grandes desafios, como por exemplo, entendermos as decisões sugeridas por esses modelos. De fato, o funcionamento interno de um modelo pode ser difícil de entender. Entretanto, a explicabilidade pode facilitar a compreensão de vários aspectos de um modelo, levando a *insights* que podem ser utilizados por várias partes interessadas (Belle e Papantonis, 2020).

Compreender por que um modelo faz uma determinada previsão pode ser tão crucial quanto a precisão da previsão em muitas aplicações. Frequentemente em aprendizado de máquina, maior precisão para grandes conjuntos de dados é alcançada por modelos complexos como modelos de *ensemble* ou *deep learning*, criando uma tensão entre precisão e interpretabilidade (Ludenberg e Lee, 2017).

A crescente disponibilidade de big data aumentou os benefícios do uso de modelos complexos trazendo à tona o *trade-off* entre precisão e interpretabilidade da saída de um modelo (Ludenberg e Lee, 2017).

Vários métodos foram propostos recentemente para ajudar os usuários a interpretar as previsões de modelos complexos, dentre eles, o SHAP (*SHapley Additive exPlanations*) (Ludenberg e Lee, 2017).

O objetivo do SHAP, neste caso, é construir um modelo linear em torno da instância a ser explicada, e então interpretar a coeficientes como a importância do recurso. No SHAP a base matemática está enraizada na teoria dos jogos de coalizão, especificamente nos valores de Shapley (Belle e Papantonis, 2020).

6. MATERIAL E MÉTODOS

6.1. Artigo 1

6.1.1 Delineamento do estudo

Realizou-se uma revisão integrativa da literatura científica a respeito dos fatores eco epidemiológicos que estão envolvidos no processo de ocorrência da febre amarela na região Sudeste do Brasil, região que teve grande relevância na ocorrência de casos de febre amarela entre 2016 e 2018 e que inclui o estado de Minas Gerais.

Realizou-se um levantamento dos artigos completos publicados na base de dados Pubmed, disponível em: <<https://pubmed.ncbi.nlm.nih.gov/>>

O quadro de busca com os descritores utilizados na pesquisa juntamente com operadores booleanos pode ser observado no quadro abaixo (Fig. 14):

```
("yellow fever"[All Fields] OR "yellow fever virus"[All Fields] OR "sylvatic yellow fever"[All Fields])
AND ("brazil"[MeSH Terms] OR "brazil"[All Fields]) AND (("disease outbreaks"[MeSH Terms] OR
("disease"[All Fields] AND "outbreaks"[All Fields]) OR "disease outbreaks"[All Fields] OR
"outbreak"[All Fields]) OR ("epidemics"[MeSH Terms] OR "epidemics"[All Fields] OR "epidemic"[All
Fields])) AND ("epidemiology"[Subheading] OR "epidemiology"[All Fields] OR "epidemiology"[MeSH
Terms]) Filters: Full text, from 2016 – 2020
```

Figura 14: Descritores de busca utilizados na base de dados Pubmed

Foram selecionados artigos que possuíssem texto completo, sem restrição de idioma, publicados entre os 01 de janeiro de 2016 a 31 agosto de 2020.

Seguindo a metodologia proposta por Souza *et al.*, (2019) de inclusão e exclusão dos trabalhos científicos relacionados à epidemiologia da FA, possíveis fatores ambientais, agente etiológico, vetores e hospedeiros contemplaram a presente pesquisa.

A seleção dos textos iniciou-se pela etapa de identificação, seguida da etapa de triagem através da leitura de título, posteriormente, etapa de elegibilidade através de leitura e análise de resumo e finalmente da etapa de inclusão dos artigos de interesse pela análise de todo o texto juntamente

com avaliação de fator de impacto (FI) Scimago Journal Rank, do periódico de publicação, o qual deveria ser ≥ 1 .

6.2 Artigo 2

6.2.1. Delineamento do estudo

Foi realizado um estudo descritivo, retrospectivo a partir da análise dos dados de febre amarela humana do Sistema de Informação de Agravos de Notificações – SINAN, disponíveis na plataforma DataSUS, da Secretaria de Estado de Saúde de Minas Gerais, entre os anos de 2007 e 2020, notificados no estado de Minas Gerais, disponíveis em: http://tabnet.saude.mg.gov.br/deftohtm.exe?def/agravos/famarnet_r.def, além de análises das epizootias suspeitas e confirmadas para febre amarela em que foram utilizadas as fichas de notificação/investigação de epizootias do SINAN.

6.2.2. Variáveis do estudo

Utilizou-se as seguintes variáveis da ficha de investigação de febre amarela para análise: ano, mês, ocupação, idade, sexo, município, UF (MG), zona de residência, escolaridade, critério de confirmação e evolução do caso.

Para análise das epizootias em PNH as variáveis utilizadas foram: data de início da epizootia, UF (MG), município de ocorrência, animais acometidos (PNH), número de animais doentes e mortos envolvidos na epizootia, zona de ocorrência, coleta de material para exame de laboratório, suspeita diagnóstica, resultado laboratorial para febre amarela.

6.2.3. Análise de dados

Foram criadas tabelas e gráficos para descrever a distribuição dos dados.

O teste de qui-quadrado segundo Sampaio (1998) foi realizado para avaliação da distribuição das frequências entre as diferentes variáveis em software Excel versão 2013.

Para o cálculo da taxa de letalidade, considerou-se o número de óbitos confirmados dividido pelo número de casos confirmados, multiplicado por 100 (Casali *et al.*, 2019).

6.2.4. Aspectos éticos

Para o presente estudo não foi necessário consentimento dos comitês de ética pertinentes uma vez que se trata de banco de dados de livre acesso disponíveis na plataforma do DataSus, segundo a lei de acesso à informação (Lei nº 12.527, de 18 de novembro de 2011). Nestes bancos de dados as informações sobre os pacientes e seus respectivos tratamentos não estão disponíveis para assegurar os direitos de sigilo dos pacientes e profissionais da saúde envolvidos nas notificações.

O banco de dados de epizootias foi obtido por meio do e-sic da Secretaria de Estado de Saúde de Minas Gerais, protocolo nº 01320000194202045 e nº 01320000004202251.

6.3. Artigo 3

6.3.1. Delineamento do estudo

Trata-se de um estudo de previsão de série temporal multivariada em que os dados pertinentes à pesquisa foram organizados em séries históricas com 14 anos, oriundos das macrorregiões de saúde centro e nordeste de Minas Gerais, selecionadas para ser realizada a previsão de febre amarela humana por meio de modelos algorítmicos de aprendizado de máquina.

A seleção do recorte do tempo foi orientada pela disponibilidade de todos os dados que fazem parte deste trabalho. As epizootias em primatas não humanos passaram a ser registradas no SINAN somente a partir do ano de 2007, sendo esse o marco para o início da série a ser estudada, dada a importância da infecção dos primatas não humanos na dinâmica da doença.

Comparou-se os modelos utilizados para previsão da febre amarela nas regiões de Minas Gerais de forma a encontrar o melhor modelo que pudesse prever os casos da doença utilizando as variáveis selecionadas.

6.3.2. Local do estudo

O estado de Minas Gerais é uma das 27 unidades da República Federativa do Brasil, na América do Sul. Está localizado na região Sudeste do Brasil, juntamente com os estados do Espírito Santo, Rio de Janeiro e São Paulo. Seu território fica entre os paralelos 14°13'58" e 22°54'00" de latitude sul e os meridianos de 39°51'32" e 51°02'35" a oeste de Greenwich. Ocupa um fuso horário correspondente a -3 horas em relação a Greenwich (Fig15.) (IBGE, 2021).



Figura 15: Mapa do estado de Minas Gerais na América do Sul.

Fonte: Governo do Estado de Minas Gerais, 2021.

Minas Gerais possui 853 municípios e de acordo com o Censo Demográfico 2010. A área territorial total oficial é de 586.222,122 Km² e apresenta uma população estimada de 21.411.923 pessoas para o ano de 2021 (IBGE, 2021).

A capital é Belo Horizonte, cuja população estimada em 2.530.701 habitantes para o ano de 2021 e densidade demográfica de 7.167,00 habitantes por quilômetro quadrado (IBGE, 2021).

A cobertura vegetal de Minas Gerais pode ser resumida em quatro tipos ou biomas principais: Mata Atlântica, Cerrado, Campos de Altitude ou Rupestres e Mata Seca. Diversos fatores, entre eles, o clima, o relevo e as bacias hidrográficas, são predominantes na constituição da variada vegetação regional (Fig. 16) (Governo..., 2021).

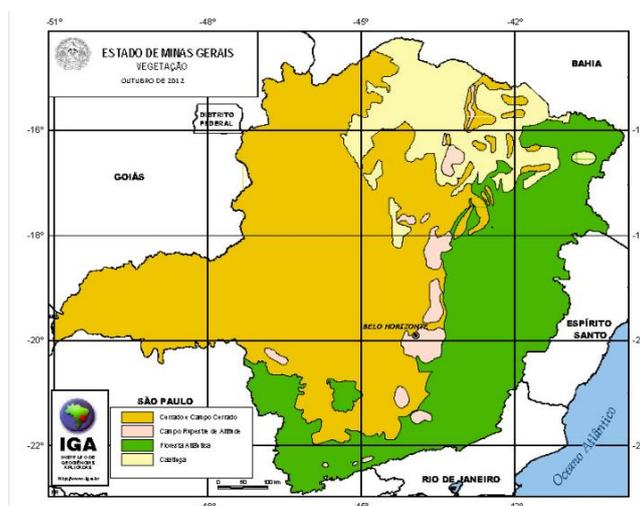


Figura 16: Principais biomas do estado de Minas Gerais

Fonte: Governo do Estado de Minas Gerais, 2021.

De acordo com o Plano Diretor de Regionalização do ano de 2014 (PDR), Minas Gerais encontra-se subdividida em 13 regiões ampliadas de saúde, compostas por 77 microrregiões e 28 unidades regionais de saúde – URS's, sendo esta, a estrutura organizacional da Secretaria de Estado de Saúde de Minas Gerais (SES-MG, 2014).

De acordo com o Plano Diretor de Regionalização (PDR) do Sistema Único de Saúde em Minas Gerais (PDR-SUS/MG), o estado foi dividido em 14 macrorregiões de saúde. O PDR é um instrumento de planejamento em saúde que estabelece uma base territorial e populacional para cálculo das necessidades, da priorização para alocação dos recursos e da descentralização programática e gerencial (Fig. 17).

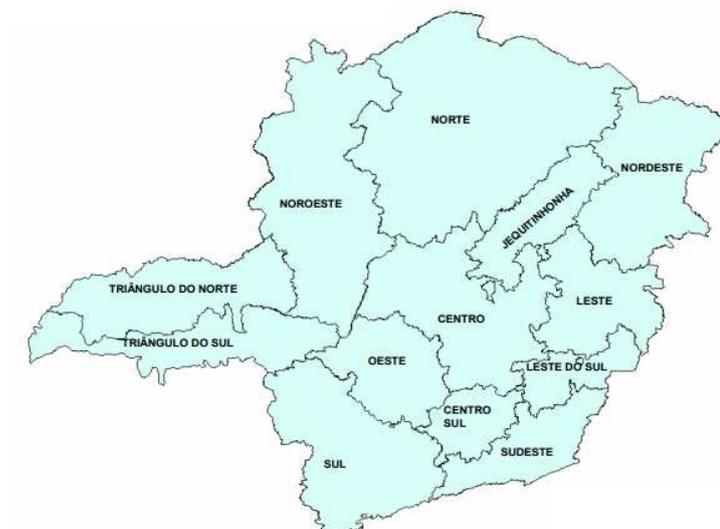


Figura 17: Divisão das macrorregiões de saúde do estado de Minas Gerais, segundo Plano Diretor de Regionalização (PDR-SUS/MG 2011 - 2014).

Fonte: SES- MG. Apresentação cartográfica PDR-MG 2011- Vigente na alteração 2014.

*Exceção do município de Divinópolis de Minas que saiu da Macrorregião de saúde Centro/microrregião de Guanhães e passou a pertencer à macrorregião de saúde Leste em 2014, que não faz parte do escopo deste trabalho.

Buscou-se identificar duas macrorregiões de saúde que obtiveram maior número de notificações para febre amarela dentro da série estudada.

A macrorregião de saúde Centro apresenta uma extensão de 58.120 Km², é composta por 103 municípios e apresenta uma população estimada de 6.480.169 habitantes e confirmaram 232 casos de febre amarela humana.

A macrorregião Nordeste apresenta uma área de 57.347,5Km², é composta por 63 municípios e tem uma população estimada de 931.946 habitantes e confirmaram 202 casos de febre amarela humana.

As duas macrorregiões a serem estudadas representam 20% do total de municípios do estado de Minas Gerais e 34,3% da população do estado de Minas Gerais (Minas Gerais, 2014).

O estudo foi conduzido com ênfase nas macrorregiões de saúde centro e nordeste, definidas pelo plano diretor da regionalização da saúde em Minas Gerais, em função da disposição dos

circuitos espaciais de produção de casos de febre amarela por local provável de infecção no estado de Minas Gerais entre 01 de janeiro de 2007 a 31 de dezembro de 2020.

6.3.3. Fonte de dados

Cobertura vacinal

Dados de cobertura vacinal anual contra febre amarela por município no estado de Minas Gerais anos 2007 a 2020, oriundos do Sistema de Informação do Programa Nacional de Imunizações (SI-PNI) foram solicitados à SES-MG via e-SIC, por meio do protocolo nº 013200002592021.

Febre amarela

Dados de notificações de casos de febre amarela em humanos e epizootias em PNH confirmados entre 01/01/2007 e 31/12/2020, oriundos de registros do SINAN, foram solicitados na SES-MG via e-SIC, Sistema Eletrônico do Serviço de Informação ao Cidadão, no site: <<http://www.acessoinformacao.mg.gov.br/>> através de cadastro pessoal prévio. Os dados foram liberados por meio dos protocolos nº 01320000194202045 e nº 01320000004202251.

Para análise dos casos de febre amarela humana entre os anos de 2007 a 2020 foram utilizadas as variáveis: data dos primeiros sintomas, classificação final (Febre amarela silvestre), local provável de infecção - UF (Minas Gerais) e município provável de infecção. Sendo excluídos casos comprovados de outros estados e casos em que a variável município provável de infecção estava vazia.

Dados meteorológicos

Dados meteorológicos foram obtidos junto ao INMET - Instituto Nacional de Meteorologia. (<https://bdmep.inmet.gov.br/>). Foram solicitados dados diários de todas as estações meteorológicas convencionais e automáticas presentes nas macrorregiões de saúde do estado de Minas Gerais trabalhadas: Centro e Nordeste (Tab. 1).

Tabela 1: Estações meteorológicas convencionais e automáticas presentes nas macrorregiões de saúde centro e nordeste segundo município de localização, tipo de estação, código da estação, latitude, longitude e altitude.

Estações pertencentes à macrorregião de saúde Centro					
Município	Tipo de estação	Código	Latitude	Longitude	Altitude
Belo Horizonte	Convencional	83587	-19,934444	-43,952222	915,47
Belo Horizonte/ Pampulha	Automática	A521	-19,883889	-43,969444	854,02
Conceição do Mato Dentro	Convencional	83589	-19,020355	-43,433948	663,02
Curvelo	Convencional	83536	-18,7475	-43,454722	668,26
Curvelo	Automática	A538	-18,747711	-44,453785	669,48
Ibirité	Convencional	83632	-20,016666	-44,084722	822,08
Ibirité	Automática	A555	-20,031457	-44,011249	1198,82
Sete Lagoas	Convencional	83586	-194854	-44,173798	763,68
Sete Lagoas	Automática	A569	-19,455278	-44173333	719
Estações pertencentes à macrorregião de saúde Nordeste					
Município	Tipo de estação	Código	Latitude	Longitude	Altitude
Araçuaí	Automática	A566	-16,848889	-42,035278	308
Teófilo Otoni	Automática	A527	-17,89284	-41,515423	467,17
Itaobim	Automática	A550	-16,575667	-41,485588	271,63
Águas Vermelhas	Automática	A549	-15,751536	-41,457787	754,07

No período de 01/01/2007 a 31/12/2020 as variáveis utilizadas foram: precipitação total, umidade média e temperatura média diários.

É comum trabalhar com dados climáticos e deparar-se com dados faltantes, seja por problemas técnicos nas estações ou mesmo humanos. Para amenizar os possíveis vieses relacionados aos dados faltantes construiu-se as séries de cada estação complementando dados faltantes das estações convencionais com dados produzidos nas estações automáticas localizadas no mesmo município. Durante o recorte no tempo estudado 01/01/2007 a 31/12/2020 as estações convencionais Pompeu e Sete Lagoas da macrorregião de saúde Centro foram desativadas permanecendo ativas somente as correspondentes automáticas sendo os dados dessas utilizados em sua plenitude a partir desse momento. Nos municípios Teófilo Otoni, Almenara e Serra dos Aimorés pertencentes à macrorregião de saúde Nordeste existem somente as estações automáticas, sendo utilizados os dados das mesmas para construção da série temporal climática. A lista das estações automáticas pode ser encontrada no link (<https://portal.inmet.gov.br/paginas/catalogoaut#>) e a lista das estações convencionais (<https://portal.inmet.gov.br/paginas/catalogoman>) ambas no site do INMET.

Após obter-se uma série mais completa, avaliou-se o percentil de dados faltantes em cada série de cada município e utilizou-se técnica de imputação multivariada para completar os dados faltantes.

Foram excluídos da análise os dados as estações meteorológicas que apresentaram mais de 20% de dados faltantes em pelo menos umas das variáveis meteorológicas, sendo elas: Florestal (82% de dados faltantes da variável umidade; 82% de dados faltantes de temperatura; 9,5% de dados faltantes de precipitação), Pompeu (27% dados faltantes da variável temperatura, 22% faltantes das variáveis umidade e precipitação), além da Belo Horizonte-Cercadinho por iniciar as operações a partir de 2013, pertencentes à macrorregião Centro. Na macrorregião Nordeste: Pedra Azul (49,2% dados faltantes na variável temperatura), Almenara (28% dados faltantes na variável precipitação) e Serra dos aimorés (27,3% de dados faltantes na variável precipitação). Não houve prejuízo para as análises a exclusão das estações metereológicas com mais de 20% de dados faltantes, já que haviam outras estações na região onde as mesmas foram instaladas e os dados então foram fornecidos para a região de interesse.

Após obter-se a série completa das estações meteorológicas por meio da imputação dos dados faltantes, foram calculadas as médias aritméticas das variáveis para todos os dias do período estudado de cada macrorregião de saúde (01/01/2007 a 31/12/2020)

6.3.4 Análise de dados

No pré-processamento dos dados são realizadas as seguintes atividades: (1) definição e separação dos conjuntos de treinamento e teste; (2) transformação de variáveis quantitativas (via padronização); (3) redução de dimensionalidade do conjunto de dados por matriz de correlação e algoritmos como Boruta (4) exclusão de variáveis/observações com dados faltantes ou utilização de técnicas de imputação (algoritmo MICE).

Um problema de série temporal pode ser tratado como aprendizado de máquina supervisionado usando passos de tempo anteriores como variáveis de entrada e passos futuros como variáveis de saída. A esse método dá-se o nome de janela deslizante ou método lag (*sliding window*), que busca trazer uma correlação entre os valores das linhas aos modelos de aprendizado de máquina permitindo que os mesmos reconheçam padrões existentes entre os valores da série temporal. O número de casos de febre amarela do dia anterior (t-1) é apresentado para o modelo (Brownlee, 2017).

Os dados de 01/01/2007 a 31/12/2020 da macrorregião de saúde centro foram utilizados para o treinamento dos modelos e como dados de teste, utilizou-se a macrorregião de saúde Nordeste afim de se avaliar a performance dos modelos frente a dados desconhecidos.

A otimização dos modelos em relação aos hiperparâmetros foi realizada por meio de *grid search*.

Os modelos utilizados para análise preditiva da febre amarela foram: RNA-MLP, SVM, Xgb, RF.

Para avaliação de desempenho dos modelos utilizou-se como métricas: MSE (*mean square error*), RMSE (*Root mean square error*), MAE (*Mean absolut error*) e R^2 coeficiente de determinação entre dados previstos e observados.

6.3.5. Aspectos éticos

O presente trabalho utilizou dados secundários de acesso livre, não sendo necessário o parecer ético para sua realização, com respaldo nas Leis n^os 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação) (Brasil, 2011) e 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais) (Brasil, 2018b), sendo todos os dados utilizados sem identificação dos envolvidos.

8. RESULTADOS E DISCUSSÃO

8.1 Artigo 1

8.1.2. Eco-epidemiologia da febre amarela na região Sudeste do Brasil entre 2016 e 2018: uma revisão integrativa da literatura científica

Eco-epidemiology of yellow fever in the Southeast of Brazil between 2016 and 2018: an integrative review of the scientific literature

Resultados e discussão

Foram encontrados na etapa de identificação 111 artigos com textos completos, sem restrição de idioma, na base de dados Pubmed e selecionados, após metodologia de exclusão e inclusão de textos, 14 artigos completos (Fig. 1).

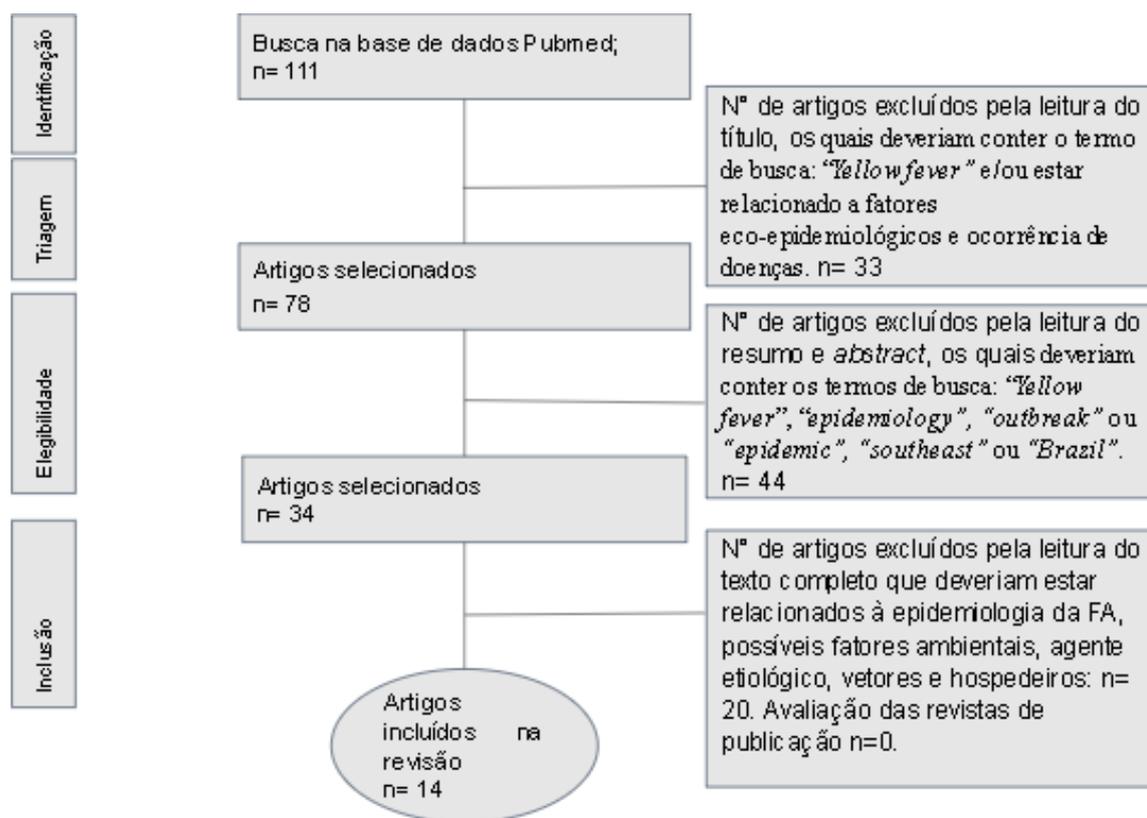


Figura 1: Fluxograma de exclusão e inclusão de artigos publicados entre 2016 e 2020 na base de dados Pubmed.

Os textos selecionados foram publicados entre janeiro de 2017 a julho de 2020, na base de dados Pubmed em 10 revistas científicas, nas quais o FI variou entre 2,280 e 41,845 (Tab. 1). Textos referenciados pelos autores e boletins epidemiológicos do Ministério da Saúde também foram utilizados na discussão.

Tabela 1: Artigos incluídos na revisão integrativa segundo fator de impacto do ano 2020

Título/ ano	Referências	Periódico, vol, n, p.	Considerações temáticas	FI 2020
Yellow fever transmission in non-human primates, Bahia, Northeastern Brazil. 2020	Goes de Jesus, <i>et al.</i>	PLoS Neglected Tropical Diseases, v.14, n.8, e:0008405.	Uso de epidemiologia molecular para identificar o genótipo do VFA que circulou em populações de PNH na Bahia.	3,885
Genomic Surveillance of Yellow Fever Virus Epizootic in São Paulo, Brazil, 2016 – 2018. 2020	Hill, <i>et al.</i>	PLoS Pathogens, v.16; n.8, e:1008699.	Análises genômicas do VFA em humanos e PNH e de dados de epizootias para avaliar a intensidade epizootica e distribuição espaço-temporal da FA.	6,210
Recent sylvatic yellow fever virus transmission in Brazil: the news from an old disease. 2020a.	Silva, <i>et al.</i>	Virol J. v.17, n.9, p.1-12.	Revisão do surto de febre amarela na região Sudeste do Brasil entre 2016 e 2019.	2,819
Yellow Fever Virus Reemergence and Spread in Southeast Brazil, 2016-2019.2020	Giovanetti, <i>et al.</i>	J Virol. v.94, n.1, p. 1-12.	Sequenciamento de genomas de VFA amostrados de PNH e pacientes infectados.	5,103
Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. 2019	Childs, <i>et al.</i>	Philos Trans R Soc Lond B Biol Sei. V.374, p.1-14.	Aprendizado de máquina utilizado para prever o risco ambiental de transbordamento da Febre Amarela.	5,680
Determination of the spatial susceptibility to Yellow Fever using a multicriteria analysis. 2019	de Paiva, <i>et al.</i>	Mem Inst Oswaldo Cruz; v.114, p.1-8.	Correlacionou os elementos físico-ambientais associados à ocorrência de FA em humanos por meio de análise multicritério	2,280
H. leucocelaenus and H. janthinomys are the primary vectors in the major YF outbreak in Brazil, 2016–2018.	Abreu, <i>et al.</i>	Emerg Microbes Infect. V.8, n.1, p. 218-231.	Levantamento entomológico de infecção natural por VFA em mosquitos capturados antes e durante o surto de 2016-2018 em cinco estados brasileiros.	5,840
Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and	Possas, <i>et al.</i>	Mem inst Oswaldo Cruz. v.113, n.10. p.1-12.	Revisão dos fatores eco-sociais envolvidos na rápida disseminação do VFA que	2,280

challenges for immunisation. 2018			aumentou o risco de reurbanização da FA.	
Spatio-temporal spillover risk of yellow fever in Brazil. 2018	Kaul, <i>et al.</i>	Parasites and vectors, v.11, n.488, p.1-12.	Previsão de risco de ocorrência de <i>spillover</i> de FA por município em todo o Brasil.	3,438
Genomic and epidemiological monitoring of yellow fever virus transmission potential. 2018	Faria, <i>et al.</i>	Science v., 361, n.6405, p.894-899.	Utilizou-se um conjunto de abordagens epidemiológicas, espaciais e genômicas para caracterizar a transmissão do VFA.	41,845
Detection and Molecular Characterization of Yellow Fever Virus, 2017, Brazil. 2018	Figueiredo, <i>et al.</i>	Ecohealth. v.15, n.4, p. 864-870.	Caracterização molecular do VFA .	2,470
Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. 2018.	Rezende, <i>et al.</i>	PLoS Negl Trop Dis. v.12, n.6, e0006538	Foram obtidas sequências de nucleotídeos de VFA detectados em humanos e NHPs no estado de Minas Gerais durante 2017-2018 com o objetivo de decifrar a origem do VFA responsável pelas epidemias recentes	2,250
The Impact of Global Environmental Changes on Infectious Disease Emergence with a Focus on Risks for Brazil. 2018	Nava, <i>et al.</i>	ILAR J, v.58, n.3, p. 393-400.	Revisão da literatura que buscou a relação entre surtos de doenças infecciosas e eventos de mudança climática e mudanças ambientais.	3,885
Genome analysis of yellow fever virus of the ongoing outbreak in Brazil reveals polymorphisms. 2017.	Bonaldo, <i>et al.</i>	Mem Inst Oswaldo Cruz. v.112, n.6. p. 447-451.	Sequenciamento genômico do VFA de <i>Alouatta clamitans</i> naturalmente infectados, obtidos no Espírito Santo.	2,280

A discussão foi dividida em temas amplos e não exclusivos.

Agente viral

O vírus da febre amarela (VFA) é membro protótipo do gênero *Flavivirus* (família *Flaviviridae*), com um genoma de RNA de fita simples e sentido positivo (Lindenbach et al. 2013; Simmonds, 2017; ICTV, 2023). Desde seu isolamento e caracterização molecular até os

dias atuais, um sorotipo e sete genótipos foram descritos na África e na América, sendo eles: África Ocidental I, África Ocidental II, África Oriental, África Central e Angola (Monath e Vasconcelos, 2015) e América do Sul I e II, sendo os últimos derivados do genótipo da África Ocidental. América do Sul I é o genótipo predominante no Brasil (Monath e Vasconcelos, 2015).

O genótipo América do Sul I possui cinco linhagens distintas. Até meados dos anos 90, as antigas linhagens (1A, 1B e 1C) co-circularam na América do Sul, mas foram substituídas pelas atuais descobertas, 1D e 1E (Souza *et al.*, 2010; Mir *et al.*, 2017), sendo a linhagem 1E, responsável pelos recentes surtos de FA no Brasil entre os anos de 2016 e 2018, a qual foi, provavelmente, originada de áreas endêmicas no Norte do Brasil (Faria *et al.*, 2018).

Por meio de análises moleculares e epidemiológicas comprovou-se a origem de ciclo silvestre do VFA, o qual circulou nos surtos do Sudeste do Brasil a partir de 2016, sendo esse, o mesmo que já havia circulado no estado de Minas Gerais entre 2000 e 2003, além de 2008 (Figueiredo *et al.*, 2018). De acordo com Cunha *et al.* (2019) a análise de cinco genomas do VFA foi consistente com a evidência de que o vírus associado ao surto de São Paulo se originou em Minas Gerais.

Análises moleculares realizadas por Rezende *et al.* (2018) combinadas com registros epidemiológicos, apoiam a ideia de que o VFA tenha sido introduzido, provavelmente, do Centro-Oeste para as regiões Sudeste e Sul do Brasil, mesma rota ocorrida nos anos 2000 e 2001 de acordo com Mir *et al.* (2017). A partir de Minas Gerais, houve uma expansão da área de circulação na região Sudeste cobrindo uma faixa de mata Atlântica, com grande potencial de atingir áreas densamente povoadas por vetores urbanos da doença (*Aedes aegypti*), sendo esta a provável rota inicial das ondas epidêmicas atuais de FA silvestre no Brasil.

Durante a duração das ondas epidêmicas de FA no Sudeste, entre os anos de 2016 e 2018, o VFA foi reintroduzido de Minas Gerais para estados do Espírito Santo e do Rio de Janeiro várias vezes (Giovanetti *et al.*, 2019).

Análises filogenéticas do VFA em PNH confirmaram a presença do mesmo em diferentes áreas urbanas de Minas Gerais em regiões próximas a populações humanas, os quais apresentavam a mesma linhagem, indicando a disseminação do VFA das áreas rurais e silvestres para as áreas

urbanas durante os surtos, com alto de risco de reintrodução da FA urbana no país (Sacchetto *et al.*, 2020).

Outro fato importante a respeito do VFA foi a observação de uma assinatura molecular distinta, em que alterações em aminoácidos localizaram-se nas duas proteínas importantes do complexo de replicase viral (NS3 e NS5) e podem estar associadas a alguma vantagem seletiva, relacionada a uma maior capacidade do vírus de infectar hospedeiros vertebrados e vetores invertebrados (Bonaldo *et al.*, 2017). Assim, essas particularidades virais podem ter desempenhado um papel na aceleração da disseminação e gravidade do surto em curso (Possas *et al.* 2018), entretanto, não se sabe se as alterações encontradas são exclusivas das linhagens pertencentes ao surto iniciado em 2016 (Bonaldo *et al.* 2017).

O VFA parece ter encontrado no Sudeste do país um ambiente com fatores eco-epidemiológicos ideais para sua multiplicação e expansão, já que a onda iniciada a partir 2016, já é considerada a de maior magnitude em 100 anos (Goes de Jesus *et al.*, 2020).

Vetores

As arboviroses são as doenças causadas pelos chamados arbovírus (ARthropod BORne VIRUS), que incluem o vírus da dengue, Zika, febre chikungunya e FA e têm sido motivo de grande preocupação em saúde pública no Brasil e outras regiões do mundo. O VFA, no Brasil, é transmitido por artrópodes silvestres (vetores), da família Culicidae, de dois gêneros: *Haemagogus* (*H. janthinomys*, *H. albomaculatus*, *H. leucocelaenus*) e *Sabethes* (*S. chloropterus*, *S. soperi*, *S. cyaneus*, *S. glaucodaemon* e *S. albiprivus*), tendo como principal fonte de infecção os PNH (Vasconcelos, 2003; De Paiva *et al.*, 2019; Abreu *et al.*, 2020).

Espécies dos gêneros *Haemagogus* e *Sabethes* tem por vezes sido encontrada compartilhado áreas de foco de FA, entretanto, tem -se sugerindo baixa capacidade adaptativa nas áreas antropicamente alteradas por *Sabethes chloropterus* (Camargo-Neves *et al.*, 2005).

As espécies *S. albiprivus* e *S. Glaucodemon* são mais comuns de serem observadas em focos epizoóticos do VFA, entretanto, suas participações nesses ciclos ainda não são bem esclarecidas (Vasconcelos *et al.*, 2001).

O gênero *Haemagogus* é considerado o vetor primário da FA e devido à sua grande distribuição espacial combinada com abundância populacional e boas taxas de infecção natural, atuaram para uma rápida disseminação e severidade do surto no Sudeste do Brasil (Abreu *et al.*, 2019).

Haemagogus janthinomys é o vetor mais importante no Brasil e em outros países. Primatófilicos, esses culicídeos picam com maior densidade por volta do meio dia animais que estejam nas copas das árvores. *Hg. Leucocelaenus* outra importante espécie que participou das ondas epidêmicas entre 2016 e 2019 no Sudeste do Brasil, apresenta atividade frequente no período da tarde e se abriga no nível do solo da floresta (Gomes *et al.*, 2010).

Haemagogus capricornii e *Hg. Janthinomys* apresentam morfologia de fêmeas e larvas aparentemente indistinguíveis, o que torna difícil determinar a participação dessas espécies no ciclo da FA, sendo necessário a diferenciação pela genitália do macho (ALENCAR *et al.*, 2002).

Não houve evidência de transmissão de VFA por *Aedes albopictus* e *Aedes aegypti*, embora o primeiro tenha sido a espécie mais difundida em diversos municípios afetados nos surtos, e mesmo havendo uma importante sobreposição entre os nichos dos vetores silvestres e os antrópicos (ABREU *et al.* 2019; Mares-Guia *et al.*, 2020).

Vale ressaltar que o VFA já foi isolado de outras espécies de culicídeos como *Psorophora ferox* no estado de São Paulo (Moreno *et al.*, 2011), na região de Ribeirão Preto, a mesma em que *Aedes scapularis* foi encontrado naturalmente infectado em 2016 (Couto-Lima *et al.*, 2017; Cunha *et al.*, 2020).

Recentemente, constatou-se infecção natural pelo VFA em *Aedes albopictus* no estado de Minas Gerais. Essa espécie é amplamente dispersa em áreas urbanas, periurbanas e rurais do país, entretanto sua capacidade vetorial precisa ser melhor avaliada (IEC, 2018).

Uma das hipóteses levantadas no meio científico sobre a dispersão do VFA na região Sudeste do Brasil foi relacionada à capacidade de o vetor ter disseminado espacialmente a doença. Entretanto, através de análise genômica do VFA em vetores e estudos sobre velocidade de mobilidade dos mosquitos, sugere-se que a ocorrência de casos em áreas distantes não foi causada pela migração natural dos PNH (Goes de Jesus *et al.*, 2020) e tampouco pela disseminação dos mosquitos mediada por seres humanos (Vasconcelos *et al.*, 2004), sendo que outras hipóteses envolvendo migração de seres humanos são mais aceitas e serão discutidas no próximo tópico.

Hospedeiros PNH e humanos

Na América Latina, de forma geral, as espécies de PNH presentes são susceptíveis ao VFA, sendo seus hospedeiros naturais. Dentre os PNH neotropicais, os gêneros *Alouatta* e *Callithrix* conhecidos popularmente como macaco Bugio e saguis, respectivamente, apresentam-se potencialmente afetados pelo VFA, sendo as espécies que mais apresentaram resultados positivos para amostras analisadas durante as ondas epidêmicas de 2016 a 2019 (Mares-Guia *et al.*, 2020). Nesses gêneros a doença apresenta elevada mortalidade, sinalizando desta forma um desequilíbrio importante para essas espécies (Moreno *et al.*, 2011).

O gênero *Ateles* sp (macaco aranha) também apresenta sensibilidade significativa ao VFA, entretanto são mais resistentes à doença, podendo curar-se adquirindo imunidade ativa. Poucos casos foram encontrados em espécies dos gêneros *Callicebus* e *Sapajus*. Os micos-leões (gênero *Leontopithecus*) estavam em sua maioria sob vigilância ativa, e apenas um animal positivo para infecção por YFV foi detectado em 2018 (Mares-Guia *et al.*, 2020).

De 2.221 óbitos de NHP, causados por FA, de 1996 até junho de 2016, no Brasil, os NHP mais afetados foram *Alouatta* sp. (85,0%) seguido por *Callithrix* sp. (8,3%) (Silva *et al.*, 2020a).

De dezembro de 2016 a junho de 2019, mais de 15.000 epizootias de PNH foram notificadas no Brasil em áreas silvestres, rurais e urbanas. Exames laboratoriais ou critérios epidemiológicos confirmaram pelo menos 1.567 eventos epizoóticos causados por FA, em 455 municípios, nas regiões Sudeste (90,9%), Sul, Nordeste, Centro-Oeste e Norte (Silva *et al.*, 2020a).

No período de julho de 2017 a junho de 2018 foram registradas no Sistema de Informação de Agravos de Notificação (SINAN), oriundas dos 4 estados da região Sudeste do Brasil, 4.908 epizootias suspeitas de FA em PNH sendo confirmadas 14,2% (699/4.908). Neste mesmo período, o estado do Rio de Janeiro registrou 619 epizootias sendo 5,6% confirmadas. Em São Paulo 2.303 foram registradas com 24,4% de casos confirmados. Minas Gerais notificou 1.832 epizootias com suspeita de FA em PNH e confirmou 5%. Espírito Santo registrou 104 epizootias sendo 1% confirmada (Brasil, 2018d).

Hill *et al.* (2020), estudando as linhagens virais circulantes no estado de São Paulo, perceberam três fases epizoóticas distintas, uma no norte de São Paulo, divisa com Minas Gerais que posteriormente se dispersou em direção ao Sul do estado para causar a segunda e terceira fases epizoóticas.

O Sul de São Paulo por ser mais densamente povoado por humanos do que o Norte, apresentou maior número de epizootias notificadas, tanto pela maior facilidade de visualização de carcaças/animais doentes por parte da população local (Hill *et al.*, 2020) como também pelas espécies predominantes na região Sul que são do gênero *Alouatta* mais sensíveis ao vírus do que as espécies predominantes no norte do estado (Culot *et al.*, 2019).

Destaca-se a transmissão silvestre de VFA em regiões florestadas altamente fragmentadas no estado de São Paulo e a importância da vigilância contínua de patógenos zoonóticos em espécies sentinela (Hill *et al.*, 2020).

Vale ressaltar a importância das epizootias de PNH como eventos sentinela para a vigilância da FA no Brasil, uma vez que a ocorrência da doença em PNH tem precedido a doença em seres humanos, sinalizando a atividade viral e, portanto, o risco do VFA infectar seres humanos (Silva *et al.*, 2020a).

Nas últimas quatro décadas, houve uma aceleração da degradação de diversos biomas brasileiros com desmatamento através de queimadas ou grandes máquinas, destruindo enormes áreas de florestas. Observa-se diante disso, um aumento das áreas de conservação combinadas com políticas de proteção da biodiversidade gerando alta densidade populacional de NHP não imune na costa Sudeste, principalmente em bioma de mata Atlântica e aumento da ocupação do ambiente modificado por NHPs adaptados, como os *Callithrix* (Saguís) e macacos pregos, que são considerados animais sinantrópicos, em áreas infestadas por *Haemagogus* o que têm favorecido a disseminação das ondas epizooticas silvestres VFA para essa região e, conseqüentemente, a infecção humana (Possas *et al.* 2018, Mares-Guia *et al.*, 2020).

A possibilidade de o tráfico de animais ser um dos fatores atuantes nos surtos de FA no Sudeste brasileiro mostrou-se plausível quando se observou que áreas florestais fragmentadas e distantes estão na rota de disseminação do vírus ao longo dos anos em diferentes estados, sendo, portanto, muito difícil a migração dos primatas entre essas áreas (Hill *et al.*, 2020). Possas *et al.* (2018) corroboram essa hipótese, uma vez que os hábitos de migração das espécies de NHP nas florestas evidenciam que eles não utilizam o solo e áreas desmatadas para migrar, ficando, portanto restritos às áreas fragmentadas de florestas.

De todas as formas de agressão que vêm sofrendo a biodiversidade brasileira, o tráfico de animais silvestres é a mais cruel. O comércio ilegal da vida selvagem é o terceiro maior negócio

ilícito do planeta, superado apenas pelo tráfico de armas e o de drogas. Acredita-se que, anualmente, esta atividade movimentada entre 10 a 20 bilhões de dólares no mundo e o Brasil participa com cerca de 10% (Lima, 2007).

De acordo com Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA), o tráfico de animais silvestres no Brasil provoca a retirada anual de cerca de 38 milhões de exemplares das florestas e matas. Os primatas são visados pelas pessoas como animais de estimação por serem inteligentes e por sua grande semelhança com os seres humanos. Por serem animais não convencionais para se ter em casa, acabam sendo, também, um objeto de ostentação (Lima, 2007).

No Brasil, a maioria dos casos humanos de FA é descrita em homens de 14 a 35 anos (Silva *et al.*, 2020b), a faixa economicamente ativa da população, trabalhadores rurais e pessoas que se expõem a atividades consideradas de risco (pesca, caça e ecoturismo) no período sazonal da doença que compreende a época do ano com elevados índices de pluviosidade e temperatura, e consequentemente, maior densidade vetorial (Saad e Barata, 2010).

Durante esse período de reemergência viral extra-Amazônica (julho/2014 a junho/2019), foram notificados cerca de 14 mil casos humanos suspeitos de FA, sendo confirmados 2.259 (6,2%) com 773 evoluindo para óbitos pela FA. A Região Sudeste concentrou a maior parte das notificações suspeitas (80,9%), assim como das confirmações (98,0%). Entre os estados, Minas Gerais registrou o maior número de casos (44,1%) e óbitos (43,1%) por FA, seguido de São Paulo, com 28,9% e 29,7%. A maior incidência foi registrada no Espírito Santo (6,49 casos por 100 mil habitantes), seguido de Minas Gerais, Rio de Janeiro e São Paulo (Brasil, 2019c).

É importante ressaltar o papel relevante do ser humano na dispersão do VFA. Estudos filogenéticos têm demonstrado que linhagens quase idênticas de VFA tem aparecido em áreas muito distantes em curto período de tempo. Acredita-se que portadores assintomáticos ou em fase prodrômica possam realizar grandes deslocamentos levando o VFA a outras regiões. O comércio global e o transporte na era moderna facilitam a movimentação de patógenos e vetores de forma rápida e a longas distâncias, alterando assim, a distribuição geográfica de doenças (Vasconcelos *et al.*, 2004).

Fatores climáticos

As doenças transmitidas por vetores, dentre elas a FA, são determinadas tanto por elementos da cadeia epidemiológica (patógeno, vetores, reservatórios e hospedeiros) quanto por variáveis ambientais, sejam estas naturais ou determinadas pela ocupação humana nos ecossistemas (padrões de uso do solo e de vegetação, umidade, temperatura, pluviosidade) (Nava *et al.*, 2017).

Os principais fatores climáticos que influenciam as biocenoses e, portanto, os seres vivos implicados no processo de transmissão de doenças, são a temperatura, umidade relativa do ar e precipitações pluviométricas (Rouquayrol, 1994).

A temperatura afeta a distribuição espaço-temporal dos vetores de doenças nas escalas global e regional, e como a temperatura continua a aumentar devido às mudanças climáticas e emissões de gases de efeito estufa, insetos vetores que habitam regiões de baixa latitude podem se expandir para novos habitats em latitudes médias ou altas regiões e em altitudes mais elevadas, levando a expansões geográficas ou mudanças nas faixas de doenças (Nava *et al.*, 2017)

As mudanças climáticas, tal como o aquecimento global, podem favorecer o surgimento de vírus e bactérias ainda desconhecidos e de pandemias, bem como, complicar os esforços de combate às doenças (Jones *et al.*, 2008).

A sazonalidade e o clima são dois fatores essenciais no controle da taxa de infecção por vírus. Estudos de previsão de *spillover* de FA em PNH têm demonstrado que os padrões sazonais da doença podem variar de acordo com as regiões do Brasil, principalmente quando relaciona-se a alta ou baixa riqueza de PNH (Kaul *et al.*, 2018). Childs *et al.* (2019) relatam que essa sazonalidade nas diferentes regiões pode ser explicada pela variação na sobrevivência, infecciosidade e abundância do vetor impulsionada pela temperatura e pela pluviosidade.

O ciclo de vida do mosquito e a replicação viral são bastante dependentes de variáveis climáticas, principalmente temperatura e índices pluviométricos (Barcellos *et al.*, 2009.)

Estudos mostram que aumentos da temperatura reduzem o período de incubação extrínseco do VFA, isto é, o período de tempo entre o mosquito estar infectado e se tornar infeccioso. A temperatura ideal para o período de incubação do mosquito tem sido estimada entre 20°C e 30°C. Os mosquitos vetores são animais poiquilotérmicos, ou seja, sua temperatura interna varia de acordo com a temperatura do ambiente. O metabolismo desses insetos aumenta com a elevação térmica, com várias implicações para a transmissão do VFA: encurtamento do estágio

de pupa; maior competência do sistema imune; e aumento da frequência de repastos sanguíneos (Hamlet *et al.*, 2008).

Uma estação chuvosa severa e prolongada associa-se com abundância de vetores e pode estar ligada ao aumento da circulação viral. Culicídeos do gênero *Haemagogus* são encontrados em áreas onde a pluviosidade anual ultrapassa 2.000 mm (Vasconcelos, 2010b; Monath e Vasconcelos, 2015).

Os gradientes térmicos secundários à altitude afetam a viabilidade do vetor e do vírus, bem como a distribuição dos PNH. Baixas temperaturas em médias ou altas latitudes, em geral, interrompem o ciclo de transmissão da doença (Hamlet *et al.*, 2008)

Um trabalho que avaliou o padrão espacial de distribuição de casos de FA e a relação entre fatores geográficos e ambientais em 13 países das Américas evidenciou que a média anual de temperatura nos locais afetados foi de 24,1°C (Hamrick *et al.*, 2017).

Os mosquitos são extremamente sensíveis às mudanças de temperatura e umidade relativa, e sua abundância está frequentemente relacionada com os níveis de pluviosidade (Kaul *et al.*, 2018).

Kaul *et al.* (2018) comentam sobre os fenômenos naturais *La Niña* e *El Niño* e suas possíveis relações com as epidemias de FA. *La Niña* caracteriza-se por um prolongamento da estação chuvosa durante a qual os vetores de VFA podem ser mais abundantes, levando a uma maior incidência de transbordamento nas populações de PNH e seres humanos, sugerindo um prolongamento da época endêmica com inclusão dos meses de julho e agosto sendo provável que mudanças na precipitação e temperatura resultantes dos eventos *La Niña* e *El Niño* impulsionem o *spillover* da FA no Brasil (Kaul *et al.*, 2018).

O entendimento do meio ambiente como um recurso a ser explorado pelo homem de acordo com seus interesses, revelou uma relação antagônica com a natureza gerando consequências danosas à todas as espécies que habitam nosso planeta. As alterações climáticas influenciam a ocorrência de doenças, principalmente, aquelas que apresentam vetores artrópodes como transmissores de agentes etiológicos, como no caso da FA.

Conclusões

A persistência do VFA na região Sudeste do Brasil, causando ondas epidêmicas de 2016 a 2018, sugere que esta região apresenta condições ecológicas e climáticas adequadas para a sua manutenção durante as estações epidêmica e inter epidêmica. Este fato apresenta riscos para o estabelecimento de ciclos enzoóticos e epidemias de FA fora da Bacia Amazônica no Brasil, em áreas densamente povoadas e densamente habitadas por *Aedes aegypti*, vetor urbano da doença, principalmente em regiões onde há presença de mata Atlântica.

O comércio ilegal de animais silvestres, desmatamento, as alterações climáticas consequentes das modificações antrópicas no meio ambiente são exemplos de ações que podem ter favorecido a ocorrência de FA em tamanha magnitude.

A conscientização de que é possível um progresso econômico em equilíbrio com a natureza e todos os seres vivos é urgente não só no Brasil, como no planeta terra como um todo. Ações políticas direcionadas a legislações ambientais severas para proteção das florestas e animais precisam ser avaliadas e os órgãos competentes necessitam ter poder de ação.

8.2 Artigo 2

8.2.1. Perfil epidemiológico dos atendimentos de febre amarela e epizootias notificados e registrados no SINAN, em Minas Gerais, entre 2007 e 2020

Epidemiological profile of yellow fever and epizootics reported and registred in SINAN, in Minas Gerais between 2007 and 2020

Resultados e discussão

As notificações de FA segundo critério de classificação dos pacientes se distribuíram de forma heterogênea ao longo da série estudada com concentração de ocorrências nos anos 2017 e 2018 (Fig. 1).

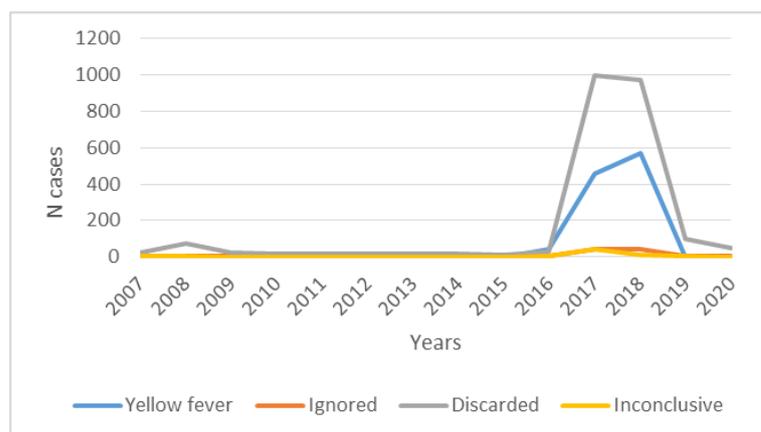


Figura 1: Distribuição das notificações de febre amarela em Minas Gerais entre os anos de 2007 e 2020 segundo a classificação final dos casos

Dos casos suspeitos de FA notificados em Minas Gerais entre os anos de 2007 e 2020, 29% foram confirmados para a doença, 66% foram descartados e 1,6% inconclusivos. Em 3,2% das notificações a variável foi ignorada.

Os anos de 2017 e 2018 representaram 95% dos casos confirmados de FA no estado de Minas Gerais, ocorrendo nos períodos sazonais dos respectivos anos as maiores epidemias da atualidade, ultrapassando todos os casos ocorridos de FA no Brasil entre 2000 e 2012.

Moussallem *et al.* (2019) no estado do Espírito Santo observou que 37% dos casos suspeitos de FA foram confirmados.

Entre os anos de 2010 e 2015 não foram registrados casos humanos de FA no estado de Minas Gerais (Minas Gerais, 2019).

De acordo com Brasil (2018), entre os anos de 2014/2015 a transmissão do VFA se deu na região Norte do país com posterior expansão para a região Centro-Oeste em 2015/2016. No ano de 2015 Minas Gerais já registrava epizootias em PNH em número discreto, entretanto isso já indicava a circulação do VFA no estado, o que culminaria com as duas maiores epidemias de FA da atualidade no Brasil.

A primeira com início no final do ano de 2016 nos Vales do Rio Doce e Mucuri e em parte da Zona da Mata e Jequitinhonha. A segunda oriunda da região Metropolitana de Belo Horizonte, Zona da Mata e partes das regiões de Campos das Vertentes, Oeste e Sul/Sudoeste mineiros (Minas Gerais, 2019).

Nos meses de janeiro e fevereiro concentraram-se a maior parte dos casos confirmados de FA silvestre no estado, representando 90% (974/1.083) da série analisada, corroborando com Escosteguy *et al.* (2019) e Cavalcante e Tauil (2016).

De acordo com Ministério da Saúde (2020), no verão o VFA encontra condições favoráveis para a transmissão. Nessa época do ano temos elevadas temperaturas e pluviosidade; alta densidade de vetores além da presença de hospedeiros primários (PNH).

De acordo com Romano *et al.* (2011), o conhecimento sobre a ocorrência sazonal da doença permitiu a intensificação de medidas de prevenção e controle no período que antecede àquele de maior ocorrência dos casos.

Em relação à distribuição espacial dos casos, observou-se que 126 municípios de Minas Gerais registraram casos confirmados de FA entre 2007 e 2020. Desses, 6 foram responsáveis por 38,5% (418/1.083), sendo eles: Belo Horizonte 12,2% (132/1.083), Caratinga 5% (53/1.083), Ladainha 4% (45/1.083), Juiz de Fora 8,1% (88/1.083), Teófilo Otoni 4,7% (51/1.083) e Mariana 4,5% (49/1.083).

Em relação ao grande aumento de casos ocorridos a partir de 2016 no estado, alguns fatores podem estar relacionados, como a baixa cobertura vacinal da população em algumas regiões associada a dificuldade de atendimento a populações mais vulneráveis antes do início dos surtos e o desequilíbrio ambiental causado por ações antrópicas (Valadares *et al.*, 2017).

Casali *et al.* (2019) mencionam o desmatamento no estado como um dos possíveis fatores que potencializaram a ocorrência da FA em Minas Gerais. Cerca de 18.433 hectares de mata Atlântica foram desmatados só no ano de 2015, com possível deslocamento de vetores para regiões com habitat mais favorável e como consequência atingindo populações de hospedeiros (PNH e seres humanos) não imunes para a FA.

Com o crescente reconhecimento de que as zoonoses sobrevivem das interações entre fatores humanos, animais e meio ambiente, entende-se que uma alta biodiversidade nos diversos ecossistemas tem a capacidade de inibir a expansão de agentes causadores de doenças. Entretanto, a perda de biodiversidade causada pelas atividades humanas, podem exacerbar o risco de ocorrência de doenças tanto para os humanos, quanto para a vida selvagem (Cardinale *et al.*, 2012). A redução da transmissão de doenças pode ser considerada, portanto, um importante serviço ecossistêmico fornecido pela alta biodiversidade (Keesing *et al.*, 2010).

Atualmente, mais do que nunca, há uma grande necessidade de entender os mecanismos que moldam os surtos e as epidemias, enquanto observa-se as modificações geradas pelas atividades humanas nos diferentes biomas e com grande perda de biodiversidade. Ciclos contínuos de disseminação de agentes infecciosos têm destacado a vulnerabilidade global das populações humanas, com ênfase nas doenças emergentes e reemergentes, as quais têm gerado impactos massivos na saúde e na economia (Carroll *et al.*, 2018).

Em relação à zona de residência, 54% dos pacientes confirmados no diagnóstico de FA silvestre residiam em zona rural e 67,2% dos casos descartados para FA residiam em área urbana. Escosteguy *et al.* (2019) observou 61,5% dos casos oriundos de zona rural. Fantini *et al.* (2021) observou que na região Sul do Brasil 80% dos casos confirmados de FA nessa região residiam em zona urbana.

Observou-se que 76% (819/1.083) dos casos confirmados de FA são casos autóctones do município de residência, corroborando com Fantini *et al.* (2021) que observou 74% de casos de FA autóctones do município de residência.

Os pilares da vigilância da FA se baseiam na vigilância de PNH que são considerados eventos sentinela (vigilância passiva) e na vigilância entomológica que se dá de forma ativa quando as ações que se baseiam no monitoramento de áreas estratégicas (sentinelas e vulneráveis/receptivas (Brasil, 2017c). Conhecer o status epidemiológico dos municípios é uma valiosa ferramenta a ser utilizada nas tomadas de decisão em saúde, principalmente as relacionadas à cobertura vacinal das populações que vivem em áreas de risco para FA.

Considerando que em Minas Gerais, no período de 2007 a 2018, foram aplicadas mais de 18 milhões de doses de vacina de febre amarela, e que se registrou, em 2018, apenas 16 casos em pacientes vacinados, a vacina tem mostrado altas taxas de segurança e efetividade na prevenção da população quanto à transmissão do vírus, sendo ainda a principal medida a ser adotada. Como medida adicional, para a população mais exposta à circulação do vírus, recomenda-se também a utilização de repelente como medida de proteção individual (Brasil, 2018c).

Em relação à idade dos pacientes observou-se que em 72,6% (787/1.083) dos casos confirmados de FA tinham idade entre 35 e 64 anos, corroborando parcialmente com Fantini *et al.* (2021) na região Sul do país que observou 80,7% de casos na faixa etária 29 a 59 anos. Crianças de 0 a 14 anos representaram 1,8% (20/1083) dos casos confirmados de FA no estado de Minas Gerais.

Cavalcante e Tauil (2016) observaram 86,7% de casos de FA em jovens adultos no Brasil entre 2000 e 20012.

Ocupações laborais relacionadas aos campos/florestas representaram 55% (342/620) dos casos confirmados nos quais essa variável foi preenchida, sendo a categoria trabalhador agropecuário em geral a mais acometida, representando 77,5% (265/342) dentre as profissões relacionadas a campo/florestas. Cavalcante e Tauil (2016) observaram 45% dos casos de FA relacionados a atividades rurais. Escosteguy *et al.* (2019) observou 40,4% de atividades laborais relacionadas ao meio rural. Moussallem *et al.* (2019) encontrou 50,8% dos casos confirmados em indivíduos que trabalham no setor rural.

Ressalta-se a importância da variável ocupação para a investigação de casos de febre amarela silvestre devido a alguns tipos de atividades serem realizadas em ambiente natural onde ocorre o ciclo silvestre da febre amarela.

Observou-se que essa variável foi preenchida em apenas 45% (620/1.083) das notificações confirmadas para a doença. De acordo com Siqueira *et al.* (2017) variáveis com elevada percentagem de ausência de respostas em notificações do Ministério da Saúde, podem ocorrer por falta de motivação, equipe mal dimensionada ou tempo insuficiente para o preenchimento completo da ficha de notificação da FA, diante da priorização de outras demandas das unidades notificadoras.

Em relação à escolaridade, observou-se que houve dentre os pacientes com febre amarela, predomínio de pessoas com pelo menos um ano cursado entre o 1º a 4º ano do ensino fundamental, representando 22% (239/1.083), 5% (57/1.083) eram analfabetos e apenas 1% (15/1.083) tinham ensino superior completo. Fantini *et al.* (2021) observou em 22,6% de casos confirmados para FA com pelo menos um ano estudado entre 1º a 4º ano do ensino fundamental. Escosteguy *et al.* (2019) observou que 73% dos casos de FA no estado do Rio de Janeiro tinham no máximo o ensino fundamental incompleto.

Em 49% das notificações confirmadas para FA a variável escolaridade dos pacientes não foi preenchida, corroborando com Fantini *et al.* (2021) que observou 45,2% da variável escolaridade vazias na região Sul do país. Gonçalves (2021) no Distrito Federal observou que 90% dos casos de FA não tinham informação sobre a escolaridade dos pacientes.

De acordo com Schettino *et al.* (2020) o baixo nível de escolaridade dos trabalhadores florestais, ocupação observada em maior frequência no presente estudo, dificulta a percepção dos problemas ocupacionais e riscos ambientais aos quais os mesmos ficam expostos durante a execução de suas atividades, dificultando assim, a aplicação de medidas de proteção que possam eliminar os riscos de acidentes e desenvolvimento de doenças com riscos ocupacionais.

Em relação ao sexo dos pacientes, o sexo masculino representou 85% (923/1.083) das notificações confirmadas para FA corroborando com Escosteguy *et al.* (2019) que observou 86,5%, Gonçalves (2021) 80%, e Moussallem *et al.* (2019) 79,5%. Fantini *et al.* (2021) observou 71% dos casos de FA em homens.

Sugere-se que homens adultos se expõem mais em ambientes onde o vetor da febre amarela está presente do que mulheres. De acordo com Albano *et al.* (2010), homens comparecem menos aos serviços de saúde de forma preventiva, quando comparados com mulheres, crianças e idosos. A demanda dos homens nos serviços de saúde limita-se a ações de cunho curativo quando já existe alguma doença já instalada, como diabetes e hipertensão.

De acordo com Brasil (2008), os elementos culturais de estereótipo e gênero masculino estão relacionados à não adesão dos homens aos serviços de saúde de atenção primária. O gênero masculino foi constituído como referência de força, trabalho e desenvolvimento da sociedade. A doença, sob essa hipótese é vista como sinal de fraqueza e fragilidade e não é devidamente reconhecida pelos homens como inerente à sua própria condição biológica reduzindo assim a percepção de risco à qual esses trabalhadores estão expostos.

Segundo Ferraz *et al.* (2012) é imprescindível que as equipes de saúde da zona rural reconheçam os riscos ocupacionais existentes nessa população e atuem com medidas de proteção à saúde e prevenção de acidentes de trabalho.

Observou-se que em 70,85% (2.637/3.721) dos casos suspeitos de FA foi realizado o exame de laboratório para diagnóstico (Tab. 1).

Tabela 1: Distribuição dos atendimentos de febre amarela segundo classificação final do caso e critério de classificação, em Minas Gerais entre 2007 e 2020.

Classificação final	Modo de confirmação			
	Laboratorial n (%)	Clinico-epidemiológico n (%)	Ignorado n (%)	Total n(%)
Ignorado	0%	0%	121 (3.2%)	121 (3.2%)
Febre amarela confirmada	951 (25.5%)	93 (2.5%)	39 (1%)	1.083 (29%)
Descartado	1.685 (45.3%)	594 (16%)	175 (4.7%)	2.454 (66%)
Inconclusivo	1 (0%)	2 (0%)	60 (1%)	63 (1.6%)
Total	2.637 (70.85%)	689 (18.5%)	395 (10.6%)	3.721 (100%)

Foram excluídos dessa análise 9 casos de febre amarela, os quais foram classificados erroneamente como febre amarela urbana, uma vez que, não foi comprovada a transmissão por vetores urbanos da doença.

Em 25,5% (951/3.721) dos casos suspeitos de FA em Minas Gerais houve a confirmação laboratorial para febre amarela. Observou-se que em 10,6% (395/3.721) dos casos suspeitos a variável modo de confirmação encontrou-se em branco. Casos inconclusivos totalizaram 1,6% (63/3.721). Esse baixo percentual encontrado reflete uma boa qualidade do banco de dados e boa qualidade do serviço prestado à população em relação a essa variável.

As variáveis critério de classificação e classificação final do paciente apresentaram taxa de completude ótimas segundo o documento *Update Guidelines for Evaluating Public Health Surveillance Systems* (Klaucke *et al.*, 1988) publicado pelo *Centers for Diseases, Control and Prevention* (CDC) do Estados Unidos uma vez que apresentaram mais de 90% dos registros preenchidos.

Em 18,5% (689/3.721) dos casos suspeitos de FA o critério utilizado para a classificação final do paciente foi critério clínico-epidemiológico. Dos pacientes confirmados para FA observou-se que 87,8% (951/1.083) foram diagnosticados pelo critério laboratorial, 8% (93/1.083) por critério clínico epidemiológico e 3,6% (39/1.083) essa informação foi ignorada. Moussallem *et al.* (2019) observaram 90% de diagnóstico por critério laboratorial e 10% por critério clínico-

epidemiológico. Fantini *et al.* (2021) observou 87% dos casos de FA na região Sul do Brasil classificados por critério laboratorial, 7% por critério clínico-epidemiológico.

Em relação ao desfecho dos casos confirmados para FA observou-se que 64% (693/1.083) dos casos evoluíram para cura, 32% (344/1.083) evoluíram para óbito, em 3% (37/1.083) essa variável foi ignorada e em 1% (9/1.083) os pacientes evoluíram para óbito por outras causas.

A taxa de letalidade observada para o período estudado foi 32%, corroborando com Moussallem *et al.* (2019) com 30,9% no estado do Espírito Santo. Cavalcante e Tauil (2016) observaram uma taxa de letalidade de 47,8% no Brasil entre 2000 e 2012. Escosteguy *et al.* (2019) observou uma taxa de letalidade de 40,4% no estado do Rio de Janeiro. Gonçalves (2021) no Distrito Federal entre 2008 e 2019, observou uma taxa de letalidade de 70%.

Takata (2017) comenta que a alta letalidade observada em surtos de FA pode ser um artefato da subnotificação, uma vez que muitos casos são assintomáticos ou produzem apenas sintomas brandos, podendo ser confundido por várias doenças e não havendo procura médica.

Deve ser considerado caso suspeito o indivíduo com exposição em área recentemente afetada (em surto) ou em ambiente rural e/ou silvestre, com até sete dias de febre aguda (febre aferida ou relatada) acompanhada de dois ou mais dos seguintes sinais e sintomas: cefaléia (localização principalmente supraorbitária), mialgia, lombalgia, mal-estar, calafrios, náuseas, icterícia e/ou manifestações hemorrágicas ser residente ou proveniente de área de risco para febre amarela, nos últimos 15 dias, que não possua comprovante de vacinação contra febre amarela ou que tenha recebido a primeira dose há menos de 30 dias. Exames laboratoriais inespecíficos podem apresentar alterações e auxiliar na identificação de formas mais graves e no manejo clínico (Brasil, 2017a).

O diagnóstico específico da febre amarela pode ser feito diretamente pela detecção do vírus em amostras clínicas (sangue e/ou tecidos) ou indiretamente pela detecção de anticorpos por meio de sorologia, RT-PCR, isolamento viral e histologia. Os exames são realizados em laboratórios de referência em vários estados brasileiros, podendo as secretarias de saúde de cada estado e município fornecer informações sobre como encaminhar o material biológico e como receber o resultado (Brasil, 2017a).

De acordo com Brasil (2004) caso confirmado por critério clínico-epidemiológico é o caso suspeito procedente de área com ocorrência de surto ou epidemia de febre amarela, cuja

evolução para óbito tenha ocorrido até 10 dias a partir da data do início dos sintomas que impossibilita a confirmação laboratorial.

Em relação às epizootias em PNH, no período de 2007 a 2020 foram registrados um total de 5.345 notificações de suspeita de epizootias, com maior incidência entre os anos de 2016 a 2018 representando 82,9%.

Observou-se que 3,85% (206/5.345) foram positivas para FA e registradas no banco de dados do SINAN, 20,80% (1.098/5.345) tiveram resultado laboratorial negativo, 61,38% (3.281/5.345) apresentavam essa variável vazia, ou seja, sem preenchimento. Em 3,35% (179/5.345) as epizootias foram consideradas inconclusivas e em 10,79% (566/5.345) foram consideradas ignoradas. Minas Gerais (2022) observou no período sazonal de 2021/2022 que 53,4% de epizootias foram indeterminadas e no período sazonal de 2022/2023, 87,2%.

É importante ressaltar que para acompanhamento das ondas epidêmicas de 2016 a 2020 foram utilizadas planilhas de monitoramento de epizootias para atualização o banco de dados do SINAN. Quando o mesmo não ocorre em tempo hábil, as análises realizadas no banco de dados do SINAN podem não demonstrar a real situação em relação ao agravo. De acordo com Minas Gerais (2021), no período de monitoramento 2016/2017 foram registrados 1.336 epizootias em PNH com a confirmação do vírus em 327 PNHs, seguida pelo o período de monitoramento 2017/2018, com 1.574 epizootias notificadas com a confirmação laboratorial em 121 primatas. Qualquer divergência observada em relação aos dados dos boletins da SES-MG e dados do SINAN sinaliza uma falta de sinocrismo entre as esferas estadual e federal em relação à febre amarela. Sugere-se que o banco do SINAN seja alimentado e atualizado em tempo real pelas esferas estaduais e municipais, gerando informações consistentes e robustas a respeito dos agravos de notificação no país para que as tomadas de decisão sejam assertivas e fieis à real situação.

Casos indeterminados (inconclusivos, ignorados e vazios) que representaram maior parte das notificações registradas, implicam investigações sem coleta de amostra para envio ao laboratório e/ou impossibilidade de se obter um diagnóstico conclusivo. Diante do alto número de notificações de epizootias registradas, ressalta-se a importância da participação da população como ponto relevante na vigilância de PNH (Brasil, 2021b).

Sugere-se que o alto percentual de resultados indeterminados observado nas fichas de suspeita de epizootias em PNH, pode estar relacionado ao mau estado de conservação das carcaças quando o animal já está morto, impossibilitando a coleta de amostras para a realização de exames laboratoriais. Carcaças de animais silvestres são mais facilmente avistadas pelas populações humanas quando o animal vem a óbito próximo de vias de trânsito e/ou quando exala odor característico e torna-se mais facilmente perceptível sua presença, a ponto de ser visualizada por alguém que passe próximo ao local onde a mesma se encontra.

Em relação à coleta de material, observou-se que em 47% (2.508/5.345) das epizootias registradas no SINAN houve coleta de amostras para envio ao laboratório. Ao se avaliar a possibilidade de coleta de amostras segundo número de animais envolvidos nas epizootias, observou-se que dos 6.777 animais doentes ou mortos, em 47,78% (2.899/6.777) foi registrada a coletada amostra para envio ao laboratório nas respectivas notificações, caso todos animais tivessem sido coletados, o que nem sempre é possível devido às condições das carcaças encontradas, mas espera-se que se colete o maior número de animais possível numa epizootia.

Em relação ao aproveitamento de carcaças de animais mortos (6.270 animais), observou-se 42,20% (2.646/6.270) de envio de amostra para laboratório, caso todos os animais envolvidos nas epizootias fossem coletados. Se pelo menos um animal fosse coletado por epizootia, teríamos 40% (2.508/6.270) dos animais coletados. Como trata-se de análise de evento epizootico, o número de animais coletados não é disponível na ficha de epizootias. De acordo com Oliveira *et al.* (2010) em somente 21,4 % das carcaças foram coletadas amostras para testes no Rio Grande do Sul. Delfino (2011), em Goiás, aproveitou somente 34,5%. Araújo *et al.* (2011) observou que no Brasil entre os anos de 2007 e 2009, 22,2% de amostras de PNH foram enviadas para análises laboratoriais. No estado de São Paulo, Mascheretti *et al.* (2013) utilizou somente 7% das carcaças de PNH encontradas. Moutinho *et al.* (2020) relata 81,1% de aproveitamento das carcaças de PNH para análises de raiva e FA.

Observou-se que em 194 notificações de epizootias em que foi registrado o envio de material para envio ao laboratório, o número de animais doentes ou mortos apresentaram-se sem preenchimento, sendo que em 8,7% o resultado laboratorial foi positivo para FA. Estima-se, portanto, que o número de animais envolvidos nas epizootias registradas no SINAN em Minas Gerais entre 2007 e 2020 ultrapasse 6.971 animais caso pelo menos um animal tivesse amostras coletadas nessas epizootias.

Segundo Romano (2011) é comum haver um grande número de carcaças não aproveitáveis devido à dificuldade de se coletar amostras seja pelas mesmas se encontrarem em locais de difícil acesso ou pelo intervalo de tempo entre a notificação e a investigação, impedindo um correto diagnóstico devido às condições que se encontram as carcaças.

A vigilância em PNH visa identificar precocemente a circulação do VFA nesses animais, de forma a possibilitar a instituição de ações preventivas e de controle da doença, reduzindo assim a morbidade e a mortalidade de humanos pela FA. Todo PNH encontrado doente ou morto deve ser considerado caso suspeito para FA, sendo necessária coleta de material biológico para diagnóstico quando for possível e notificação compulsória imediata no SINAN para início de investigação de epizootia (Brasil, 2017c).

A vigilância da FA através de corredores ecológicos tem permitido acompanhar o deslocamento do VFA no Brasil. Obara *et al.* (2012) alertam em seu estudo que a formação de corredores ecológicos favorece a mobilidade de vetores da febre amarela silvestre.

A participação da população na vigilância de PNH é fundamental para controle da febre amarela silvestre no Brasil. A implementação do software SISS-Geo para a notificação de epizootias em PNH tem possibilitado a notificação simultânea e em tempo real de eventos e rapidez na verificação da epizootia. Durante o monitoramento 2019/2020, o SISS-Geo foi utilizado pelos estados da região Sul (PR, SC e RS), com redução de 82,9% na proporção de eventos sem identificação do gênero/espécie dos animais envolvidos, e de 76,5% na proporção de eventos sem registro das coordenadas geográficas (Brasil, 2021b).

Em relação ao número de animais envolvidos nas epizootias vivos ou mortos segundo a área de ocorrência da epizootia, observou-se que 49,9% (3.381/6.777) dos animais encontrados, estavam mortos e localizados em área rural (Tab. 2).

Tabela 2: Distribuição do número de PNH doentes e mortos envolvidos nas epizootias registradas no SINAN, segundo a zona de ocorrência das epizootias em Minas Gerais entre 2007 e 2020.

Estado dos animais envolvidos nas epizootias			
Zona de ocorrência	Animais doentes n(%)	Animais mortos n(%)	Total n(%)
Urbana	252 (3,7%)	2540 (37,5%)	2792 (41,2%) ^a
Rural	238 (3,5%)	3381 (49,9%)	3619 (53,4%) ^b
Peri urbana	14 (0,2%)	279 (4,1%)	293 (4,3%) ^b
Ignorado	0 (0%)	8 (0,1%)	8 (0,1%) ^b
Vazio	3 (0%)	62 (0,9%)	65 (0,9%) ^b
Total n(%)	507 (7,5%)	6270 (92,5%)	6777 (100%)

*Letras iguais entre linhas apresentam igualdade estatística no Teste de Qui-Quadrado corrigido para valores de $p \leq 0,05$

Em relação ao número de animais envolvidos nas epizootias, observou-se uma média de 1,26 (5.345/6.777) animais por epizootia, sendo que 92,5% estavam mortos. Araújo *et al.* (2011) encontrou 1,8 animais/epizootias e Moutinho *et al.* (2020) observou 1,03 animais/epizootia.

Na América Latina, de forma geral, as espécies de PNH presentes são susceptíveis ao VFA, sendo seus hospedeiros naturais. Dentre os PNH neotropicais os gêneros *Alouatta* e *Callithrix* conhecido popularmente como macaco Bugio e Saguis, respectivamente, apresentam-se potencialmente afetados pelo VFA, nos quais a doença apresenta elevada mortalidade, sinalizando um desequilíbrio importante para essas espécies (Moreno *et al.*, 2011).

Observou-se 41,2% das epizootias com ocorrência em zona urbana. A presença da circulação do VFA em áreas urbanas do estado de Minas Gerais fez ressurgir no meio científico o debate sobre os riscos de reurbanização da FA com transmissão por *A. aegypti*. De acordo com Sacchetto *et al.* (2020) análises filogenéticas do VFA em PNH confirmaram a presença do vírus em diferentes áreas urbanas de Minas Gerais, próximas a populações humanas, os quais apresentavam a mesma linhagem indicando a disseminação do VFA das áreas rurais/silvestres para as áreas urbanas durante os surtos, com alto de risco de reintrodução da FA urbana no país.

De acordo com Brasil (2021b) observou-se novamente a ocorrência de epizootias em PNH por febre amarela na região Centro-Oeste do país no período sazonal 2020/2021, mesmo padrão de ocorrência da doença no período sazonal 2015 que antecedeu as grandes epidemias no estado de Minas Gerais.

Conclusão

O estudo descritivo do perfil epidemiológico de casos humanos confirmados para FA e epizootias no estado de Minas Gerais é imprescindível para que políticas públicas sejam direcionadas às populações mais vulneráveis e sob maior risco de adoecerem. A alta complexidade da FA exige comprometimento de todos os atores envolvidos para seu controle eficaz. Ações sócio educativas junto à população sobre a importância da vacinação da FA e vigilância de PNH devem ser realizadas periodicamente. Contar com um sistema de informações de qualidade, confiável, com os campos das fichas de notificação preenchidos de forma correta e com atualização constante, é extremamente importante para a adoção de medidas preventivas de forma adequada a fim de minimizar os efeitos de futuros surtos da doença. Políticas públicas ambientais mais rígidas devem ser priorizadas no Brasil devido à complexidade do contexto que envolve a saúde humana, animal e ambiental.

8.3. Artigo 3

8.3.1. Comparação de modelos de aprendizado de máquina para análise preditiva da febre amarela em Minas Gerais/Brasil.

Comparasion of machine learning models for predictive analyses of yellow fever in Minas Gerais/Brazil.

Resultados e discussão

Para avaliação de multicolineariedade entre as variáveis de entrada ou independentes afim de evitar redundância de informações construiu-se uma matriz de correlação na qual valores de coeficiente de Pearson (r) foram calculados para cada par de covariáveis (Fig. 1).

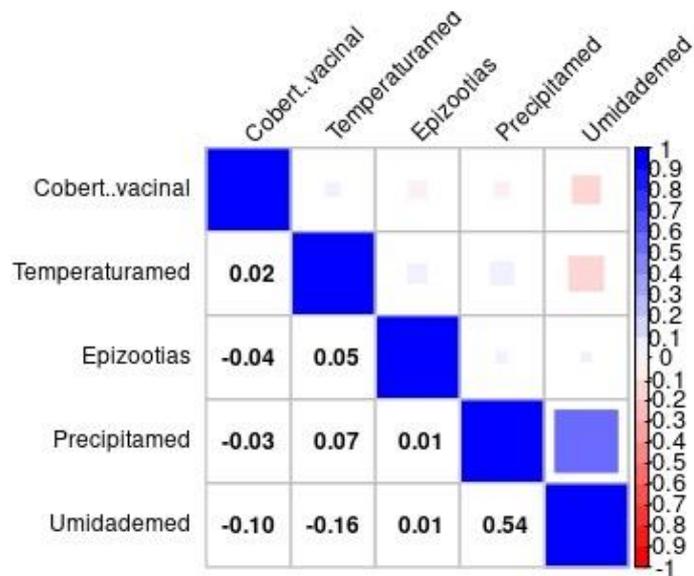


Figura 1: Matriz de correlação (r) entre as covariáveis utilizadas no estudo de previsão da FA em Minas Gerais entre 2007 e 2020.

Considerando um ponto de corte de $r \geq 0,8$, nenhuma variável foi excluída por apresentar redundância de informações. A relação entre variáveis de entrada com maior r observado foi umidade média diária e precipitação com $r=0,54$, entretanto, a força dessa relação é considerada média e não há, portanto, risco de redundância entre as informações.

O algoritmo Boruta, um seletor de variáveis, foi utilizado para selecionar variáveis importantes para os modelos de previsão. Observou-se que nenhum recurso ou variável foi excluída pelo método realizado. Todas as variáveis apresentaram-se melhores que as sombras criadas pelo algoritmo para comparar com as variáveis entrada (Fig. 2).

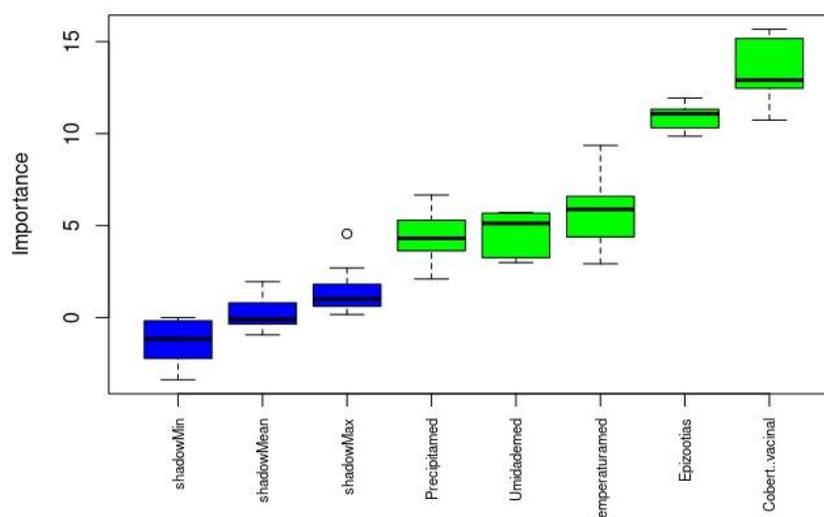


Figura 2: Resultado da análise do seletor de variáveis Boruta utilizado na pesquisa.

Na figura 2 pode-se observar que cobertura vacinal, epizootias e temperatura média diárias foram as variáveis com maior relevância para o algoritmo Boruta, entretanto, todas as variáveis avaliadas foram consideradas relevantes.

Diversos trabalhos que utilizam modelos de aprendizado de máquina realizam a seleção de recursos (variáveis) por meio do algoritmo Boruta. Leong e Addulah (2019) utilizaram com êxito o algoritmo Boruta para selecionar recursos em trabalho de previsão de Alzheimer assim como Kaur e Kumari (2020) para previsão de diabetes.

A seleção de variáveis ou recursos geralmente é uma importante etapa na aplicação de métodos de aprendizado de máquina. Normalmente, quando um grande número de variáveis é apresentado ao modelo, uma parte significativa pode se apresentar como irrelevante para previsão ou classificação, e obviamente sua relevância não é conhecida antecipadamente (Kursa e Rudnicki, 2010).

A permanência de recursos irrelevantes pode aumentar o custo computacional devido a presença de ruídos desnecessários. Trabalhar com um conjunto de recursos minimamente possível tendo melhores resultados é conhecido como problema mínimo-ideal (Nilsson *et al.*, 2007).

Lidar com grande número de variáveis ou atributos podem tornam os algoritmos mais lentos e muitos algoritmos de aprendizado de máquina apresentam uma diminuição de precisão, quando o número de variáveis é significativamente maior do que o ideal (Kursa e Rudnicki, 2010).

A Figura 3 apresenta o comportamento mensal das variáveis independentes ou de entrada da macrorregião de saúde centro em Minas Gerais, entre os anos de 2007, utilizados na fase de treinamento dos modelos (Fig.3).

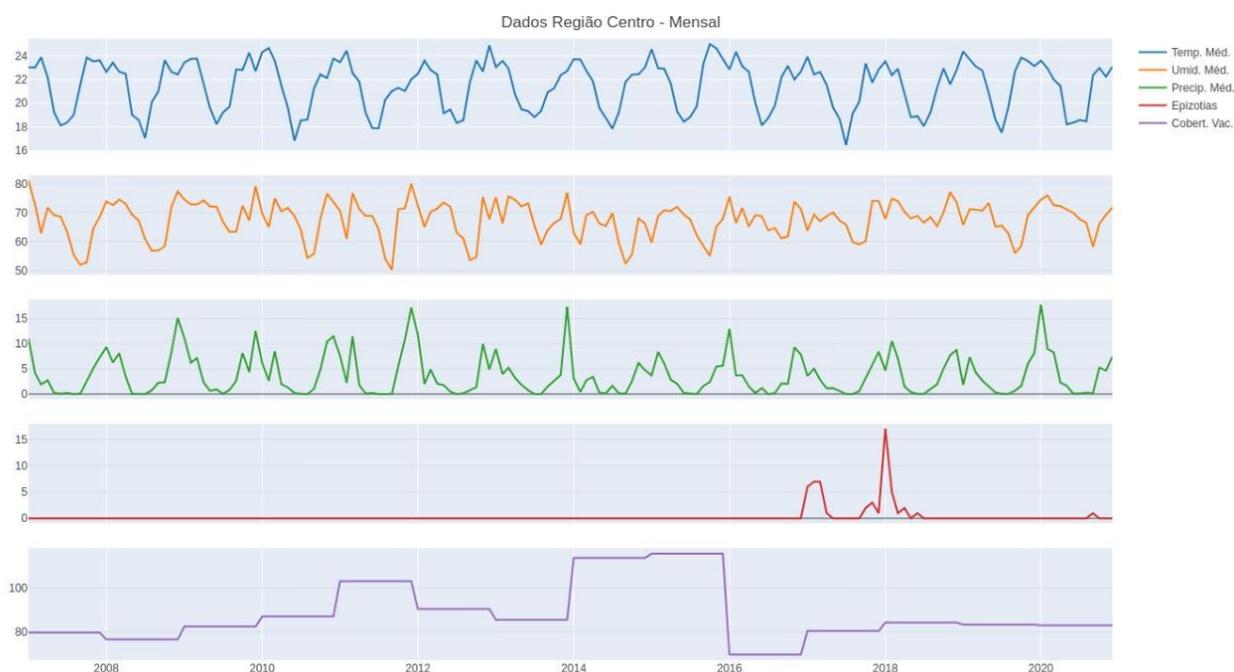


Figura 3: Comportamento mensal das variáveis independentes do conjunto de treino (macrorregião de saúde centro) utilizados no sistema de previsão para a série histórica entre 2007 e 2020

De acordo com as curvas dispostas na Figura 3, é possível observar o comportamento das variáveis de entrada utilizadas no treinamento dos modelos de aprendizado de máquina, sendo que as variáveis meteorológicas apresentam características cíclicas para os mesmos períodos dos anos e dados bem distribuídos. Epizootias em PNH com ocorrência em dois momentos da série, um antecedendo a ocorrência de casos humanos e outro no mesmo momento de ocorrência. Cobertura vacinal oscilou ao longo da série com aumento em 2014 e 2016, após queda manteve estável até o fim da série.

Na tabela 1 pode-se observar as medidas de média (MED), valor máximo (MAX), valor mínimo (MIN), desvio padrão (DP) e coeficiente de variação (CV) dos valores das variáveis meteorológicas, epizootias, cobertura vacinal e FA do conjunto de treinamento da macrorregião de saúde centro (Tab.1).

Tabela 1: Estatística dos valores das variáveis observadas utilizadas nos modelos de previsão no período de 2007 a 2020 em Minas Gerais do conjunto de treino (macrorregião de saúde centro)

Variáveis	MED	MAX	MIN	DP	CV(%)
Temperatura média diária	21,46	28,23	13,42	2,54	11,87
Precipitação média diária	3,72	104,60	0,00	8,32	223,39
Umidade média diária	67,61	93,80	36,60	9,59	14,18
Cobertura vacinal média anual	88,20	115,53	69,75	13,42	15,20
Epizootias confirmadas diárias	0,01	3,00	0,00	0,12	1150,27
Febre amarela confirmados diários	0,04	10,00	0,00	0,47	1046,98

O coeficiente de variação (CV) mede a variabilidade de um conjunto de dados. Indica o quanto um conjunto se afasta ou aproxima da média, ou seja, indica homogeneidade ou heterogeneidade nos dados.

Observou-se um alto CV para epizootias, uma das variáveis de entrada e para série de FA variável de saída dos modelos, demonstrando um grande grau de dispersão em relação à média e, portanto, alta heterogeneidade.

Em relação aos dados meteorológicos observou-se ausência de medição de dados discrepantes e homogeneidade, com exceção da precipitação média diária que apresentou um maior CV indicando maior heterogeneidade em relação aos outros dados.

Na Figura 4 observa-se o comportamento das variáveis independentes ou de entrada do conjunto de teste da macrorregião de saúde nordeste entre os anos de 2007 e 2020 (Fig.4).

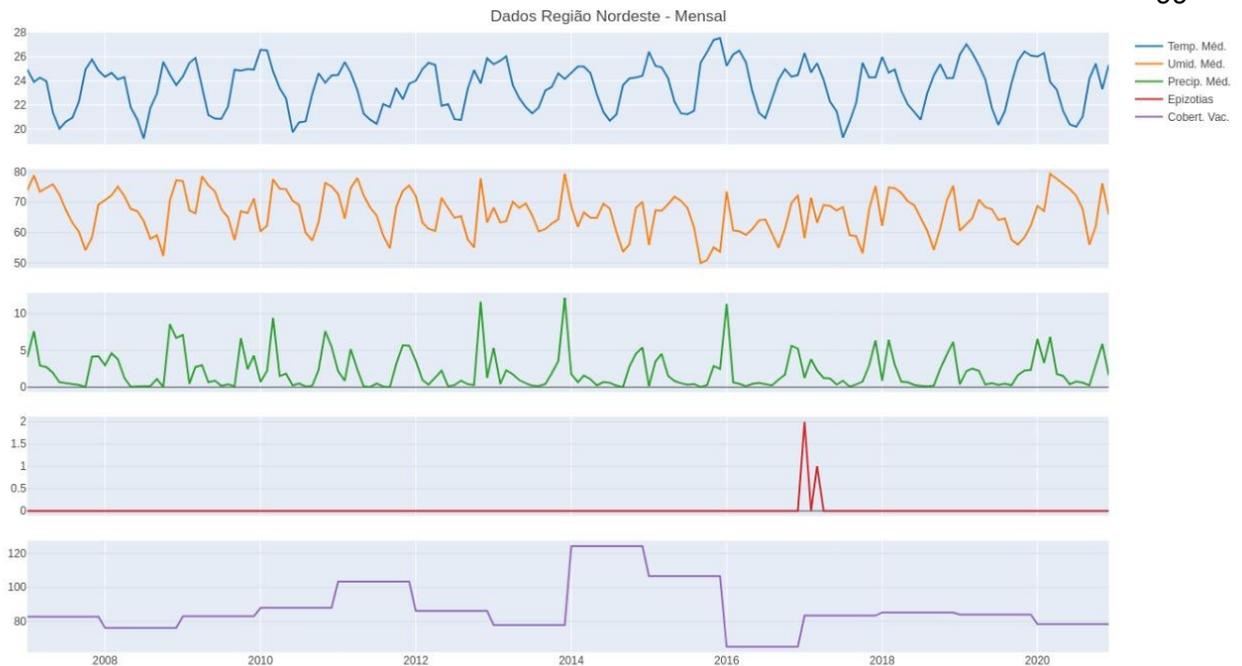


Figura 4: Comportamento das variáveis independentes dos grupos de teste (macrorregião de saúde nordeste) utilizados no sistema de previsão para a série histórica entre 2007 e 2020

De acordo com as curvas dispostas na figura 5, é possível observar o comportamento das variáveis de entrada e saída utilizadas na fase de teste dos modelos de aprendizado de máquina, sendo que as variáveis meteorológicas apresentam características cíclicas para os mesmos períodos dos anos e dados bem distribuídos. Epizootias em PNH com distintos padrões de ocorrência entre duas regiões de teste. Número e época de ocorrência de casos de FA humana foram verificados em divergência entre os dois conjuntos de teste. Cobertura vacinal oscilou ao longo da série apresentado aumento acentuado entre 2014 e 2016 e manteve-se estável entre 2017 e 2020.

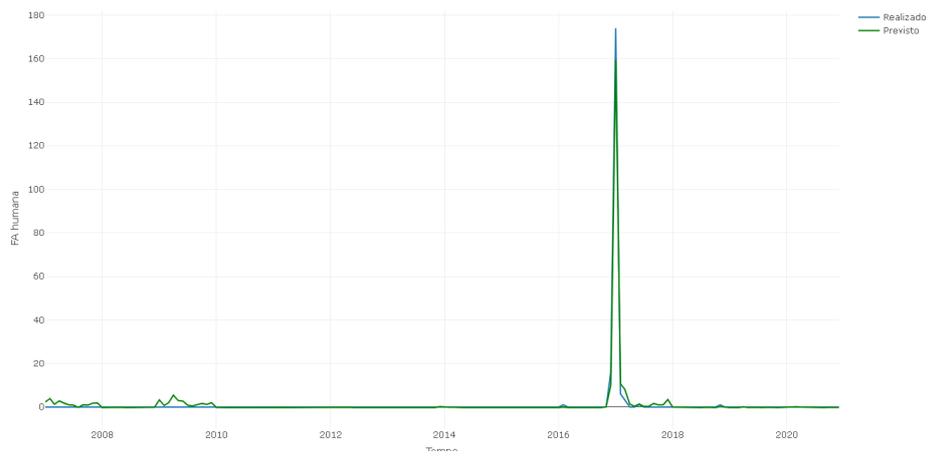
Na tabela 2 pode-se observar as medidas de média (MED), valor máximo (MAX), valor mínimo (MIN), desvio padrão (DP) e coeficiente de variação (CV) dos valores das variáveis meteorológicas, epizootias, cobertura vacinal e FA para os dados da macrorregião de saúde nordeste utilizados na fase de teste dos modelos (Tab.2).

Tabela 2: Estatística dos valores das variáveis observadas utilizadas nos modelos de previsão no período de 2007 a 2020 em Minas Gerais para os dados de teste (macrorregião de saúde nordeste)

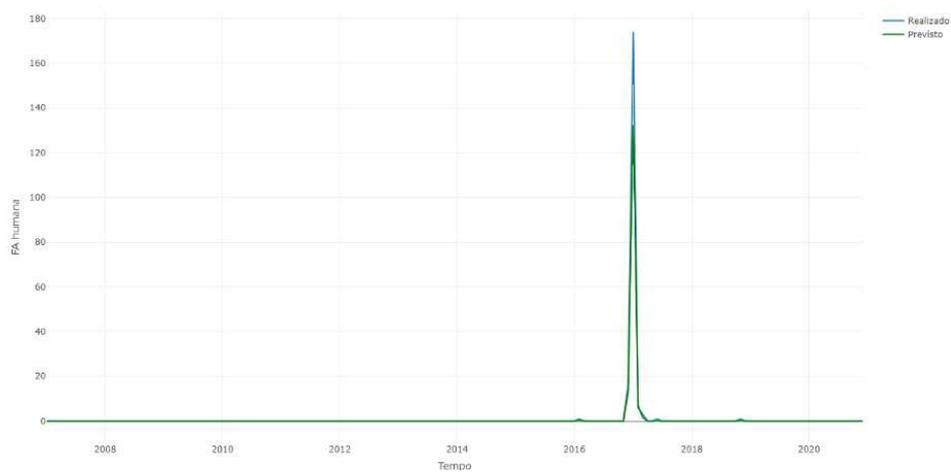
Variáveis	Média	Máx.	Mín.	Dp	CV(%)
Temperatura média diária	23,60	31,25	15,75	2,39	10,15
Precipitação média diária	2,14	63,75	0,00	5,77	269,06
Umidade média diária	66,36	94,00	33,25	9,89	14,09
Cobertura vacinal média anual	0,874	124,14	65,06	14,81	16,94
Epizootias confirmadas diárias	0,00	1,00	0,00	0,02	4127,95
FA confirmados diários	0,03	14,00	0,00	0,51	1295,09

Assim como nos dados de treinamento observou-se um alto CV para as variáveis epizootias e FA silvestre, nos grupos de teste o mesmo ocorreu. Na macrorregião de saúde Nordeste o CV para epizootias foi extremamente alto indicando alta dispersão dessa variável em relação à média.

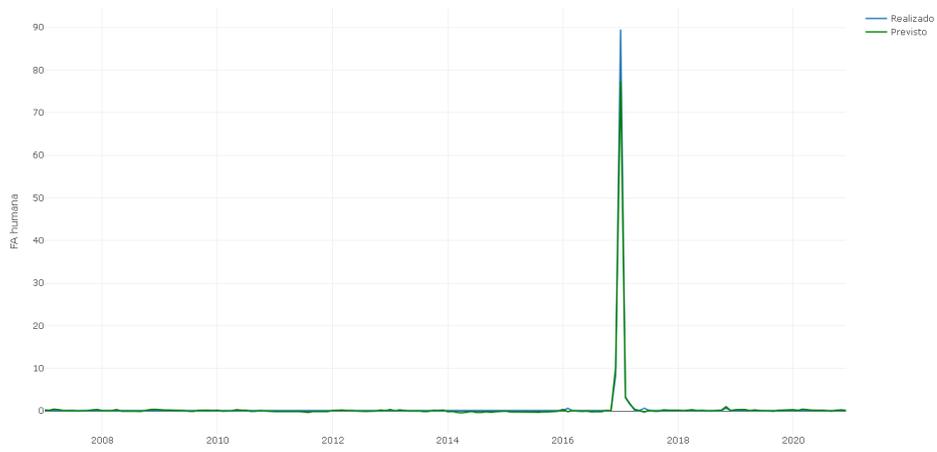
Os resultados das previsões na fase de teste obtidos por cada um dos modelos de aprendizado de máquina podem ser observados na Figura 5, em que o random forest obteve melhor desempenho de previsão com a amostra de dados da macrorregião nordeste em Minas Gerais entre 2007 e 2020 (Fig.5).



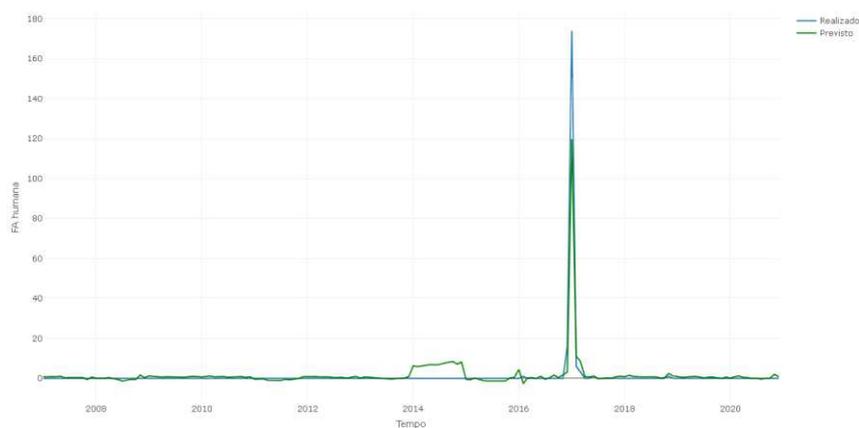
(a) RF



(b) Xgb



(c) MPL



(d) SVM

Figura 5: Valores observados e previstos mensais pelas modelagens de RF (a), MLP (b), Xgboost (c) e SVM (d) para previsão de casos de febre amarela em Minas Gerais entre 2007 e 2020.

Observa-se que nenhum dos modelos foi capaz de prever o número de caso máximo diário de FA para macrorregião de saúde nordeste, entretanto seguiram muito bem a tendência de aumento na época em que os mesmos ocorreram.

Para melhor visualização da previsão dos casos de FA um zoom no gráfico de previsão diária (Fig. 6).

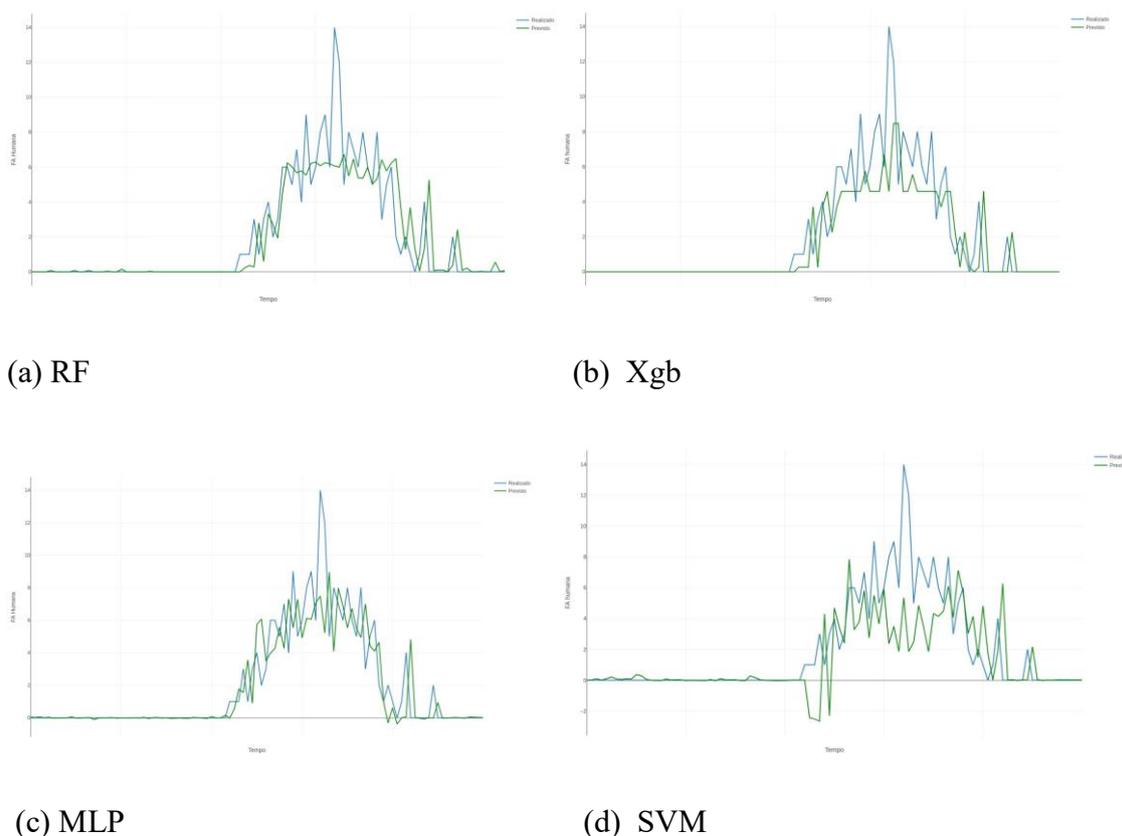


Figura 6: Valores observados (azul) e previstos (verde) pelas modelagens: RF (a), Xgb (b), RNA (c) e SVM (d) para previsão de casos de febre amarela em Minas Gerais diários com zoom na época de ocorrência de casos de FA.

Comparando-se as métricas de avaliação de performance dos modelos de aprendizado de máquina otimizados para previsão da FA, observou-se que o modelo RF (a) na Figura 6, apresentou melhor desempenho na previsão de casos de FA na macrorregião de saúde nordeste apresentando $MSE= 0,2232149$, $RMSE= 0,4724563$, $R^2= 0,7768282$ e $MAE= 0,07545407$. O modelo com pior desempenho na previsão de casos de FA para região nordeste foi o SVM (Tab. 3).

Tabela 3: Métricas de avaliação de desempenho dos modelos de aprendizado de máquina utilizados na previsão da febre amarela em Minas Gerais entre 2007 e 2020 para região nordeste.

Métrica	RNA-MLP	XGboost	SVM	Random Forest
MSE	0,2407423	0,2423376	0,4484465	0,2232149
R ²	0,7593042	0,7577092	0,5516402	0,7768282
RMSE	0,490655	0,492278	0,6696614	0,4724563
MAE	0,07935186	0,0439027	0,138301	0,07545407

Apresentando o maior coeficiente de determinação R^2 o modelo random forest otimizado explica 77,68% da variância dos dados. Xgboost e RNA-MLP também apresentaram resultados satisfatórios. O algoritmo SVM apresentou pior performance de previsão para os dados disponibilizados.

Devido à natureza de “caixa preta” dos algoritmos de aprendizado de máquina, é difícil explicar quais características específicas do problema levam a uma determinada previsão, o que seria de certa forma uma desvantagem de uso de aprendizado de máquina na área da saúde, principalmente quando se trabalha com previsão envolvendo diagnóstico clínico.

Uma alternativa para explicar as previsões e resolver essas desvantagens, é a combinação da previsão por meio algoritmo de aprendizado de máquina com uma estrutura baseada em SHAPley Additive exPlanations (SHAP) (Ludenberg e Lee, 2017). O método baseia-se na teoria dos jogos de coalizão denominado valor de Shapley (Shapley, 1953).

Para determinar os principais preditores de febre amarela calculou-se a importância dos recursos de regressão do modelo final (Fig. 7).

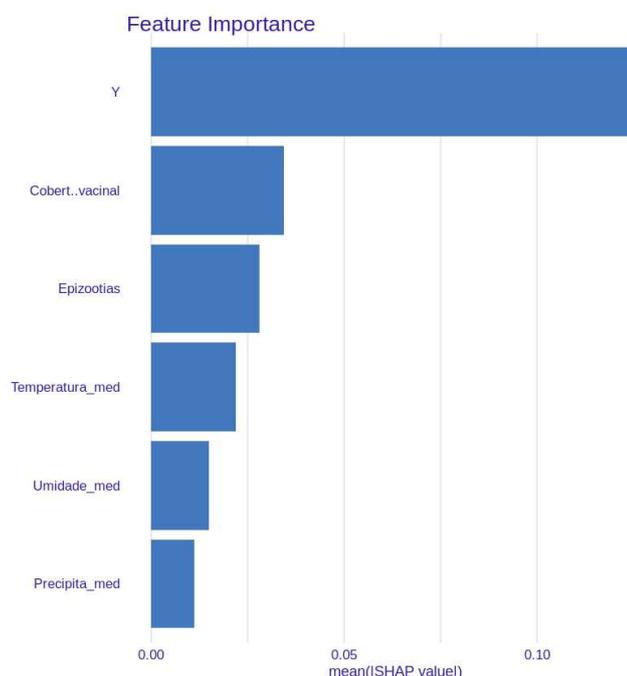


Figura 7: Análise de importância das variáveis – Shap value para o modelo random forest com melhor desempenho de previsão dos casos de febre amarela em Minas Gerais entre 2007 e 2020.

Os valores SHAP mostram a contribuição de cada recurso para a previsão final. A *lag feature* (Y) criada para tornar o problema de aprendizado de máquina supervisionado em um problema de previsão de série temporal foi aquela com maior importância para o problema em questão (Fig. 7). Cobertura vacinal, epizootias e temperatura média também contribuíram para previsão de forma significativa e com baixa discrepância no Shap value.

As epizootias são consideradas pelo sistema de vigilância da FA como eventos sentinela da doença e sua ocorrência norteia atividades preventivas como intensificação de vacinação da população sob risco de adoecer.

A temperatura afeta a distribuição espaço-temporal dos vetores de doenças nas escalas global e regional, e como a temperatura continua a aumentar devido às mudanças climáticas e emissões de gases de efeito estufa, insetos vetores que habitam regiões de baixa latitude podem se expandir para novos habitats em latitudes médias ou altas regiões e em altitudes mais elevadas, levando a expansões geográficas ou mudanças nas faixas de doenças (Nava *et al.*, 2017).

Limitações da pesquisa

O presente trabalho retrata uma análise experimental de previsão de série temporal multivariada que utiliza como um dos dados de entrada além de dados meteorológicos e de cobertura vacinal, as notificações de epizootias em PNH confirmados para FA. Apesar da importância da vigilância em PNH, o baixo número de amostras positivas presentes nos registros do SINAN para FA em Minas Gerais, não tem permitido demonstrar a real situação das populações de PNH acometidos pela doença. Certamente a má qualidade dessas informações não demonstrando a real situação impacta de forma negativa nos modelos de previsão.

Conclusão

Problemas de previsão de série temporal podem ser trabalhados como problema de aprendizado de máquina supervisionado. Trabalhar com previsão de doenças como a febre amarela são de extrema importância para a tomada de decisão em saúde pública e pode impactar de forma positiva a vida das populações sob risco de adoecerem. Os modelos de previsão em aprendizado de máquina mostraram-se uma boa alternativa para trabalhos envolvendo várias séries preditoras. Em geral, várias contribuições foram feitas neste estudo. Desenvolver e comparar o desempenho de diferentes algoritmo de aprendizado de máquina incluindo XGBoost que tem atraído atenção por sua rápida velocidade de cálculo, forte capacidade de generalização e alto

desempenho preditivo, random forest com melhor desempenho no presente estudo, RNA-*MLP* e SVM demonstra que para cada conjunto de dados trabalhado um modelo apresentará o melhor desempenho. Em combinação com outros conhecimentos avançados de aprendizado de máquina, como preenchimento de valores ausentes baseado no algoritmo MICE e seleção de recursos baseada no Boruta, otimização de hiperparâmetros baseada em GridSearch também foram usadas. O uso de algoritmos de aprendizado de máquina para previsão de séries temporais apresenta-se como um grande desafio. Desenvolver modelos de aprendizado de máquina com alto desempenho preditivo que possam prever doenças auxiliarão as tomadas de decisão em saúde pública e trarão impactos positivos na vida das populações sob risco de adoecerem.

9. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Buscou-se através de uma revisão integrativa da literatura (artigo 1) encontrar fatores eco-epidemiológicos que possam estar envolvidos nas ondas epidêmicas de febre amarela entre 2016 e 2018 na região Sudeste do Brasil.

Condições ecológicas e climáticas demonstraram ser adequadas para manutenção do VFA durante as estações epidêmica e inter epidêmica baseado em estudos relacionados à região Sudeste do Brasil. Este fato apresenta riscos para o estabelecimento de ciclos enzoóticos e epidemias de FA fora da bacia Amazônica em áreas densamente povoadas e densamente habitadas por *Aedes aegypti*, vetor urbano da doença, principalmente em regiões onde há presença de mata Atlântica.

O comércio ilegal de animais silvestres, desmatamento, as alterações climáticas consequentes das modificações antrópicas no meio ambiente são exemplos de ações que podem ter favorecido a ocorrência de FA em tamanha magnitude.

Algumas barreiras ambientais/ecológicas ao inibir os movimentos de vetores e hospedeiros podem prevenir o transbordamento de FAV de um ciclo silvestre.

No artigo 2 descreveu-se as notificações do SINAN de casos suspeitos e confirmados de febre amarela humana e de epizootias em PNH no estado de Minas Gerais entre 2007 e 2020.

Apesar do alto número de noificações registradas de PNH com suspeita de FA, poucas epizootias foram registradas no SINAN, indicando possivelmente falha no sistema de vigilância de PNH. Além disso a baixa qualidade das carcaças, indicado pelo alto percentual de casos

indeterminados, não tem permitido demonstrar a real situação das populações de PNH acometidos pela doença. Certamente a má qualidade dessas informações, não demonstrando a real situação, impacta de forma negativa nos modelos de previsão. Vale ressaltar que epizootias em PNH por febre amarela são considerados eventos sentinela para vigilância e devem ser utilizados como alerta para bloqueio vacinal em regiões em que a cobertura da população está abaixo do ideal. Ações sócio educativas junto à população sobre a importância da vacinação da FA e vigilância de PNH devem ser realizadas periodicamente.

No artigo 3 utilizou-se dados das notificações confirmadas de febre amarela humana e de epizootias em PNH do SINAN, além de cobertura vacinal da população de estudo e variável relacionados ao ambiente (temperatura, umidade e precipitação médias diárias) prever casos de febre amarela em uma macrorregião de saúde de Minas Gerais.

Técnicas supervisionadas de aprendizado de máquina foram aplicadas a problema de regressão para prever resposta quantitativa envolvendo série temporal multivariada de interesse para saúde pública. Trabalhar com inferência causal não foi o propósito do presente trabalho, diante disso não se realizou estudos de associação entre as variáveis preditoras e a variável predita, embora seja relevante para os estudos epidemiológicos. As variáveis ou séries temporais preditoras foram selecionadas inicialmente através de uma revisão bibliográfica da literatura científica a respeito dos fatores eco-epidemiológicos envolvidos nas ondas epidêmicas e epizoóticas de febre amarela ocorridos na região sudeste do Brasil entre os anos de 2016 e 2018.

A aplicação das técnicas de aprendizado de máquina seguiu um roteiro metodológico apresentado no capítulo 5 da tese. As variáveis qualitativas foram padronizadas, dados faltantes foram imputados por técnica multivariada, algoritmo Boruta para seleção de recursos foi utilizado e hiperparâmetros dos modelos foram otimizados na fase de aprendizado.

Para tratar o problema de aprendizado de máquina em problema de previsão de série temporal, realizou-se o método de janela deslizante criando a *lag feature*, o que traz uma ordem temporal ao banco de dados. Foram realizados cálculos das estatísticas descritivas das séries temporais que compõe o banco de dados para auxiliar na percepção sobre distribuição e dispersão dos valores utilizados no trabalho.

Dos quatro modelos treinados e testados com dados desconhecidos, o RF apresentou melhor desempenho na previsão de febre amarela humana.

È importante destacar que mesmo que um modelo seja otimizado e apresente boa performance preditiva, as tomadas de decisão e ações junto às populações é que farão a diferença na redução de casos de febre amarela humana.

O uso de algoritmos de aprendizado de máquina e métodos estatísticos para produção de conhecimento a partir de dados tem se tornado componente útil e imprescindível para área da saúde, seja ela pública ou saúde de pacientes por meio de diagnóstico utilizando aprendizado de máquina para classificação ou regressão.

Estudos tem evidenciado que modelos de aprendizado de máquina apesar de serem aptos a trabalharem com problemas de regressão de classificação, não há um modelo ideal, sendo necessário comparar vários modelos de acordo com a performance demonstrada em um determinado banco de dados.

Espera-se que esse trabalho sirva como norteador para pesquisas de envolvam previsão de séries temporais e algoritmos de aprendizado de máquina no contexto da saúde com objetivo de melhorar a vida das populações, sobretudo aquelas sob risco de adoecerem através de tomadas de decisão baseadas em análises de dados.

10. REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, F.V.S. et al. *Haemagogus leucocelaenus* and *Haemagogus janthinomys* are the primary vectors in the major yellow fever outbreak in Brazil, 2016-2018. *Emerg Microbes Infect.* v.8, n.1, p.218-231, 2019.
- ALBANO BR, BASÍLIO MC, NEVES JB. Desafios para a inclusão dos homens nos serviços de atenção primária à saúde. *Revis. Enferm. Int.*3(2), 2010.
- ALENCAR, J. A. de; LOPES, C. M. et al. Ausência da diferenciação isoenzimática entre *Haemagogus capricornii* e *Hg. janthinomys* (Díptera: Culicidae).19 *Congresso Brasileiro de Entomologia*, 2002.
- APRESENTAÇÃO CARTOGRÁFICA PDR 2011. Disponível em: <https://www.saude.mg.gov.br/images/documentos/APRESENTACaO%20CARTOGRAFICA%20PDR-2011.pdf>. Acesso em 01 maio 2019.
- AUGUSTIN G. History of yellow fever. Searcy & Pfaff ltd, New Orleans, USA; 1909.
- BARCELLOS, C.; MONTEIRO, A. M. V.; CORVALÁN, C. *et al.* Climate and environmental changes and infectious diseases: scenarios and uncertainties for Brazil Health surveillance. *Epidemiol. Serv. Saúde*, Brasília, v.18, n.3, p.285-304, 2009.
- BELLE V, PAPANTONIS I. Principles and Practice of Explainable machine learning. *Front Big Data*. Jul 1;4:688969 2021.
- BENCHIMOL, J. Febre amarela: a doença e a vacina, uma história inacabada. Rio de Janeiro: FIOCRUZ; 2001. 469 p.
- BEN-HUR, A.; WESTON, J. A user's guide to support vector machines. v. 609, 2008.
- BÉRENGER-FÉRAUD, L.-J.-B. 1890 *Traité théorique & clinique de la fièvre jaune* par L.-J.-B. Bérenger Féraud, directeur du Service de Santé, président du Conseil Supérieur de Santé de la Marine, membre correspondant de l'Academie de Médecine etc. Paris, Octave Doin.
- BICCA-MARQUES J. C.; FREITAS D. S. The role of monkeys, mosquitoes and human in the occurrence of a yellow fever outbreak in a fragmented landscape in south Brazil: protecting howler monkeys is a matter of public health. *Trop Conserv Sci*.v.3,n.1, p.78-89, 2010.

BICCA-MARQUES, J. C. Outbreak of yellow fever affects howler monkeys in southern Brazil, *Oryx*, v.43 p. 173, 2009.

BONALDO, M. C.; GOMEZ, M. M.; SANTOS, A. AC. *et al.* Genome analysis of yellow fever virus of the ongoing outbreak in Brazil reveals polymorphisms. *Mem. Inst. Oswaldo Cruz.* v.112, n.6, p. 447-451, 2017.

BOSER., B. E.; GUYON, I. M.; VAPNIK V. N. “A Training Algorithm for Optimal Margin Classifiers”. In: Proceedings of the fifth annual workshop on Computational learning theory, p.144–152, 1992.

BRAGA, A. P.; LURDEMIR, T. B.; CARVALHO, A. C. P. L. F. *Redes Neurais Artificiais: Teoria e Aplicações.* LTC - Livros Técnicos e Científicos Editora S.A. Belo Horizonte, Brasil, 2000.

BRASIL. Ministério da Saúde. Manual de vigilância epidemiológica da Febre Amarela. Ministério da Saúde Brasília: 2004. 63p. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/manual_vigilancia_epid_febre_amarela.pdf. Acessado em: 17 abr 2020.

BRASIL. Ministério da Saúde. Política nacional de atenção integral à saúde do homem. Brasília: MS; 2008.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. In: Diário Oficial da República Federativa do Brasil, Brasília, DF, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm . Acesso em 10 jan 2022.

BRASIL. Ministério da Saúde. (Brasil, 2017a). Secretaria de Atenção à Saúde. Febre Amarela: Guia para profissionais de saúde. Ministério da Saúde, Secretaria de Atenção à Saúde. Brasília: Ministério da Saúde, 2017a .67: il. ISBN 978-85-334-2480-7.

BRASIL. Ministério da Saúde. (Brasil, 2017b). Secretaria de Vigilância em Saúde. Informe especial febre amarela no Brasil nº 01/2017. Brasília: Ministério da Saúde, 2017b. Disponível em: <<http://portalarquivos.saude.gov.br/images/pdf/2017/marco/18/Informe-especial-COES-FA.pdf>> Acessado em: 24 agosto de 2021.

BRASIL. Ministério da Saúde. (Brasil, 2017c). Guia de vigilância de epizootias em primatas não humanos e entomologia aplicada à vigilância da Febre Amarela. Brasília: Ministério da Saúde, 2.ed atualizada. Brasília, p.100, 2017c.

BRASIL. Ministério da Saúde. (Brasil, 2018a). Febre Amarela: Ministério da Saúde atualiza casos no país. 2018a.. Disponível em: <https://www.saude.gov.br/noticias/agencia-saude/42940>. Acesso em: 26 ago. 2021.

BRASIL. (Brasil, 2018b). LEI nº 13.709 de 14 de agosto de 2018. Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. In: Diário Oficial da República Federativa do Brasil, Brasília, DF, 14 ago. 2018b. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm> . Acesso em 10 jan 2022.

BRASIL. Ministério da Saúde. (Brasil, 2018c). Secretaria de Vigilância em Saúde. Monitoramento do Período Sazonal da Febre Amarela Brasil – 2017/2018. Informe nº 15 | 2017/2018. 2018c. Disponível em:<https://saude.campinas.sp.gov.br/doencas/febre_amarela/2018/Informe_MS_15_F_Amarela_monitoramento.pdf>. Acessado em: 12 ago 2022.

BRASIL. Ministério da Saúde. (Brasil, 2018d). Informe n. 21. 2017/2018. Monitoramento do período sazonal da febre amarela. Brasil 2017/2018. Ano 2018d. Disponível em: <https://sbim.org.br/images/files/informe-fa-21-11abr18-c.pdf>. Acesso em: 1 ago. 2022.

BRASIL. Ministério da Saúde. (Brasil, 2019a). Secretaria de Vigilância em Saúde. Guia de Vigilância em Saúde: volume único [recurso eletrônico] / Ministério da Saúde. 3ª ed. – Brasília: Ministério da Saúde, 2019a. 740 p.

BRASIL. Ministério da Saúde. (Brasil, 2019b). Secretaria de Vigilância em Saúde. Monitoramento sazonal da febre amarela. Informe nº 2, 2018/2019. Brasília: Ministério da Saúde, 2019b.

BRASIL. Ministério da Saúde. (Brasil, 2018c). Saúde Brasil 2019 uma análise da situação de saúde com enfoque nas doenças imunopreveníveis e na imunização. Brasília: Ministério da Saúde, 2019c. 520 p.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Boletim epidemiológico nº12. Ministério da Saúde, Brasília. Março, V.51. Ano 2020.

BRASIL. Ministério da Saúde. (Brasil, 2021a). Secretaria de Vigilância em Saúde. Boletim Epidemiológico nº 04. Brasília. vol. 52, fev. 2021a.

BRASIL. Ministério da Saúde. (Brasil, 2021b). Febre amarela. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/f/febre-amarela-1>. Acesso em: 1 ago 2022.

BREIMAN, L. Out-of-bag estimation. 1996. Disponível em: <<https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf>>. Acesso em ago. 2022.

BREIMAN, L. Random forests. *Machine learning*, v.45, n.1, p.5–32, 2001.

BROWNLEE J. Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future. *Machine learning Mastery*, 2017 – 367p.

BRYANT, J. E.; HOLMES, E. C; BARRETT, A. D. T. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog* v.3, n.5: e75, 2007.

CAMARGO, A. P; SENTELHAS, P. C. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. *Rev. Bras. de Agromet.*, v.5, n.1, p.89–97, 1997

CAMARGO-NEVES, V.L.F.; POLETTO, D.W.; RODAS, L.A.C.; *et al.* Entomological investigation of a sylvatic yellow fever area in São Paulo State, Brazil. *Cad. de Saú. Púb.* v. 21, p. 1278-1286, 2005.

CARDINALE, B., DUFFY, J., GONZALEZ, A. *et al.* Biodiversity loss and its impact on humanity. *Nature* v.486, p.59–67, 2012.

CARVALHO, B., RANGEL, E., VALE, M. Evaluation of the impacts of climate change on disease vectors through ecological niche modelling. *Bull. of Entomol. Res.*, v.107, n.4, p.419-430, 2017.

CASALI, F. T.; MUNICCI, G. S.; MEIRA, A. L. P. *et al.* Análise das características epidemiológicas da febre amarela em um estado da Região Sudeste do Brasil. *Rev. Enf. act. Cost Rica*. 37, pp. 50-65, 2019.

CAVALCANTE K. R. L. J.; TAUIL P. L. Características epidemiológicas da febre amarela no Brasil, 2000-2012. *Epidemiol Serv Saúde*. v. 25,n.3, p. 11-20, 2016.

CHILDS M. L.; NOVA N.; COLVIN J.; MORDECAI E. A. Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. *Philos Trans R Soc Lond B Biol Sci*. v.374, p.1-14, 2019.

CHIPPAUX, J. E.; CHIPPAUX, A. Yellow fever in Africa and the Americas: a historical and epidemiological perspective. *Jour. of Vem. AniM. and Toxins incl. Trop. Diseases*, p.24:20, 2018.

CIVES. Centro de informação de saúde do viajante. Febre amarela no Brasil: 2008. <http://www.cives.ufrj.br/informacao/fam/fam-2008.html>. Acesso em: 02 ago. 2022.

COELHO, F.G.F. *Semi-supervised feature selection*.2013, 127p. (PhD Thesis) Universidade Federal de Minas Gerais and Université catholique de Louvain. Belo Horizonte; Louvain.

COSTA Z. G. A. *et al.* Evolução histórica da vigilância epidemiológica e do controle da febre amarela no Brasil. *Rev Pan-Amaz Saude*; v.2, n.1. p.11-26, 2011.

COUTO-LIMA, D.; MADEC, Y.; BERSOT, M.I. *et al.* Potential risk of re-emergence of urban transmission of Yellow Fever virus in Brazil facilitated by competent *Aedes* populations. *Sci Rep* v.7, n.4848, p.1-12, 2017.

CULOT L, PEREIRA L A, AGOSTINI I, *et al.* ATLANTIC-PRIMATES: um conjunto de dados de comunidades e ocorrências de primatas na Mata Atlântica da América do Sul. v. 100. *Ecologia*. 2019.

CUNHA S. M.; FARIA N. R.; CALEIRO G. S. *et al.* Genomic evidence of yellow fever virus in *Aedes scapularis*, southeastern Brazil, 2016. *Acta Tropica*, v.205, 2020.

CUNHA, M. S.; DA COSTA, A. C.; DE AZEVEDO FERNANDES, N.C.C. *et al.* Epizootics due to Yellow Fever Virus in São Paulo State, Brazil: viral dissemination to new areas (2016–2017). *Sci Rep* v.9, 5474, p.1-13 2019.

CVE – Centro de Vigilância Epidemiológica “Prof. Alexandre Vranjac”. A. Informe técnico febre amarela. Centro de vigilância epidemiológica. São Paulo. Disponível em: <https://www.ribeiraopreto.sp.gov.br/files/ssaude/pdf/inf-tec-febre-amarela.pdf>. Acesso em: 01 ago. 2022.

DE PAIVA C. A. *et al.* Determination of the spatial susceptibility to Yellow Fever using a multicriteria analysis. *Mem Inst Oswaldo Cruz*. v.114, p.1-8, 2019.

DELFINO, D. A. A. *Investigação epidemiológica de epizootia de Febre Amarela em primatas não-humanos em Aparecida de Goiânia, Goiás, Brasil*. 2011. 36f. (Dissertação de Mestrado Ciência Animal). Universidade Federal de Goiás, Goiânia.

DÍAZ, S.; SETTELE, J.; BRONDÍZIO E. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). 2019. Disponível em: <https://www.luke.fi/wp-content/uploads/2019/05/190506-ipbes-Summary-for-Policymakers-IPBES-Global-Assessment.pdf> . Acessado em 25 out 2020.

EHLERS, R. S. *Análise de Séries Temporais*. 2009. Acessado em: 01/10/2021. Disponível em: www.icm.usp.br/ehlers/stemp/stemp.pdf.

ESCOSTEGUY, C. C.; PEREIRA A. G. L.; MARQUES, M. R. V. E. *et al.* . Febre amarela: perfil dos casos e fatores associados ao óbito em hospital referência no estado do Rio de Janeiro, 2017–2018. *Rev Saude Publ.* v.53, n.89. p. 1-12, 2019

ESPINOSA, L. A.; PORTELA, M. M.; RODRIGUES, R. Spatio-temporal variability of droughts over past 80 years in Madeira Island. *Jour.of Hydro: regional studies*, v. 25, p. 1-18, 2019.

FACELI, A. C. K.; LORENA, J.; GAMA *et al.* Inteligência artificial: Uma abordagem de aprendizado de máquina. 2011.

FAHRIG, L. Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.* v.34, p 487–515, 2003.

FANTINI, D. C.; ALVES, V.; PASTOR, M. V. D. *et al.* Perfil epidemiológico da febre amarela da região Sul do Brasil, de 2007 a 2019. *Braz. Jour. of Health Review*, v.4, n.1, p.891-907, 2021.

FARIA N. R.; KRAEMER M. U. G.; HILL S. C. *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*. v. 361, n.6405, p.894-899, 2018.

FERRAZ L, TRINDADE LL, BEVILAQUA E, *et al.* As demandas do homem rural: informações para a assistência nos serviços de saúde da atenção básica. *Rev Min Enferm.* v.17, n.2, p. 349-55. 2013.

FIGUEIREDO P. O.; SILVA A. T. S.; OLIVEIRA J. S.; *et al.* Detection and Molecular Characterization of Yellow Fever Virus, 2017, Brazil. *Ecohealth*. v.15, n. 4, p. 864-870, 2018.

FINLAY C. El mosquito hipotéticamente considerado como agente de transmisión de la fiebre amarilla. *An Acad Cien Med La Habana*; v.18, p.147-169; 1881.

FIOCRUZ. Agência Fiocruz de notícias. Uma breve história da febre Amarela. 2008. Disponível em: <https://agencia.fiocruz.br/uma-breve-hist%C3%B3ria-da-febre-amarela>. Acesso em: 24 de agosto 2021.

FRANCO, O., História da Febre Amarela no Brasil. Rio de Janeiro: Divisão de Cooperação e Divulgação, Departamento Nacional de Endemias Rurais, Ministério da Saúde, 1969.

FRIEDMAN, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*. *Elsevier*, v. 38, n. 4, p. 367–378, 2002.

GEMAN, S.; BIENENSTOCK E.; DOURSAT, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*. 4, p. 1-58. 1992.

GHALEHKHONDABI, I.; ARDJMAND, E.; YOUNG, W.N.; *et al.* Water demand forecasting: review of soft computing methods. *Env Monit and Asses*. v.189, n.7, p.313, 2017.

GIOVANETTI M.; DE MENDONÇA M. C. L.; FONSECA V.; *et al.* Yellow Fever Virus Reemergence and Spread in Southeast Brazil, 2016-2019. *J Virol*. v.94, n.1, p.1-12, 2019.

GOES DE JESUS J, GRÄF T, GIOVANETTI M, *et al.* Yellow fever transmission in non-human primates, Bahia, Northeastern Brazil. *PLoS Negl Trop Dis*.v.14, n.8, e:0008405, 2020.

GOMES, A. C.; TORRES M.A.N.; DE PAULA, M.B.; *et al.* Ecologia de *Haemagogus* e *Sabethes* (Diptera: *Cilicidae*) em áreas enzooticas do vírus da febre amarela, Rio Grande do Sul, Brasil, *Epidemiol. Serv. Saúde*, v.19, n.2, p.101-113, 2010.

GONÇALVES, V. R. G. *Febre amarela (FA) no Distrito Federal : estudo descritivo dos casos confirmados de febre amarela no DF, 2008 a 2019*. 2021.Trabalho de conclusão de curso (especialização) — Universidade de Brasília, Faculdade de Ciências da Saúde, Brasília.

GÓNZALEZ-BANDALA, D. A.; CUEVAS-TELLO, J. C.; NOYOLA, D. E., *et al.* Computational Forecasting Methodology for Acute Respiratory Infectious Disease Dynamics. *Int J Environ Res Public Health*. 2020.v.17, n.12:4540, 2020.

GOO, T.; APIO, C.; HEO, G.; *et al.* Forecasting of the COVID-19 pandemic situation of Korea. *Genomics Inform*.v.19, n.1:e11, 2021.

GOVERNO DO ESTADO DE MINAS GERAIS. Conteúdo. Conheça Minas. 2021. Disponível em: <<https://www.mg.gov.br/conteudo/conheca-minas/geografia/clima-vegetacao-e-relevo>>. Acesso em: 16 setembro 2021.

GRATIVOL, A. D.; RUIZ-MIRANDA, C. R.; TEIXEIRA, A. M. G.; SCHMIDT, M. A. R. Abordagem de metapopulação para conservação dos micos-leões-dourados na paisagem fragmentada da bacia do rio São João, RJ. *Conservação do mico-leão-dourado: enfrentando os desafios de uma paisagem fragmentada*, p. 136-159, 2008.

GRYNSZPAN, D. Febre Amarela e saúde: questões entre sociedade e biodiversidade. *Fundação Oswaldo Cruz*. 2017. Disponível em: <http://www.ioc.fiocruz.br/abcnaciencia/html/word/?page_id=154>. Acessado em: 20 agos. 2020.

GUAN, G.; ENGELHARDT B. *Proceedings of the 4th machine learning for Healthcare Conference*, PMLR v.106:p.271-287, 2019.

GUMAEI, A.; AL-RAKHAMI, M.; AL RAHHAL, M. M. et al. Prediction of COVID-19 Confirmed Cases Using Gradient Boosting Regression Method. *Computers, Materials & Continua*. Vol. 66, n.1,p.315-329, 2021.

HAMLET A.; JEAN, K.; PEREA, W. *et al.* The seasonal influence of climate and environment on yellow fever transmission across Africa. *PLoS Negl Trop Dis*.v.12, n.3, e0006284, 2008.

HAMRICK P. N.; ALDIGHERI, S.; MACHADO, G., *et al.* Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. *PLoS Neglected Tropical Diseases*. v.11, n.9, e0005897, 2017.

HARRELL, FE JR. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. *Springer-Verlag*, New York, 2001.

HASTIE T.; TIBSHIRANI R.; FRIEDMAN J. The elements of statistical learning: data mining, inference, and prediction. New York: *Springer*; 2008.

HILL S. C.; DE SOUZA R.; THÉZÉ J.; *et al.* Genomic Surveillance of Yellow Fever VirusEpizootic in São Paulo, Brazil, 2016 - 2018. *PLoS Pathog*. v. 16, n. 8, e1008699, 2020.

IBGE. Instituto Brasileiro de Geografia e Estatística. IBGE cidades@. 2021. Disponível em: <<https://cidades.ibge.gov.br/brasil/mg/belo-horizonte/panorama>>. Acesso em: 15 setembro 2021.

ICTV. International Committee on Taxonomy of Viruses. 2023. Available: <<https://ictv.global/taxonomy/>>. Accessed: 23 jan 2023.

IEC. INSTITUTO EVANDRO CHAGAS . Instituto Evandro Chagas detecta vírus da Febre Amarela em mosquito *Aedes albopictus* no Brasil [Internet]. Inst. Evandro Chagas. 2018. Disponível em: <https://antigo.iec.gov.br/descoberta/instituto-evandro-chagas-detecta-virus-da-febre-amarela-em-aedes-albopictus2/>. Acesso em: 12 mar 2021.

INMET. INSTITUTO NACIONAL DE METEOROLOGIA. Glossário. Disponível em: <https://portal.inmet.gov.br/glossario/glossario>. Acessado em: 9 nov 2021.

JAMES, G.; WITTEN, D.; HASTIE, T. *et al.* An introduction to statistical learning: with applications in R. *Springer*. New York, 2013. 426p.

JONES, K.E. et al. Global trends in emerging infectious diseases. *Nature*, v. 451, p. 990-993, 2008.

KAUL R. B. et al. Spatio-temporal spillover risk of yellow fever in Brazil. *Parasit Vectors*. v.11,n.488, p. 1-12, 2018.

KAUR, H.; KUMARI, V. Predictive modelling and analytics for diabetes using a machine learning approach, *Appl. Comp. and Informatics*, v. 18 N. 1/2, p. 90-100. 2022.

KEESING, F., BELDEN, L., DASZAK, P. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468, 647–652, 2010 <https://doi.org/10.1038/nature09575>.

KLAUCKE, D. N.; BUEHLER, J. W.; THACKER, S. B. *et al.* Center for Diseases Control and Prevention. Guidelines for evaluating surveillance systems: recommendations from the guidelines working group. *Morb Mortal Wkly Rep*. v.37, n.S-5, p.1-18. 1988

KOU, D.; ZHANG, Y.; ZHENG, H. A parameters selection method of svm. In: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on. p. 1–4, 2010

KURSA, M. B.; RUDNICKI, W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software*, v.36, n.11, p.1–13. 2010.

LANTZ, B. Machine learning with R. Packt Publishing Ltd, 2013.

LEONG, L. K.; ABDULLAH, A. A. Prediction of Alzheimer's Disease (AD) Using Machine Learning Techniques with Boruta Algorithm as Feature Selection Method. *J. Phys. Conf. Ser.* 1372:012065, 2019;.

LIMA, R. O tráfico de animais silvestres. In: Diagnóstico do tráfico de animais silvestres na mata Atlântica. Corredores Central e Serra do mar. Renctas. Brasília, 2007.

LINDENBACH, B. D.; MURRAY, C. L.; THIEL, H. J.; RICE C. M. Flaviviridae. D.M. Knipe, P.M. Howley (Eds.), *Fields Virology*, Lippincott Williams & Wilkins, Philadelphia, pp. 712-746, 2013.

LINDENBACH, B. D.; MURRAY, C. L.; THIEL, H. J.; RICE C. M. Flaviviridae. D.M. Knipe, P.M. Howley (Eds.), *Fields Virology*, Lippincott Williams & Wilkins, Philadelphia, pp. 712-746, 2013.

LITTLE, R.J.; RUBIN, D.B. Missing Data. *International Encyclopedia of the Social and Behavioral Sciences*, v. 15, p. 602-607, 2015.

LIU, B. C.; BINAYKIA, A.; CHANG PC *et al.* Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang. *Plos one* v.12, n.7 :e0179763, 2017.

LOBÃO L. M.; RODRIGUES, B. S DE S. L. Mudanças ambientais de origem antrópica e sua relação com o adoecimento humano. *Saúde dinâmica – Revista Científica Eletrônica* 1ª Edição 2019 | Ano I, nº 1, 2019.

LÖWY I. *Vírus, mosquitos e modernidade: a febre amarela no Brasil entre ciência e política.* Rio de Janeiro: Fiocruz; 2006.

LÖWY, I.: Representação e intervenção em saúde pública: vírus, mosquitos e especialistas da Fundação Rockefeller no Brasil. *História, Ciências, Saúde Manguinhos*, V(3): 647-77, 1999.

LUDBERG, S. M.; LEE S. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, p.4768–4777, 2017.

MARES-GUIA M. A. M. D. M.; HORTA M. A.; ROMANO A. *et al.* Yellow fever epizootics in non-human primates, Southeast and Northeast Brazil (2017 and 2018). *Parasites and Vectors*. v.13, n.90, p.1-8, 2020.

MASCHERETTI, M; TEGAN, C. H.; SATO, H. K. *et al.* Febre amarela silvestre: reemergência de transmissão no estado de São Paulo, Brasil, 2009. *Revista de Saúde Pública*. v. 47, n.5, p.881-889, 2013

McCARTHY, M. A century of the US army yellow fever research. v.357 (9270) *Lancet*, 2001.

MINAS GERAIS. Secretaria de Estado de Saúde. DELIBERAÇÃO CIB-SUS/MG Nº 1.979, DE 18 DE NOVEMBRO DE 2014. Ajuste do PDR. Lista de macrorregiões de saúde. Disponível em: <<https://www.saude.mg.gov.br/images/documentos/Del%201979%20-%20PDR%202014.pdf>>. Acesso em 01 maio 2019.

MINAS GERAIS. Secretaria do Estado de Saúde. Boletim epidemiológico – 19/02/2019 Febre amarela Silvestre em Minas Gerais Período sazonal: Julho/2018 a junho/2019. 2019. Disponível em:

<https://www.saude.mg.gov.br/images/noticias_e_eventos/000_2019/jane_fev_mar/Febre_Amarela/Boletim_atualiza%C3%A7%C3%A3o_FA_12-02-2019.pdf>. Acesso em: 26 ago. 2021.

MINAS GERAIS. Secretaria do Estado de Saúde. Nota informativa SES/SUBVS-SVE-DVAT-CEVARB 1877/2021. 2021. Disponível em: <[file:///D:/DOWNLOADS/SEI_1320.01.0038355_2021_82%20NI%20SES%201877-2021%20-%20ALERTA%20FA%20\(1\)%20\(1\).pdf](file:///D:/DOWNLOADS/SEI_1320.01.0038355_2021_82%20NI%20SES%201877-2021%20-%20ALERTA%20FA%20(1)%20(1).pdf)>. Acesso em: 20 fev. 2023.

MINAS GERAIS. Secretaria de Estado de Saúde. CEVARB/DVAT/SVEAST – Sub VPS. Atualização epidemiológica - Febre Amarela. 2022. Disponível em: <https://www.saude.mg.gov.br/images/ppt-apresenta%C3%A7%C3%A3o_periodo_eleitoral%20An%C3%A1lise%20Febre%20Amarela%2028-07-2022.pdf>. Acesso em: 3 ago 2022.

MINISTÉRIO DA SAÚDE. Saúde de A a Z. Febre amarela. Ano: 2020. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/f/febre-amarela>> Acesso em: 26 ago. 2021.

MIR, D., DELATORRE, E., BONALDO, M. *et al.* Phylodynamics of Yellow Fever Virus in the Americas: new insights into the origin of the 2017 Brazilian outbreak. *Sci. Rep.* v.7, n.1:7385. 2017.

MONATH T. P.; VASCONCELOS P. F. C. Yellow Fever. *Jour. of Clin. Virol.* v.64: p.160–173, 2015.

MORAES, A. M., C. R.; RUIZ-MIRANDA, P. M. GALETTI, B. B. *et al.* Landscape resistance influences effective dispersal of endangered golden lion tamarins within the Atlantic Forest. *Biol. Cons.*, v.224, p. 178–187, 2018.

MORENO E. S.; ROCCO I. M.; BERGO E. S. *et al.* Yellow Fever Working Group. Reemergence of yellow fever: detection of transmission in the State of São Paulo, Brazil, 2008. *Ver. Soc. Bras. Med. Trop.* 2011, v.44, n.3, p.290-296., 2011.

MORIASI, D. N.; ARNOLD, J. G.; LIEW, M. W. VAN; BINGER, R. L.; HARMEL, R. D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, v.50, p.885-900, 2007.

MOUSSALLEM, T. M.; GAVA, C.; ARDISSON, K. S. *et al.* Yellow fever outbreak in a rural urban mixed community of Espírito Santo, Brazil: epidemiological aspects. *Rev Panam Salud Publica*; v.43:e29, . 2019

MOUTINHO, F. F B.; CORREA, D. M S; MARCANTH, N. *et al.* Vigilância de epizootias em primatas não humanos em Niterói, RJ, Brasil. *Hygeia – Rev. Bras. de Geo. Méd. e da Saú*, v. 16, p. 37–48, 2020.

MUSSUMECI, E.; COELHO, F. Machine-learning forecasting for Dengue epidemics - Comparing LSTM, Random Forest and Lasso regression. *Medrxiv*, 2020.

NAVA A.; SHIMABUKURO J. S.; CHMURA A. A.; LUZ S. L. B. The Impact of Global Environmental Changes on Infectious Disease Emergence with a Focus on Risks for Brazil. *ILAR J.* v.58, n.3, p.393-400, 2017.

NILSSON, R.; PENA J.; BJORKEGREN, J.; TEGNER, J. “Consistent Feature Selection for Pattern Recognition in Polynomial Time.” *The Jour. of mach. Lear. Res*, v.8, n. 612, 2007.

OBARA, Marcos Takashi et al. Infecção natural de *Haemagogus janthinomys* e *Haemagogus leucocelaenus* pelo vírus da febre amarela no Distrito Federal, Brasil, 2007-2008. Instituto Evandro Chagas *Epid. e Serv. de Saú.*, v. 21, n. 3, p.457-463, 2012.

OLIVEIRA, S. V.; LOPES, J. T. S.; BAIRROS, S. Vigilância da Febre Amarela em Caçapava do Sul, Rio Grande do Sul, Brasil: um relato da epizootia 2008-2009. *Rev. Pan-Ama. de Saú.*, v. 1, n.1, p. 181-186, 2010.

OPAS. ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. Alerta Epidemiológica Fiebre amarilla. 2017. Disponível em: <http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&Itemid=270&gid=37642&lang=es>. Acessado em: 24 agos 2021.

PAAIJMANS K. P, HEINIG R. L, SELIGA R. A, BLANFORD JI, *et al.* Temperature variation makes ectotherms more sensitive to climate change. *Global Chang Biol.* V.19: p.2373-80, 2013.

PANDEY, M.; KARTHIKEYAN. Performance Analysis of Time Series Forecasting Using Machine learning Algorithms for Prediction of Ebola Casualties. In: Deka, G., Kaiwartya, O., Vashisth, P., Rathee, P. (eds) Applications of Computing and Communication Technologies. ICACCT 2018. Communications in Computer and Information Science, vol 899. Springer, Singapore. 2018.

PARKS, J. M. Cluster analysis applied to multivariate geologic problems. *Jour. of Geo.*, Chicago, v.74, n.5, p.703 715, 1966.

PARMEZAN, A. R. S. *Predição de séries temporais por similaridade*. 219p. (Dissertação de mestrado). Instituto de Ciências Matemáticas e de computação- USP. Universidade de São Paulo. São Carlos, 2016.

PATHAK, M. (2018). Feature Selection in R with the Boruta R Package. Disponível em: Acesso em 6 de dezembro de 2020.

PIERCE D.A, “Seasonal adjustment when both deterministic and stochastic seasonality are present”, In Seasonal Analysis of Economic Time Series. Arnold Zellner Ed. 242-269. Washington, D.C., U.S. Dept. of Commerce, Bureau of the Census. 1979.

POSSAS, C.; LOURENÇO-DE-OLIVEIRA, R.; TAUIL, P.L.; et al. Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. *Mem. do Instit. Oswaldo Cruz*, v. 113, n. 10, p. 1-12, 2018.

PRIETO JT, SCOTT K, MCEWEN D, *et al.* The Detection of Opioid Misuse and Heroin Use From Paramedic Response Documentation: machine learning for Improved Surveillance. *J Med Inter. Res.* 2020 v.22, n.1:e15645.

RASCHKA, S.; MIRJALILI, V. Python machine learning. Third edition. Packt. Birmingham, UK. 2019. 741p.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria, 2022. URL: <https://www.R-project.org/>.

REZENDE I. M.; SACCHETTO L.; MUNHOZ DE MELLO É. *et al.* Persistence of Yellow fever virus outside the Amazon Basin, causing epidemics in Southeast Brazil, from 2016 to 2018. *PLoS Negl Trop Dis.* v.12, n.6, e:0006538, 2018.

RIBEIRO. M. C.; METZGER, J. P.; MARTENSEN, A. C. *et al.* The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? *Implic. for conser.. Bio. Conser.* v.142, n.6, p. 1141 – 1153, 2009.

ROCCO, I. M; KATZ, G; TUBAKI, R. M. Febre amarela silvestre no Estado de São Paulo, Brasil: casos humanos autóctones . *Rev. Inst. Adolfo Lutz* . v. 62, n.3, p. 201-206, 2003.

ROIG C.; MIRET J.; ROJAS A. *et al.* Estudio de Fiebre Amarilla en primates en áreas de brote de los departamentos de San Pedro y Central del Paraguay. *Mem. Del Inst Investig. en Ciencias la Salud.*v.7, p.40–5, 2009.

ROMANO, A. P. M.; RAMOS D. G.; ARAÚJO F. A. A.; *et al.* Febre amarela no Brasil: recomendações para a vigilância, prevenção e controle. *Epidemiol. Serv. Saúde*, Brasília, v.20, n.1, p.101-106, 2011.

ROMANO, A.P.M.; COSTA, Z.G.A.; RAMOS, D.G.*et al.* (2014) Yellow Fever Outbreaks in Unvaccinated Populations, Brazil, 2008–2009. *PLoS Negl Trop Dis* v.8, n.3. e2740, 2014.

ROUQUAYROL, M. Z. Epidemiologia e saúde. 4.ed. Rio de Janeiro: MEDSI, 1994.

ROUQUAYROL, M. Z. Rouquayrol: epidemiologia & saúde. - 8. ed. - Rio de Janeiro: Medbook, 2018. 752 p..

ROYAL SOCIETY. (GB). Machine learning: the power and promise of computers that learn by example. Rep. DES4702, *R. Soc. G. B.*, Londres. 2017

RUIZ-CÁRDENAS, R; KRAINSKI, E. T. Preenchimento de falhas em bancos de dados meteorológicos diários: uma comparação de abordagens. XVII Congresso Brasileiro de Agrometeorologia – 18 a 21 de Julho de 2011 – SESC Centro de Turismo de Guarapari, Guarapari - ES. 2011.

SAAD, L. C.; BARATA, R. B. Surtos de febre amarela no estado de São Paulo, 2000-2010. *Epidemiol. Serv. Saúde*. v.25, n.3, pp.531-540. 2016.

SACCHETTO L.; SILVA N. I. O.; REZENDE I. M. D. *et al.* Neighbor danger: Yellow fever virus epizootics in urban and urban-rural transition areas of Minas Gerais state, during 2017-2018 yellow fever outbreaks in Brazil. *PLoS .Negl .Trop. Dis.* v.14, n.10, e8658, 2020.

SAMPAIO, I.B.M. Estatística aplicada à experimentação animal. Belo Horizonte: FEPMVZ, 1998. 221p. STATISTICAL ANALYSES SYSTEM – SAS, 1998.

SANDERSON, M.; BULLOCH, A. G.; WANG, J.; *et al.* Predicting death by suicide using administrative health care system data: Can recurrent neural network, one-dimensional convolutional neural network, and gradient boosted trees models improve prediction performance? *J Affect Disord.* v.264, p.107-114, 2020.

SANTOS, H. G. dos. *Comparação da performance de algoritmos de machine learning para análise preditiva em saúde pública e medicina*. 206p. 2018. (Tese de doutorado). Universidade de São Paulo. São Paulo.

SÃO PAULO. Prefeitura de São Paulo. Secretaria Municipal de Saúde. Alerta febre amarela. Ano 2017. Disponível em: <https://sbim.org.br/images/files/alertan001-ccd.pdf>. Acesso em 01 ago. 2022.

SCHAPIRE, R. E. The strength of weak learnability. [S.l.]: Machine learning. 197-227p, 1990.

SCHETTINO, S.; GUIMARÃES, N. V.; SILVA, D. L. *et al.* Relação entre a ocorrência de acidentes de trabalho e a baixa escolaridade dos trabalhadores no setor florestal. *Braz. J. of Develop.* Curitiba, v. 6, n.4, p.22567-22589, 2020.

SHAPLEY, L. A Value for N-Person Games. *Contributions to the Theory of Games*, pages 307–317, Boris Sharchilev, Yury Ustinovski. 1953.

SHIH, H.; RAJENDRAN, S. Comparison of Time Series Methods and machine learning Algorithms for Forecasting Taiwan Blood Services Foundation's Blood Supply. *Jour. of Health. Eng.*. p.1-6. 2019.

SILVA N. I. O.; SACCHETTO L.; DE REZENDE I. M.; *et al.* (Silva *et al*, 2020a). Recent sylvatic yellow fever virus transmission in Brazil: the news from an old disease. *Viol J.* v.17, n.9, p. 1-12, 2020.

SILVA C. F.; CARDOSO S C. C.; MENGUETE P. S. (Silva *et al*, 2020b). Distribuição das epizootias em primatas não humanos por febre amarela no estado do Rio de Janeiro nos anos ciclos de 2017 e 2018, segundo família. *Ações de saúde e geração de conhecimento nas ciências médicas* Ponta Grossa, PR: Atena, 2020.

SIMMONDS P.; BECHER P.; BUKH J. *et al.* ICTV.Virus Taxonomy Profile: *Flaviviridae*. *Jour. Of Gen.viol.* 98, p.1-2, 2017.

SIQUEIRA, P. C.; NOIA, M. E. L; CATÃO, R. de C. *et al.* Completude das fichas de notificação de febre amarela no estado do Espírito Santo, 2017. *Epidemiol. Serv. Saúde.* v..29, n.3, e2019402, 2020

SOPER F, PENNA E, SERAFIM J, FROBISHER M, PINHEIRO J. Yellow fever whithout *Aedes aegypti*. Study of a rural epidemic in the Valle do Chanaan, Espirito Santo, Brazil, 1932. *Am J Hyg.* v.18, n.3, p. 555-87, 1933.

SOUZA R. P. *Filogeografia da Febre Amarela na América do Sul.* 134p. Tese (Doutorado) Universidade Federal de São Paulo. São Paulo. 2013.

SOUZA R. P.; FOSTER P. G.; SALLUM M. A. *et al* Detection of a new yellow fever virus lineage within the South American genotype I in Brazil. *J Med Virol.* v.82, n.1 p.175-85. 2010

SOUZA T. DOS S.; COELHO, E. G. A.; OLIVEIRA, S. R. T. *et al.* Ocorrência de febre amarela no Brasil: uma revisão integrativa da literatura (2014-2018). *Revi. Eletr. Acer. Saúde*, v.28, 2019.

TAKATA, R. Novas técnicas contra uma antiga ameaça. *Cienc. Cult.* .v..69, n.2, p.09-11, 2017.

TAPAK, L.; HAMIDI, O.; FATHIAN, M. *et al.* Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC Res Notes*. v.12, n.1, 2019

TAUIL PL. Aspectos críticos do controle da febre amarela no Brasil. *Rev Saude Publica*. v. 44, n. 3, p. 555-588, 2010.

TEIXEIRA L. Da transmissão hídrica a culicidiana: a febre amarela na sociedade de medicina e cirurgia de São Paulo. *Rev Bras Hist*. v.21, n.41, p.217-242. 2001;

VALADARES G.; CAMPOS K.; RODRIGUES L.; *et al.* Febre amarela: análise estatística no período de janeiro a junho de 2017 nos municípios de Caratinga, Piedade de Caratinga, Imbé de Minas, Ubaporanga e Santa Bárbara do Leste -Minas Gerais. *Bras. Jour. Surg. Cli.Rese.*. v. 20, n.3 p.171-181, 2017.

VAN BUUREN, S.; GROOTHUIS-OUDSHOORN, K. MICE: Multivariate Imputation by Chained Equations in R. *Jour. of Statis. Software*, v. 45, n. 3, p. 1–68, dez. 2011.

VASCONCELOS P. F. C.; TRAVASSOS DA ROSA A. P. A.; RODRIGUES S. G. *et al.* Yellow fever in Pará State, Amazon Region of Brazil, 1998-1999: entomologic and epidemiologic findings. *Emerging Infectious Diseases* v.7, p.565-569, 2001.

VASCONCELOS P. F. C. Febre amarela. *Rev. Soc. Bras. Med. Trop*. v.36, n.2. 2003.

VASCONCELOS P. F. C.; SPERB A. F.; MONTEIRO H. A. O.; *et al.* Isolations of yellow fever vírus from *Haemagogus leucocelaenus* in Rio Grande do Sul State, Brazil. *Trans R Soc Trop Med Hyg*, v.97, n.1. p.60-2, 2003.

VASCONCELOS P. F. C.; BRYANT J. E.; TRAVASSOS DA ROSA, A. P. A. Genetic Divergence and Dispersal of Yellow Fever Virus, *Brazil. Emerg. Inf. Disea*. v.10, n.9, p.1578-1584, 2004.

VASCONCELOS P. F. C. (Vasconcelos, 2010a) Febre Amarela no Brasil: reflexões e hipóteses sobre emergência em áreas previamente livres. *Rev. Saúde Pública.* v.44, n. 6, p.1144-9, 2010.

VASCONCELOS P. F. C. (Vasconcelos, 2010b) Yellow fever emergence in Brazil. *Public Health Rev.* 2010.

VELÁSQUEZ, R. M. G. *Seleção de Variáveis de Entrada para Previsores Neurais de Carga.* 2006, 125 p. UFRJ, D.Sc., Engenharia Elétrica, 2006 (Tese doutorado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.

VILLELA, P. M. *Impactos ambientais da modernização agropecuária em Goiás.* 2016. 99 f. Dissertação (Mestrado em Ciências Ambientais) - Universidade Federal de Goiás, Goiânia.

VOLPATO, G. et al. Baby pangolins on my plate: possible lessons to learn from the COVID-19 pandemic. *Jour. of Ethnobiology and Ethno.*, v. 2020, p.16-19, 2020.

VOPHAM, T.; HART, J. E.; LADEN, F. *et al.* Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environ Health.* v.17, n.40, 2018.

WICKHAM H, GROLEMUND G. *R for data science.* Sebastopol, Canadá: O'Reilly Media, Inc. 2016.

WIEMKEN T. L.; KELLEY R. R. machine learning in Epidemiology and Health Outcomes Research. *Annu Rev Public Health.* v.41, p.21-36. 2020.

WILCOVE, D. S., MCLELLAN, C. H., DOBSON, A. P. Habitat fragmentation in the temperate zone. In: *Conservation biology. The science of scarcity and diversity* (ed. M. E. Soulé). Sinauer, Sunderland, MA, 1986.

WILLMOTT, C. J. On the validation of models. *Phys. Geog.*, v.2, n. 2, p. 184-194, 1981.

WHO. WORLD HEALTH ORGANIZATION. Regional office for Africa. Yellow fever. Factsheet. Transmission. 2021. Disponível em: <https://www.afro.who.int/health-topics/yellow-fever>. Acesso em: 24 de agosto de 2021.

ZAMBRANO-BIGIARINI, M. *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series.* doi: 10.5281/zenodo.839854, R package

version 0.4-0, 2020. Disponível em: <<https://github.com/hzambran/hydroGOF>>. Acesso em: 10 maio 2021.

ZHANG, L.; ZOU, Z.; SHAN, W. Development of a method for comprehensive water quality forecasting and its application in Miyun reservoir of Beijing, China. *Jour. of Envir. Sciences*, v.56, p.240-246, 2017.

10. ANEXOS

10.1 Lista de municípios das macrorregiões de saúde centro e nordeste de acordo com Plano Diretor de regionalização da saúde de Minas Gerais 2014

Código	Municípios	Extensão territorial (KM²)	População Estimada (IBGE/TCU 2014)
Macro:	CENTRO (BELO HORIZONTE/SETE LAGOAS)	58.120	6.480.169
Micro:	BELO HORIZONTE/ NOVA LIMA/ CAETÉ	4428,4	3.358.067
310620	Belo Horizonte	331	2491109
310640	Belo Vale	365,4	7803
311000	Caeté	541,1	43395
313460	Jaboticatubas	1113,8	18785
314230	Moeda	154,2	4903
314480	Nova Lima	428,4	88672
313660	Nova União	171,5	5766
315390	Raposos	71,9	16144
315460	Ribeirão das Neves	154,2	319310
315480	Rio Acima	230,1	9816
315670	Sabará	303,6	133528
315780	Santa Luzia	233,8	214830
316830	Taquaraçu de Minas	329,4	4006
Micro:	BETIM	3667,5	681.860
310670	Betim	345,9	412003
310810	Bonfim	301,2	7014
310900	Brumadinho	640,2	37314

312060	Crucilândia	166,5	4991
312410	Esmeraldas	909,6	66237
312600	Florestal	194,4	7137
313010	Igarapé	109,9	39045
313665	Juatuba	96,8	24662
314015	Mário Campos	35,2	14427
314070	Mateus Leme	302,6	29873
315040	Piedade dos Gerais	260,6	4897
315530	Rio Manso	232,1	5636
316292	São Joaquim de Bicas	72,5	28624
Micro:	CONTAGEM	329,5	844.678
311860	Contagem	194,6	643476
312980	Ibirité	73	171932
316553	Sarzedo	61,9	29270
Micro:	CURVELO	15633,3	183.638
310480	Augusto de Lima	1250,7	5051
310920	Buenópolis	1611	10586
311910	Corinto	2524,5	24457
312090	Curvelo	3295,9	78373
312570	Felixlândia	1553,4	14973
313110	Inimutaba	529,1	7349
314250	Monjolos	652,1	2365
314360	Morro da Garça	414	2649
315320	Presidente Juscelino	696,6	3901
316060	Santo Hipólito	430,8	3261

Código	Municípios	Extensão territorial (KM²)	População Estimada (IBGE/TCU 2014)
316935	Três Marias	2675,2	30673
Micro:	GUANHÃES	5357,3	117.198
311380	Carmésia	258,5	2584
311750	Conceição do Mato Dentro	1671,5	18235
312260	Dom Joaquim	406,7	4624
312310	Dores de Guanhões	381,8	5335
312800	Guanhões	1076	33297
314060	Materlândia	282,3	4659
315600	Rio Vermelho	987,2	13674
315680	Sabinópolis	919,4	16014
316610	Senhora do Porto	382,3	3598
316650	Serra Azul de Minas	222,7	4358
317180	Virginópolis	440,4	10820
Micro:	ITABIRA	5868,1	231.571
310540	Barão de Cocais	340,7	30893
310770	Bom Jesus do Amparo	195,5	5871
311535	Catas Altas	240,2	5184
312590	Ferros	1090,2	10707
313170	Itabira	1256,5	116745
313280	Itambé do Mato Dentro	381,1	2261
314370	Morro do Pilar	476,5	3399
314750	Passabém	94,5	1764
315720	Santa Bárbara	684,2	29888
315800	Santa Maria de Itabira	509,7	10942
316050	Santo Antônio do Rio Abaixo	107,2	1818

316190	São Gonçalo do Rio Abaixo	364,6	10488
316480	São Sebastião do Rio Preto	127,2	1611
Micro:	JOÃO MONLEVADE	1688,4	138.797
310600	Bela Vista de Minas	108,6	10362
313620	João Monlevade	99,3	78040
314470	Nova Era	363,2	18000
315570	Rio Piracicaba	370,4	14590
316100	São Domingos do Prata	746,9	17805
Micro:	OURO PRETO	2981,4	181.136
313190	Itabirito	543	49203
314000	Mariana	1193,3	58233
314610	Ouro Preto	1245,1	73700
Micro:	SETE LAGOAS	14880,9	436.612
310020	Abaeté	1816,9	23494
310320	Araçai	185,8	2344
310500	Baldim	554	8082
310700	Biquinhas	457,2	2652
310960	Cachoeira da Prata	61,2	3727
310990	Caetanópolis	156,2	11046
311250	Capim Branco	94,1	9461
311560	Cedro do Abaeté	279,9	1222
311890	Cordisburgo	823,2	8981
312640	Fortuna de Minas	198,1	2872

312720	Funilândia	201,7	4153
313100	Inhaúma	244,3	6114
313570	Jequitibá	446	5310

313970	Maravilhas	260,4	7674
314350	Morada Nova de Minas	2084,6	8712
314640	Paineiras	637,8	4692
314690	Papagaios	552,8	15144
314740	Paraopeba	625,1	23940
314960	Pequi	204,1	4314
315200	Pompéu	2557,7	30943
315360	Prudente de Moraes	125,8	10287
315370	Quartel Geral	555,5	3493
315850	Santana de Pirapama	1221	8068
316720	Sete Lagoas	537,5	229887
Micro:	VESPASIANO	3284,9	306.612
311787	Confins	42	6409
313760	Lagoa Santa	232	58702
314110	Matozinhos	252,9	36382
314930	Pedro Leopoldo	291	62473
315900	Santana do Riacho	676,8	4235
316295	São José da Lapa	48,6	21905
317120	Vespasiano	70,1	116506

Macro	ORDESTE (TEÓFILO)	57347,5	931.946
Micro:	ÁGUAS FORMOSAS	4896	60.173
310090	Águas Formosas	817,7	19248
310660	Bertópolis	425,6	4656
312015	Crisólita	969,8	6467
312705	Fronteira dos Vales	317,8	4765

313890	Machacalis	329,8	7210
314850	Pavão	599,4	8755
315765	Santa Helena de Minas	276,9	6336
317030	Umburatiba	404,9	2736
Micro:	ALMENARA	15452,2	182.523
310170	Almenara	2300,8	41028
310520	Bandeira	484,7	5039
312245	Divisópolis	566,1	10027
312560	Felisburgo	594	7288
313470	Jacinto	1390,5	12524
313580	Jequitinhonha	3517,5	25260
313600	Joáima	1667,7	15523
313650	Jordânia	549,2	10786
314055	Mata Verde	230,2	8363
314675	Palmópolis	436,5	6537
315510	Rio do Prado	479,4	5324
315660	Rubim	968,7	10304
315710	Salto da Divisa	943,6	7096
315810	Santa Maria do Salto	442,1	5400
316030	Santo Antônio do Jacinto	497,4	12024
Micro:	ARAÇUAÍ	5257	91.183
310340	Araçuaí	2235,7	37220
310650	Berilo	586,8	12469
311950	Coronel Murta	813,9	9394
312650	Francisco Badaró	463,8	10546
313545	Jenipapo de Minas	284,9	7531

317160	Virgem da Lapa	871,9	14023
Micro:	ITAOBIM	5632,4	82.851

311700	Comercinho	656,6	7962
313330	Itaobim	679,9	21566
313400	Itinga	1640,7	15012
314140	Medina	1440	21485
314315	Monte Formoso	383,8	4877
315217	Ponto dos Volantes	1215,2	11949
Micro:	NANUQUE	4929,5	70.686
311370	Carlos Chagas	3198,9	20090
314430	Nanuque	1515,4	41852
316670	Serra dos Aimorés	215,2	8744
Micro:	PADRE PARAÍSO	2686	62.527
311300	Carai	1240,2	23458
311545	Catuji	421	6721
313230	Itaipé	480,9	12489
314630	Padre Paraíso	543,9	19859
Micro:	PEDRA AZUL	3668,6	53.796
310100	Águas Vermelhas	1257,6	13378
310270	Cachoeira de Pajeú	673,8	9372
312235	Divisa Alegre	118,5	6398
314870	Pedra Azul	1618,7	24648
Micro:	TEÓFILO OTONI / MALACACHETA / ITAMBACURI	14.825,8	328.207
310285	Angelândia	184,9	8416
310470	Ataléia	1838,4	14188

311080	Campanário	441,4	3720
312675	Franciscópolis	715,9	5784
312680	Frei Gaspar	628,3	6031
313270	Itambacuri	1418,6	23557
313700	Ladainha	865,2	17881
313920	Malacacheta	719,3	19209
314490	Nova Mógica	377	3811
314530	Novo Cruzeiro	1700,6	31760
314535	Novo Oriente de Minas	754,1	10773
314620	Ouro Verde de Minas	175	6138
315000	Pescador	317,6	4284
315240	Poté	632,7	16428
316330	São José do Divino	326,2	3943
316555	Setubinha	535,7	11717
316860	Teófilo Otoni	3242,8	140567

10.2 Ficha de notificação de investigação de Epizootia

República Federativa do Brasil Ministério da Saúde		SINAN SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO		Nº	
FICHA DE NOTIFICAÇÃO/ INVESTIGAÇÃO EPIZOOTIA					
Definição do caso: Animal ou grupo de animais encontrados doentes e/ou mortos, incluindo ossadas, sem causa definida, que podem preceder a ocorrência de doenças em humanos					
Dados Gerais	1	Tipo de Notificação		2- Individual	
	2	Agravado/doença		EPIZOOTIA	
	3	Data da Notificação			
	4	UF	5	Município de Notificação	Código (IBGE)
Dados de Ocorrência	6	Unidade de Saúde (ou outra fonte notificadora)		Código	
	7	Data do início da epizootia			
	8	Fonte da informação		9 (DDD) Telefone da fonte da informação	
	10	UF	11	Município de Ocorrência	Código (IBGE)
	12	Distrito			
	13	Bairro		14 Logradouro (rua, avenida, ...)	Código
	15	Número	16	Complemento (apto., casa, ...)	17 Geocampo 1
	18	Geocampo 2		19 Ponto de Referência	20 CEP
	21	(DDD) Telefone		22 Zona	23 Ambiente
			1 - Urbana 2 - Rural 3 - Periurbana 9 - Ignorado		1 - Domicílio 2 - Parque, praça ou zoológico 3 - Área silvestre 4 - Reserva ecológica 5 - Outro
	24	Houve coleta de material para exame laboratorial		1-Sim 2-Não 9-Ignorado	25 Se houve coleta, informar a data
	26	Se houve coleta, qual material		1-Sim 2-Não 9-Ignorado	
		<input type="checkbox"/> fígado <input type="checkbox"/> rim <input type="checkbox"/> baço <input type="checkbox"/> cérebro <input type="checkbox"/> coração <input type="checkbox"/> fezes <input type="checkbox"/> soro <input type="checkbox"/> sangue total			
		<input type="checkbox"/> outro material Qual _____			
27	Animais acometidos		<input type="checkbox"/> Doentes _____ <input type="checkbox"/> Mortos _____		
		1-Ave 3-Canino 5-Felino 7-Primata não humano 9-Outros. 2-Bovídeo 4-Equídeo 6-Morcego 8-Canídeo selvagem Especificar _____		<input type="checkbox"/> Doentes _____ <input type="checkbox"/> Mortos _____	
28	Suspeita diagnóstica		4-Encefalite Espongiforme Bovina	<input type="checkbox"/> 1ª suspeita diagnóstica	
		1-Raiva 5-Febre Amarela 2-Encefalite Equina 6-Influenza Aviária 3-Febre do Vírus do Nilo Ocidental 7-Outro. Especificar: _____		<input type="checkbox"/> 2ª suspeita diagnóstica <input type="checkbox"/> 3ª suspeita diagnóstica	
29	Resultado laboratorial		1-Positivo 2-Negativo 3-Inconclusivo 9-Ignorado		
		<input type="checkbox"/> Raiva <input type="checkbox"/> Encefalite espongiforme bovina <input type="checkbox"/> Outro Especificar _____			
		<input type="checkbox"/> Encefalite equina <input type="checkbox"/> Febre amarela			
		<input type="checkbox"/> Febre do Nilo <input type="checkbox"/> Influenza aviária			
Observações:					
Investigador	Município/Unidade de Saúde		Código da Unid. de Saúde		
	Nome		Função	Assinatura	
	Sinan NET		SVS 21/08/2008		

10.3 Ficha de investigação de febre amarela

República Federativa do Brasil Ministério da Saúde		SINAN		Nº				
SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO								
FICHA DE INVESTIGAÇÃO DE FEBRE AMARELA								
CASO SUSPEITO: Indivíduo com quadro febril aguda (até 7 dias), de início súbito, acompanhado de icterícia e/ou manifestações hemorrágicas, residente ou procedente de área de risco para febre amarela ou de locais com ocorrência de epizootias em primatas não humanos ou isolamento de vírus em vetores, nos últimos 15 dias, não vacinados contra febre amarela ou com estado vacinal ignorado.								
Dados Gerais	1	Tipo de Notificação		2 - Individual				
	2	Agravado/doença		Código (CID10)	3 Data da Notificação			
	FEBRE AMARELA		A 9 5.9					
	4 UF	5	Município de Notificação	Código (IBGE)				
6	Unidade de Saúde (ou outra fonte notificadora)		Código	7 Data dos Primeiros Sintomas				
Notificação Individual	8		Nome do Paciente		9	Data de Nascimento		
	10 (ou) Idade	1 - Hora 2 - Dia 3 - Mês 4 - Ano	11	Sexo M - Masculino F - Feminino 1 - 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 a 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	
Dados de Residência	17 UF	18		Município de Residência	Código (IBGE)	19	Distrito	
	20		Bairro		21		Logradouro (rua, avenida,...)	
	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)		31		Data da Investigação	
	32		Ocupação		33		Informar os dados da investigação entomológica (mosquitos) e de epizootias	
Antecedentes Epidemiológicos	1 - Sim 2 - Não 9 - Ignorado					<input type="checkbox"/>	Ocorrência de Epizootias (Mortandade de macacos: conhecidos como guariba, bugio, saguis, micos, macaco aranha, macaco prego, guigó, soim, etc.)	
						<input type="checkbox"/>	Isolamento de vírus em mosquitos	
						<input type="checkbox"/>	Presença de mosquito <i>Aedes aegypti</i> em área urbana (Observar período de viremia do paciente)	
Antecedentes Clínicos	34		Vacinado Contra Febre Amarela		1 - Sim 2 - Não 9 - Ignorado	<input type="checkbox"/>	35	Caso Afirmativo. Data
	36		UF		37		Município	
	38		Unidade de Saúde		39		Sinais e Sintomas 1 - Sim 2 - Não 9 - Ignorado	
Atendimento	40		Ocorreu Hospitalização?		1 - Sim 2 - Não 9 - Ignorado	<input type="checkbox"/>	41	Data da Internação
	42		UF		43		Município	
	44		Unidade de Saúde		45		Exames Inespecíficos (anotar o maior valor encontrado, independente da data de coleta)	
Dados do Laboratório	45		Bilirrubina Total		mg/dl		AST (TGO)	UI
	45		Bilirrubina Direta		mg/dl		ALT (TGP)	UI

