

**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
DEPARTAMENTO DE BIOLOGIA
CIÊNCIAS BIOLÓGICAS-BACHARELADO**

DULCE MARIA TRINDADE VILANOVA

**ALERTA VERMELHO:
SUBSIDIANDO ESTRATEGICAMENTE A PREVENÇÃO
DE DENGUE COM IA EM SERGIPE**

**São Cristóvão
2025**

DULCE MARIA TRINDADE VILANOVA

**ALERTA VERMELHO:
SUBSIDIANDO ESTRATEGICAMENTE A PREVENÇÃO
DE DENGUE COM IA EM SERGIPE**

Trabalho de Conclusão de Curso II apresentado ao Departamento da Biologia requisito para obtenção do grau de bacharel em Ciências Biológicas.

Orientador: Prof. Pablo Ariel Martinez

**São Cristóvão
2025**

RESUMO

A dengue, uma arbovirose transmitida pelo mosquito *Aedes aegypti*, causa sérios problemas de saúde pública, exigindo controle rigoroso dos surtos. Métodos tradicionais de amostragem enfrentam limitações devido à falta de uma abordagem multifocal e dificuldades de acesso. Todavia, dados de satélites diários fornecem informações climáticas e de uso do solo para qualquer região. No Brasil, há notificação compulsória dos casos de dengue no Sistema de Informação de Agravos de Notificação (SINAN). Dessa forma, a combinação desses dados com algoritmos de Machine Learning (ML) permite identificar padrões e fazer previsões de forma autônoma e eficiente. Este trabalho objetivou construir modelos preditivos de dengue no estado de Sergipe por meio de ML, e averiguar a importância dos fatores envolvidos na incidência de dengue estadual por meio de uma análise com Shapley Additive Explanations. Inicialmente, foram gerados modelos dinâmicos do *A. aegypti* para todos os municípios sergipanos, de 2016 a 2022, a partir do software MaxEnt, fornecendo a adequabilidade climática (AC) mensal do mosquito. Para a criação dos modelos preditivos, a AC foi utilizada, juntamente de dados mensais epidemiológicos do SINAN, dados mensais climáticos do TerraClimate, dados anuais de uso de solo do Map Biomas, referentes aos anos de 2016 a 2022, e variáveis socioeconômicas pelo IBGE. Destes dados, foram selecionadas sete variáveis preditoras (AC, temperatura mínima-TM, precipitação-PP, escoamento superficial-ES, área florestal-AF, de pastagem-AP, Produto Interno Bruto-PIB e esgotamento sanitário-ES). A partir dessas variáveis, foram treinados (80% dos dados) e testado (20% dos dados) os algoritmos de Extreme Gradient Boosting Classifier-XGB, Random Forest Classifier-RF, Máquina de Suporte de Vetores-SVM e Deep Learning- DL para prever a incidência de dengue em três categorias (Baixa, Média ou Alta). Os algoritmos foram calibrados nos treinos com os melhores hiperparâmetros (100 iterações, cv=5) e avaliados nos testes com base em matrizes de confusão, e nas métricas: acurácia, precisão, recall e F1. O melhor modelo foi submetido a análise Shapley para avaliar a importância de cada variável na incidência. O modelo XGB teve o melhor desempenho, acertando 73,4% para previsões de alta incidência e métricas acima de 0,5, podendo ser utilizado como ferramenta para o monitoramento da dengue em Sergipe. Com a análise Shapley obtivemos que as variáveis que tiveram maior impacto sobre a incidência em todos municípios sergipanos foram: PIB, AF e a PP. O PIB apresentou relação inversa com a incidência de dengue, valores mais baixos contribuem positivamente nos modelos, o que geralmente está atrelado a infraestrutura urbana deficiente. Municípios com menores AF apresentam influência positiva maior sobre a incidência. Os distúrbios nas AF podem motivar a migração de vetores para as áreas urbanas. A precipitação possui relação positiva com a incidência devido ao aumento de umidade e o maior desenvolvimento de criadouros, o que pode expandir a população do *Aedes* e o seu contato com a população, aumentando a incidência. Este projeto enfatiza como o uso de ML poderá contribuir no controle, análise e prevenção da arbovirose em Sergipe a partir da criação de um sistema de monitoramento contínuo espacial.

Palavras-chave: *Aedes aegypti*; dengue; extreme gradient boosting classifier; machine learning; Shapley.

SUMÁRIO

| | |
|--|----|
| 1. INTRODUÇÃO | 5 |
| 1.1 Objetivo Geral..... | 7 |
| 1.1.2 Objetivos Específicos | 7 |
| 2. METODOLOGIA | 8 |
| 2.1 Modelos dinâmicos de nicho para <i>Aedes aegypti</i> | 8 |
| 2.2 Coleta de dados | 9 |
| 2.3 Construção dos modelos preditivos | 10 |
| 2.4 Análise de SHAP | 11 |
| 3. RESULTADOS | 12 |
| 4. DISCUSSÃO..... | 17 |
| 5. CONCLUSÃO | 21 |
| 6. REFERÊNCIAS | 22 |
| Apêndice A..... | 29 |
| Apêndice B | 30 |
| Apêndice C | 31 |
| Apêndice E | 33 |
| Apêndice F..... | 34 |
| Apêndice G..... | 35 |

1. INTRODUÇÃO

Historicamente, desde os primeiros assentamentos humanos, a interação entre o homem e o meio ambiente tem sido complexa e problemática. Com o advento da revolução industrial, a revolução verde e a rápida expansão urbana, os seres humanos ampliaram sua ocupação em novos ecossistemas, muitas vezes de maneira desordenada e sem planejamento adequado (Amuasi *et al.*, 2020; Mackenzie; Jeggo, 2019). Isso intensificou o contato com novos animais e patógenos, seja pela invasão de habitats naturais ou pelo desenvolvimento de criadouros em áreas urbanas e rurais (Cabrera *et al.*, 2022; Silva, Jesiel Souza; Mariano; Scopel, 2008). Como consequência, surgiram doenças antes desconhecidas, ressaltando a necessidade de uma prática mais integrada para controlar e preveni-las. Nesse sentido o conceito "One Health" surgiu como uma abordagem essencial para enfrentar os desafios de saúde no século XXI, promovendo a integração entre a saúde humana, animal e ambiental em um sistema unificado (Mackenzie; Jeggo, 2019). Esta perspectiva holística e transdisciplinar busca soluções conjuntas e inovadoras para problemas de saúde e ambientais, destacando a importância da conectividade entre setores (Adisasmito *et al.*, 2023; Cabrera *et al.*, 2022; De Castañeda *et al.*, 2023; Elnaiem *et al.*, 2023). Esta abordagem tem sido incorporada em estratégias da Organização Mundial da Saúde (OMS, 2024; FaO; Who; Woah, 2024) para combater as Doenças Tropicais Negligenciadas (DTN) (World Health Organization, 2020; Mulakoli *et al.*, 2023).

Em 2016, a OMS elencou 20 enfermidades como Doenças Tropicais Negligenciadas, as quais são conhecidas por sua distribuição vinculada à zona tropical do planeta (Ministério da Saúde, 2024; Valverde, 2013; Rosa Meurer; Soares Coimbra, 2022). Dentre as principais DTNs, está a dengue, uma arbovirose endêmica em mais de 100 países tropicais, transmitida principalmente pelo mosquito *Aedes aegypti*, quando infectado por um dos quatro tipos do flavivírus (DENV-1, DENV-2, DENV-3, DENV-4) (Halasa; Shepard; Zeng, 2012). Esta enfermidade, mundialmente, atinge cerca de 400 milhões de pessoas por ano, resultando em até 22 mil mortes, o que sobrecarrega os sistemas de saúde (Bhatt *et al.*, 2013; Gubler, 2012; Roy; Bhattacharjee, 2021). Os sintomas da dengue podem variar bastante entre formas mais brandas, como febre alta, irritações na pele, dor atrás dos olhos, cefaleia e mialgia; e mais graves, como vômitos e hemorragias, os quais são capazes de levar à morte do indivíduo

(Whitmire *et al.*, 1987; OMS, 2024). Seus efeitos impactam a vida de milhares de pessoas, gerando um grande ônus à sociedade.

Além dos gastos para o tratamento dos enfermos, cerca de 8,9 bilhões de dólares anualmente em todo o planeta (Paz-Bailey *et al.*, 2024), muito também é investido no combate e prevenção dessa arbovirose, pelo controle e monitoramento do vetor. Atualmente, os métodos mais empregados no monitoramento do *A. aegypti* são as análises das presenças de larvas e pupas, e a utilização de ovitrampas (Vieira; Rocha; Justo, 2023), que permitem conhecer os locais com maior infestação do mosquito, auxiliando na tomada de medidas para controle da situação. Contudo, essas amostragens apenas determinam que uma residência possui foco de dengue, mesmo havendo vários recipientes infestados (multifocal), minimiza a gravidade da situação e a eficiência da amostragem. (Paz-Bailey *et al.*, 2024). Ademais, há uma grande dificuldade em acessar alguns espaços geográficos, gerando um empecilho para a obtenção dos dados (Tauil, 2002). Deste modo, a aplicação de novos métodos para monitoramento da dengue sob a ótica One Health, que utilizem a correlação de fatores climáticos e globais quanto a sua influência direta e indireta no ciclo das arboviroses, poderia contribuir na compreensão da complexa dinâmica dessa enfermidade.

Atualmente há uma grande geração de dados diários, principalmente proveniente de satélites (De Castro; De Oliveira, 2021; Junior *et al.*, 2022). Esses dados são de fácil acesso e podem ser aplicados na saúde pública para compreender melhor a dinâmica das enfermidades e formular melhores estratégias para lidar com a situação (Chen, Yuyang *et al.*, 2024). Os dados obtidos a partir do sensoriamento remoto podem ser utilizados para alimentar e treinar algoritmos de Machine Learning (ML), sendo possível identificar padrões e fazer previsões de forma autônoma e dinâmica. O sensoriamento, que leva em conta as variáveis atreladas à sazonalidade da incidência da arbovirose, pode ser uma importante ferramenta para auxiliar na tomada de decisões governamentais, facilitando o monitoramento e otimizando gastos a partir da utilização de métodos inovadores como modelos de distribuição de espécies (Chen, Yuyang *et al.*, 2024).

Dentre os países afetados pela dengue, o Brasil destaca-se no ranking dos países tropicais com maior número de casos (OMS, 2024). Segundo o Ministério da Saúde, até a semana epidemiológica 4 de 2024, o Brasil já contabilizava 260 mil casos de dengue. A presença de características climáticas e ambientais favoráveis à proliferação do seu principal vetor, o *A. aegypti*, e os desafios para o estabelecimento de políticas públicas efetivas para o controle,

contribuem para a sua alta incidência. Por meio do Decreto 78.231, de 12 de Agosto de 1976, tornou-se compulsória a notificação das suspeitas e casos confirmados da doença (SINAN, 2019), são incorporados ao Sistema de Informações de Agravos de Notificação (SINAN) e disponibilizados no DATASUS. Esta plataforma abriga diversos dados de saúde pública brasileira. A disponibilidade desses e sua constante atualização permite que seja possível o desenvolvimento de novos estudos e metodologias para controle e monitoramento da dengue e de outras doenças, ao longo do tempo, de maneira eficiente e autônoma, podendo inclusive ser usada para alimentar e treinar algoritmos de ML.

No Brasil, o estado de Sergipe tem o terceiro maior coeficiente de letalidade de dengue do país, segundo o boletim epidemiológico da semana 1 a 24 de 2024 do Ministério da Saúde. O estado possui uma pequena extensão territorial e apresenta autonomia administrativa municipal bastante evidente, frente a problemáticas de saúde pública (SERGIPE, 1996). Assim, todos os municípios possuem um bom sistema de notificação desta doença, mesmo que entre eles existam características heterogêneas. Isso facilita a identificação e o discernimento acerca dos diversos fatores que afetam a incidência da arbovirose nesse estado, tornando Sergipe um bom modelo de análise.

1.1 Objetivo Geral

Construir modelos preditivos mensais da incidência de dengue no estado de Sergipe e avaliar a importância dos fatores envolvidos na incidência de dengue a nível estadual e municipal.

1.1.2 Objetivos Específicos

- Construir modelos preditivos mensais da incidência de dengue no estado de Sergipe por meio do ML utilizando variáveis ambientais, de uso de solo e socioeconômicas, testando algoritmos diferentes.
- Avaliar o desempenho de diferentes algoritmos nos modelos preditivos.
- Observar quais os fatores que mais impactam na incidência de dengue no estado e em diferentes municípios, por meio da análise Shapley Additive exPlanations

2. METODOLOGIA

2.1 Modelos dinâmicos de nicho para *Aedes aegypti*

O vetor de transmissão da dengue, *Aedes aegypti*, possui diferentes estágios de desenvolvimento ao longo do ano, os quais variam com os meses. Levar em conta essas variações para estudar *A. aegypti* pode auxiliar no delineamento de estratégias de mitigação mais acuradas. Nesse sentido, os modelos dinâmicos de nicho permitem integrar dados de ocorrência e variáveis ambientais, estabelecendo o tempo como aspecto chave em suas predições, ajudando a entender sobre a variação da distribuição da espécie ao longo dos meses (Dobson *et al.*, 2023). Esta pode ser uma forma eficiente de prever o risco de transmissão da dengue, por tornar possível identificar áreas e momentos com maior risco de transmissão de arboviroses.

Para a construção de um modelo dinâmico de *Aedes aegypti* foi utilizada a abordagem de calibração multitemporal proposta por Eduardo *et al.* (2022). Para isto, foram obtidos dados de ocorrência da espécie, do período de 2016 a 2022, a partir do Global Biodiversity Information Facility (GBIF) (<https://www.gbif.org/>). Esses dados foram organizados e limpos, corrigindo sinonímias, nomenclaturas desatualizadas, erros espaciais, duplicatas e dados ausentes. Cada registro incluía informações sobre o ano e o mês de coleta, permitindo construir modelos para predições mensais. Para evitar a autocorrelação espacial nos modelos, utilizamos o pacote "spThin" para rarefação (Aiello-Lammens *et al.*, 2015), mantendo apenas um ponto a cada 50 km². Após a triagem, restaram 1.387 pontos de ocorrência de *A. aegypti*, distribuídos em todo o Brasil

Foram utilizados dados climáticos mensais provenientes do TerraClimate (<https://www.climatologylab.org/terraclimate.html>), que foram submetidos a um teste de correlação de Pearson (Pearson; Henrici, F.R.S., 1896), mantendo, nos casos que duas ou mais variáveis apresentaram correlação de $r > |0.7|$, apenas a variável com maior importância biológica para *A. aegypti*, para evitar a sobreparametrização dos modelos. Dessa forma, seis variáveis foram selecionadas: (i) precipitação; (ii) temperatura máxima; (iii) escoamento superficial; (iv) velocidade do vento; (v) pressão de vapor e (vi) radiação solar.

O modelo de distribuição dinâmica de *A. aegypti* foi construído utilizando o software MaxEnt (Phillips; Anderson; Schapire, 2006). Os parâmetros padrão do MaxEnt

foram empregados para modelar mensalmente o nicho ecológico de *A. aegypti*. Para isto, foram gerados 1.000 pontos de pseudo-ausências por mês ao longo de um ano. Ao final, construímos uma base de 12.000 pontos de pseudo-ausências, gerados para caracterizar a variação climática ao longo do ano na região de estudo. Foram extraídas as variáveis explicativas espaciotemporalmente dinâmicas para os pontos de ocorrência e pontos de pseudo-ausências. Pela aplicação padrão do MaxEnt, foram usadas diferentes combinações de classes de “features”, a regularização foi mantida em 1, para prevenir o sobreajuste dos modelos. A validação cruzada k-fold foi realizada com 10 partições (Jung; Hu, 2015). Foram então geradas as projeções espaço-temporalmente dinâmicas. Uma vez gerados os mapas mensais de adequabilidade climática para *A. aegypti* para o Brasil (Apêndice A), foram extraídos os valores de adequabilidade climática média para cada município de Sergipe.

2.2 Coleta de dados

Foram utilizadas 16 variáveis preditoras do período de 2016 a 2022, que incluíram variáveis climáticas mensais, variáveis anuais de uso de solo, variáveis socioeconômicas e a adequabilidade climática mensal do *A. aegypti*. As variáveis climáticas mensais utilizadas foram a precipitação (1), a temperatura máxima (2), a temperatura mínima (3), a pressão de vapor (4), a evapotranspiração potencial (5), o escoamento superficial (6), a velocidade do vento (7) e a incidência da radiação solar (8), obtidas do TerraClimate (<https://www.climatologylab.org/terraclimate.html>). Para as variáveis anuais de uso de solo, foi calculada a proporção das áreas de floresta nativa (9), de manguezal (10), de pastagem (11), destinadas à aquicultura (12) e das áreas urbanas (13), todas a nível municipal, obtidas na plataforma MapBiomas (<https://plataforma.brasil.mapbiomas.org/>). Já as variáveis socioeconômicas foram obtidas a partir do Instituto Brasileiro de Geografia e Estatística (IBGE) (<https://www.ibge.gov.br/>), sendo elas o Produto Interno Bruto anual (PIB) (14), em que o ano de 2022 foi calculado a partir da variação média dos últimos 6 anos somando com o valor do último ano e a Taxa de esgotamento sanitário (15). E por fim a adequabilidade climática mensal do *A. aegypti* (16), aferida neste trabalho (ver em 2.1 Modelos dinâmicos de nicho para *Aedes aegypti*).

Para estimar a variável resposta, incidência média mensal de dengue a cada 100.000 habitantes por município, foi feito um cálculo de proporcionalidade a partir de dados mensais de ocorrência de dengue para todos os municípios do estado de Sergipe, no período de 2016 a 2022, coletados do SINAN

(<http://tabnet.datasus.gov.br/cgi/defthtm.exe?sinannet/cnv/denguebse.def>), e de dados populacionais dos municípios de Sergipe obtidos do IBGE (<https://www.ibge.gov.br/>). Para evitar problemas com excessos de 0 nos modelos os dados foram filtrados, mantendo somente os meses com incidência > 0 . Posteriormente os valores de incidência foram classificados em três categorias, de acordo com os quartis: (i) baixa incidência (primeiro quartil; abaixo de 5 casos a cada 100mil habitantes); (ii) média incidência (até o terceiro quartil; de 5 até 26 casos a cada 100mil habitantes) e (iii) alta incidência (acima do terceiro quartil; a partir de 27 casos a cada 100mil habitantes).

2.3 Construção dos modelos preditivos

As variáveis preditoras foram submetidas a dois testes de seleção. Primeiramente, foi realizado um teste de correlação de Pearson (Pearson; Henrici, F.R.S., 1896), no qual, quando duas ou mais variáveis apresentaram $r > |0.7|$, foi selecionada aquela com maior importância epidemiológica a fim de evitar a sobreparametrização dos modelos. Após esta análise, foi realizada a segunda seleção de variáveis utilizando o algoritmo de Random Forest, com a biblioteca “sklearn” do Python (Pedregosa *et al.*, 2011), selecionando todas as preditoras com $\geq 5\%$ de importância para explicação da variável resposta. Ao fim desses testes, foi selecionado um total de sete variáveis explicativas para a incidência de dengue no estado de Sergipe, a saber: (i) precipitação; (ii) escoamento superficial; (iii) temperatura mínima; (iv) área de floresta; (v) área de pastagem; (vi) adequabilidade climática do *A. aegypti*; (vii) PIB; e (viii) taxa de esgotamento sanitário.

Para as análises de ML, foram aplicados 4 algoritmos de classificação no Python: eXtreme Gradient Boosting Classifier (XGBoost), com a biblioteca “XGBoost” (Chen, Tianqi; Guestrin, 2016); Random Forest Classifier (RandomForestClassifier) com a biblioteca “Sklearn” (Pedregosa *et al.*, 2011), Deep Learning (KerasClassifier), com a biblioteca “TensorFlow” (Abadi *et al.*, 2015) e Máquina de Suporte de Vetores (SVM), com o processo gaussiano “rbf” e a biblioteca “Sklearn” (Pedregosa *et al.*, 2011). Os modelos utilizaram as variáveis explicativas para prever valores de incidência de forma categórica em Baixa incidência, Média incidência ou Alta incidência. Inicialmente, os dados foram divididos em um conjunto de treino (80% dos dados) e um conjunto de teste (20% dos dados) a partir de uma amostragem estratificada. No conjunto de treino, os algoritmos foram submetidos a uma busca randomizada para encontrar os melhores hiperparâmetros, usando 100 iterações e 5 blocos de validação cruzada. Os modelos com os melhores hiperparâmetros

foram então avaliados no treino e validados no teste, com base nas matrizes de confusão, e nas métricas de acurácia, precisão, recall e F1. Todas as análises foram realizadas no Python, utilizando a plataforma Google Colab (<https://colab.google>). Por fim, com o pacote “dplyr” (Wickham et al., 2023) do R (R Core Team, 2016) foram gerados mapas preditivos para o mês de maior incidência dos anos de 2019, 2020, 2021 e 2022, a partir da combinação do shapfile dos municípios de Sergipe (IBGE, 2023) com as previsões do modelo de maior desempenho, gerados no Python.

2.4 Análise de SHAP

Em análises preditivas de Machine Learning, apesar de ser possível avaliar a qualidade das previsões a partir de métricas de desempenho, muitas vezes, a interpretação da lógica interna dos modelos torna-se complexa (Azodi; Tang; Shiu, 2020; Petch *et al.*, 2022; Watson *et al.*, 2019). Diante da dificuldade de explicar suas previsões, os modelos de ML tornam-se “caixas-pretas”, prejudicando a tomada de decisões acerca do problema em questão (Ekanayake; Meddage; Rathnayake, 2022; Štrumbelj; Kononenko, 2014; Watson *et al.*, 2019). Nesse contexto, o método de SHapley Additive exPlanations (SHAP) surge como uma forma de desvendar o funcionamento interno de qualquer algoritmo de ML, ao computar a importância situacional de variáveis a partir de um valor de Shapley, o que permite a visualização de quanto cada variável contribui em valor e em sentido (positiva ou negativamente) para a previsão (Štrumbelj; Kononenko, 2014). Nos últimos anos, o método SHAP tem sido usado em previsões de qualidade do ar (Vega García; Aznarte, 2020), em autoencoders (Antwarg *et al.*, 2021), na engenharia civil (Ekanayake; Meddage; Rathnayake, 2022) e na saúde (Nohara *et al.*, 2022), demonstrando um grande potencial explicativo aplicável à epidemiologia.

Neste trabalho, foi executada uma análise de SHAP sobre as previsões gerais do Estado de Sergipe, feitas com o melhor modelo, calibrado com os melhores hiperparâmetros. Além disso, foram analisados separadamente os oito territórios de estado (Agreste Central Sergipano, Alto Sertão Sergipano, Baixo São Francisco Sergipano, Centro Sul Sergipano, Grande Aracaju, Leste Sergipano, Médio Sertão Sergipano e Sul Sergipano). Em cada território, foi selecionado o município que apresentou o maior valor de incidência mensal, e foi aplicado o SHAP a essa previsão para compreender os principais fatores que estão associados com os altos níveis de incidência nos municípios.

3. RESULTADOS

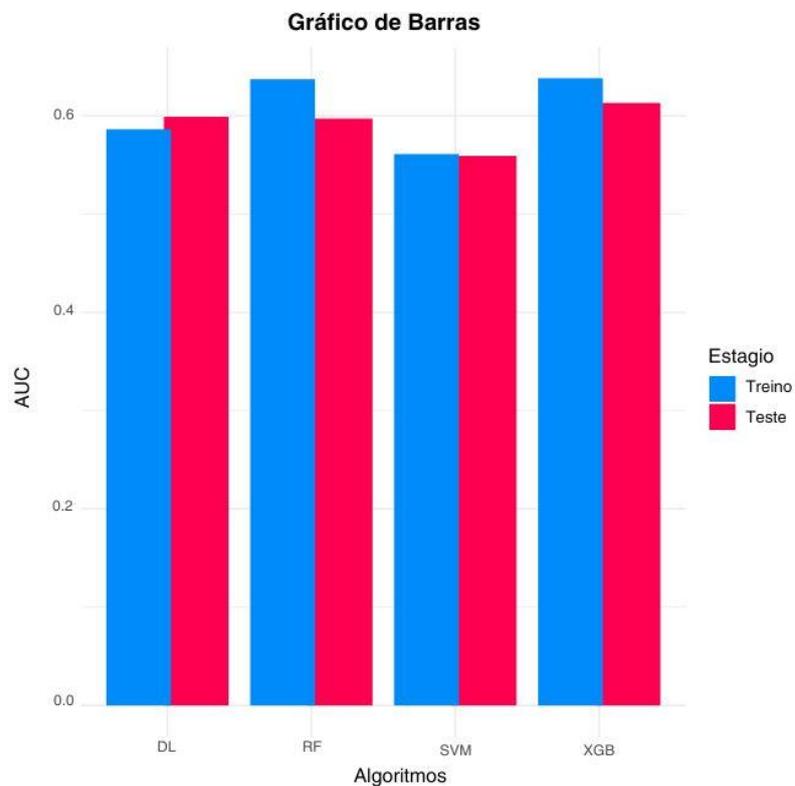


Figura 1: Gráfico de Barras feito a partir da acurácia dos modelos na fase treino e teste utilizando os algoritmos: Deep Learning (KerasClassifier), Randon Forest Classifier, Máquina de Suporte de Vetores (Gaussiano) e eXtreme Gradient Boosting Classifier

| Métricas | Treino | Teste |
|--------------|--------|-------|
| Precision | 0.644 | 0.621 |
| Recall | 0.619 | 0.584 |
| F1-score | 0.625 | 0.597 |
| Accuracy | 0.638 | 0.613 |
| Macro avg | 0.629 | 0.601 |
| Weighted avg | 0.639 | 0.613 |

Tabela 1- Métricas de desempenho no conjunto de treino e teste do algoritmo eXtreme Gradient Boosting Classifier.

Nossos modelos preditivos apresentaram desempenhos satisfatórios, com todas as métricas acima de 0,5 (Apêndice D). Os melhores modelos foram gerados pelo eXtreme Gradient Boosting Classifier, possuindo as melhores métricas, com a acurácia a cima de 0.6 no treino e no teste (Figura 1, Tabela 1, Apêndice D)

A matriz gerada na fase treino (Apêndice E) com 80% dos dados conseguiu classificar corretamente 947 predições de 1484. Para cenários com baixa incidência (<5), nosso modelo preditivo classificou corretamente 64,6% dos dados (247/382), errando com maior frequência a classificação para casos medianos (127) do que para altos (8). Para os cenários medianos (5 a 27), houve um acerto de 70,0% das classificações (510/728), sendo a categoria de incidência com maior número de dados e de acertos. Na predição para cenários mais altos (>27), houve 50,8% de acertos de classificação (190/374), sendo observada uma tendência de subestimação dessa categoria, sendo a maioria dos erros classificando esses dados como média incidência (156). O modelo acertou a maioria dos casos, sendo 58,8% dos resultados previstos como “baixo” de fato meses de baixa incidência (247/420); 64,3% dos resultados previstos como “médio”, meses de média incidência (510/793); e 70,1% dos previstos como “alto” (190/271).

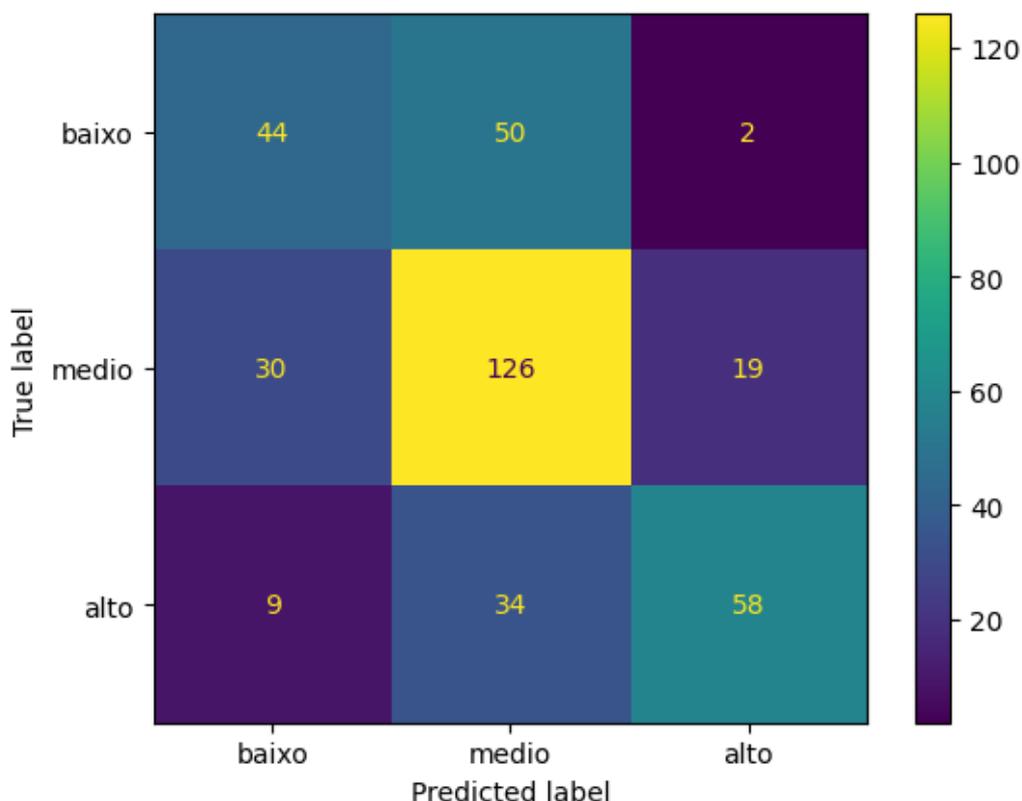


Figura 2- Matriz de confusão gerada na fase de teste a partir do eXtreme Gradient Boosting Classifier.

Na matriz de teste (Figura 2), gerada com 20% dos nossos dados, obtivemos 45,8% (44/96) de acerto na classificação de cenários baixos (>5), com maior erro devido à superestimação de 50 cenários como médios e 2 como alto; 72% (126/175) de acerto em previsões medianas (5 a 27), com erros maiores devido à subestimação de 30 cenários; e 57,4% (58/101) de acerto para os cenários altos (>27), havendo uma subestimação de 34 cenários para médios e 9 para baixos. Apesar de uma menor taxa de acerto nas classificações do teste, as qualidades de previsão dentro de cada categoria se mantiveram satisfatórias, de modo que, em o modelo acertou 53% (44/83) das suas previsões de baixa incidência; 60% (126/210) das previsões de média incidência; e 73,4% (58/79) das previsões de alta incidência.

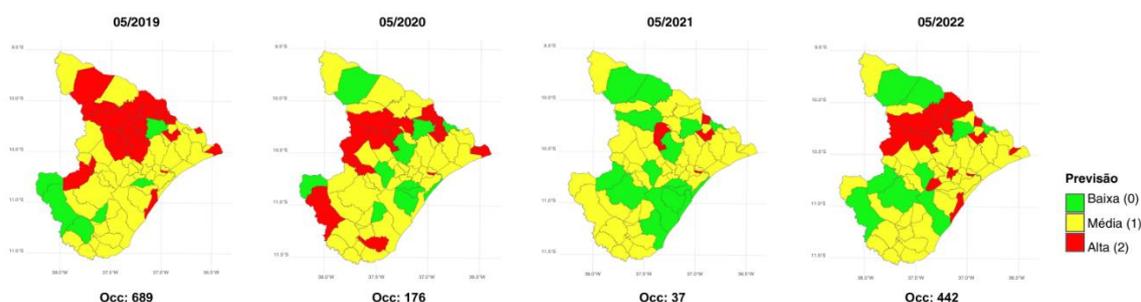


Figura 3: Mapas de previsões do estado de Sergipe com as previsões do melhor modelo (XGBoost) para os meses de maio para os anos de 2019, 2020, 2021 e 2022. Os municípios em vermelhos são aqueles que possuem uma previsão de alta incidência de dengue, em amarelos são os que apresentam média incidência de casos e em verde aqueles que possuem previsões de baixa incidência.

Mapas de previsão relacionando o número de municípios em alerta com a quantidade real de ocorrências nos meses de maio dos últimos quatro anos (Figura 3). O ano de 2019 apresentou uma alta ocorrência de casos de dengue dentro do estado, com 689 casos totais para o mês de maio, nosso modelo gerou 24 previsões de alta incidência, 45 previsões de média incidência e 6 prevendo baixa incidência. Em 2020 tiveram 176 ocorrência totais de casos, prevendo 21 municípios com alta incidência, 43 com média incidência e 11 com baixa incidência. No ano de 2021 foram notificadas 37 ocorrências, e o modelo gerou 7 previsões de alta incidência, 52 previsões médias e 16 de baixa incidência. Em 2022 Sergipe apresentou 442 casos em maio, 18 previsões foram altas, 48 médias, 9 baixas.

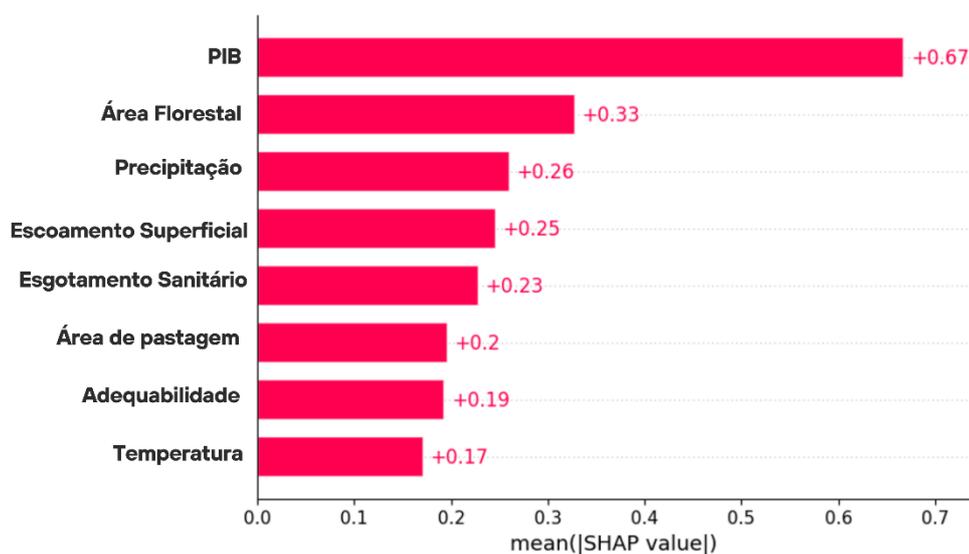


Figura 2- Gráfico *barplot* (gráfico de barras) com a sequência de variáveis na ordem de maior impacto para as de menor impacto sobre a incidência da dengue, gerado com todo o conjunto de dados do estado de Sergipe.

Quanto à importância das variáveis no *output* do modelo (Figura 4), pode-se observar que o Produto Interno Bruto (PIB) e a “Área Florestal” apresentaram maior influência sobre a incidência com valores médios próximos à 0,67 e 0,33. Seguidos pela variável de precipitação com valores médios de impacto em torno de 0,26, Escoamento Superficial com média em torno de 0,25; Esgotamento Sanitário e Área de Pastagem com médias de 0,23/0,2 e Adequabilidade com valores próximos a 0,19; e, por último Temperatura Mínima do mosquito com média próxima a 0,17.



Figura 5- Gráfico *waterfall* (cascata) mostrando as variáveis listadas em ordem de importância e suas contribuições para o output do modelo dos municípios: a) Graccho Cardoso (Médio Sertão Sergipano; dezembro de 2021), b) Santa Rosa de Lima (Leste Sergipano; dezembro de 2020), c) Pedra Mole (Agreste Central Sergipano; maio de 2019), d) Riachão do Dantas (Centro Sul Sergipano; dezembro de 2021), e) Santa Luzia do Itanhy (Sul Sergipano; julho de 2019), f) Aracaju (Grande Aracaju; junho de 2021), g) Brejo Grande (Baixo São Francisco Sergipano; junho de 2019) e h) Nossa Senhora de Lourdes (Alto Sertão Sergipano; maio de 2019). Ao lado esquerdo dos modelos estão os valores de cada variável.

As seis variáveis que apresentaram maior influência sobre a incidência de dengue dentre os municípios selecionados (Figura 5) foram “PIB”, aparecendo em sete municípios de oito (sendo a influência positiva em seis municípios: Gracho Cardoso, Santa Rosa de Lima, Pedra Mole, Santa Luzia do Itanhy, Brejo Grande e Nossa Senhora de Lourdes, e uma influência negativa em Riachão do Dantas), “Temperatura mínima” em 6 (positiva em cinco: Santa Rosa de Lima, Pedra Mole, Riachão do Dantas, Brejo Grande e Nossa Senhora de Lourdes; e negativa em Aracaju) “Área Florestal” em cinco (positiva em quatro: Gracho Cardoso, Pedra Mole, Santa Luzia do Itanhy e Nossa Senhora de Lourdes; e influência negativa em Santa Rosa de Lima), “Adequabilidade” em 5 (positiva em: Santa Luzia do Itanhy e Aracaju; Negativa em: Gracho Cardoso, Pedra Mole e Riachão,), “Esgotamento Sanitário” em 5 (positivo em: Gracho Cardoso, Santa Luzia do Itanhy e Aracaju; Negativo em: Riachão do Dantas, Brejo Grande) e “Precipitação” influente em 5 (positiva em: Santa Rosa de Lima, Riachão do Dantas, Santa Luzia do Itanhy e Aracaju; Negativo em: Brejo Grande)

4. DISCUSSÃO

O comportamento sazonal da dengue, combinado com as mudanças causadas por atividades humanas no ambiente, destaca a importância de integrar diferentes setores para estudar sua epidemiologia. Com a disponibilidade de dados diários, como imagens de satélite e registros obrigatórios de casos, é possível aplicar algoritmos de Machine Learning para analisar esses fatores de forma eficaz. Nossos modelos preditivos que apresentaram o melhor desempenho foi eXtreme Gradient Boosting (Figura 1) com todas as métricas de qualidade acima de 0,5 no cenário de treino e no teste (Tabela 1; Figura 1 e 2). Com a aplicação do método SHAP, foi possível identificar quais as variáveis mais influentes para a incidência de dengue no estado de Sergipe e para os municípios com maior incidência dentro do cenário de cada território do estado.

Apesar dos erros de classificação nos nossos modelos visualizados nas matrizes de confusão (Apêndice E, Figura 2), podemos observar que, quando o modelo prevê uma determinada categoria (e.g. prevê que um mês terá baixa incidência), ele acerta na maioria dos casos, principalmente quando as previsões são para alta incidência (Figura 2), com acerto de 73,4%. Esse fato também pode ser visualizado nos mapas de previsão (Figura 3), em que a quantidade de previsões para a alta incidência acompanha o número de ocorrências reais dentro

do estado. Desse modo, nosso modelo consegue ser um bom indicativo para cenários mais preocupantes de dengue.

A variável mais influente na incidência da dengue no estado foi o PIB (+0,67 mean SHAP value) (Figura 4). Conforme apontado na literatura, a dengue está fortemente associada a fatores socioeconômicos. Regiões com maior densidade populacional e menor PIB tendem a apresentar uma menor infraestrutura urbana, o que, aliado a variáveis ambientais, pode favorecer a proliferação de criadouros do mosquito e aumentar a exposição da população aos vetores, contribuindo para o crescimento dos casos da doença (Cabrera *et al.*, 2022; Cao *et al.*, 2017; Delmelle *et al.*, 2016; Qi *et al.*, 2015). A Área Florestal foi a segunda variável mais influente (Figura 4). Alguns estudos também mostraram a associação das variáveis vegetacionais sobre a incidência de doenças transmitidas por vetores dípteros e esse mesmo padrão se repete para a dengue (Machault *et al.*, 2010; Qi *et al.*, 2015). Essas áreas são fontes de criadouros e podem influenciar na dinâmica populacional do vetor por apresentar ótimas condições para o seu desenvolvimento, podendo ser uma fonte de mosquitos para as proximidades quando sofrem perturbações. Em seguida, a precipitação foi a terceira variável com maior impacto (Figura 4). As taxas de precipitação e da variação da temperatura influenciam diretamente a dinâmica populacional e o ciclo de vida do *A. aegypti*, a precipitação geralmente é relacionada ao acúmulo de água e a geração de novos habitats para o desenvolvimento larval do *A. aegypti* (Li *et al.*, 2020; Mendonça, 2009). Além disso, estudos mostram que a umidade do ar provocada pelas precipitações também pode aumentar a atividade de voo do vetor (Brunkard; Cifuentes; Rothenberg, 2008; Cabrera *et al.*, 2022; Junior *et al.*, 2022; Lima; Firmino; Gomes Filho, 2008; Silva, Fabrício Drummond *et al.*, 2016; Viana; Ignotti, 2013)

Com análise mais detalhada de como as variáveis influenciam na dinâmica da dengue nos municípios (Figura 5) observamos que a influência dos valores do PIB é inverso ao da incidência, ou seja, municípios como Pedra Mole que apresentam um menor valor de PIB (42.873) (Figura 5a) mostrou a maior influência positiva sobre a incidência da dengue. O inverso também é exposto, Riachão do Dantas apresenta o maior valor de PIB (328.904) e uma influência negativa da variável sobre a incidência. Esses resultados corroboram com a literatura, que aponta que os valores mais baixos de PIB são comumente relacionados com alta incidência (Qi *et al.*, 2015; Cabrera *et al.*, 2022; Cao *et al.*, 2017; Delmelle *et al.*, 2016; Qi *et al.*, 2015). Esta relação geralmente é atribuída a infraestrutura, contudo, nossa outra variável socioeconômica, Esgotamento Sanitário, não apresentou uma relação consistente

nem intensa com a incidência; como é possível visualizar em Luzia do Itanhy e Riachão do Dantas (Figura 5e, Figura 5d), onde baixos valores de esgotamento apareceram com influência positiva e negativa na incidência. Em Sergipe, somente 12 municípios apresentaram taxas de esgotamento sanitário acima de 50%, e no estudo de Andrade (2017), foi observado que a inadequação do esgotamento sanitário no estado influenciou em um crescimento de 54,50% na incidência. Entretanto, o mesmo estudo revelou a influência de outras variáveis de infraestrutura na incidência de dengue, como a coleta de lixo total, coleta direta de lixo, abastecimento de água e abastecimento canalizado, contribuindo para um aumento de 2,77% (Andrade, 2017). Dessa forma, sugerimos que a variável de esgotamento não é suficiente para representar a complexidade da infraestrutura total urbana a nível de município, e logo seus impactos sobre a dinâmica da doença. Além disso, a variável de esgotamento foi coletada em 2010, de modo que a falta de atualização desses valores pode não representar de maneira clara a situação atual.

A Temperatura mínima apareceu entre as cinco variáveis mais influentes em 6 municípios (Figura 5). No geral de forma positiva (Figura 5b, 5c, 5d, 5g, 5h), com exceção a Aracaju (Figura 5f). Assim como nossos resultados, outros estudos mostraram uma forte influência da temperatura com relação a incidência de dengue, o que se deve principalmente à ecologia do vetor e seus processos fisiológicos, influenciando desde o tamanho do *A. aegypti* e sua necessidade ao repasto sanguíneo até o processo de incubação do vírus (Cabrera *et al.*, 2022; Carvajal *et al.*, 2018). A variação média de temperatura mínima apresentada na literatura também foi similar ao nosso estudo, variando entre 17° a 22°C (Campbell *et al.*, 2015; Gomes; Nobre; Cruz, 2012; Woongkoon *et al.*, 2013). O impacto negativo da temperatura mínima na incidência em Aracaju pode ser intensificado pelo efeito de “ilha de calor” (Carrington *et al.*, 2013; Misslin *et al.*, 2016; Seah *et al.*, 2021). Esse efeito é observado em grandes centros urbanos, com baixa Área florestal, com alto número de construções (asfaltos, prédios) e com alta densidade populacional, assim como Aracaju (Carrington *et al.*, 2013; Misslin *et al.*, 2016; Seah *et al.*, 2021). Ele provoca um aumento da temperatura local, conseqüentemente a da temperatura mínima, podendo assim ultrapassar a faixa ótima de temperatura mínima para o vetor, e influenciando de forma negativa a incidência no estado e logo no nosso modelo (Misslin *et al.*, 2016; Seah *et al.*, 2021).

Valores menores de Área florestal apresentara maior influência positiva sobre a incidência (Figura 5a, 5c, 5d) enquanto maiores áreas florestais tiveram uma influência

positiva mais baixa (Figura 5e) ou negativa (Figura 5h). As florestas contribuem para as condições climáticas locais, logo, sua diminuição pode aumentar a temperatura da região em escala microclimática provocando uma maturação mais rápida dos vetores (Afrane *et al.*, 2005; Morais, 2011). Além disso, essas áreas também podem ser refúgios para os vetores quando as condições urbanas não estiverem adequadas a depender do equilíbrio da comunidade presente eles podem permanecer ali (McMichael; World Health Organization, 2003; Schmidt; Ostfeld, 2001). Contudo, o desmatamento e a diminuição da cobertura vegetal geram distúrbios na comunidade, levando à migração de mosquitos para os centros urbanos (Galeana-Pizaña *et al.*, 2024; Vallera, 2024). Além disso, as remoções da vegetação podem criar ambientes ainda mais propícios para a proliferação do *A. aegypti* (Galeana-Pizaña *et al.*, 2024; Vallera, 2024), como encontrado por um estudo no México em que a perda de 1% de cobertura florestal a nível municipal resultou em um aumento de 16,9% nos fatores de risco de dengue (Galeana-Pizaña *et al.*, 2024).

Nos modelos a precipitação contribuiu positivamente para a incidência (Figura 5b, 5d, 5e e 5f), de modo que taxas mais elevadas (Figura 5) apresentam um maior impacto. A precipitação tem grande influência na incidência de dengue por favorecer o ciclo biológico do vetor *Aedes aegypti*, e as chuvas somadas a recipientes artificiais das cidades, geram criadouros em que as fêmeas da espécie depositam seus ovos (Gubler, 2012). A precipitação não somente gera água parada, mas também influencia a temperatura e umidade do local, que ajudam na sobrevivência do vetor, prolongando seu ciclo de vida e capacidade de transmitir o vírus DENV às populações humanas (Giesen *et al.*, 2020). Chuvas frequentes também podem dificultar ou atrasar práticas de controle do mosquito, o que explica a relação positiva da precipitação com a incidência da dengue para esses municípios. Isto, somado ao desenvolvimento urbano sem planejamento e a falta de ferramentas de prevenção e monitoramento, tem afetado negativamente o bem-estar da população e o sistema público de saúde (Secretária de Estado da Saúde de Sergipe, 2024).

Com relação à adequabilidade climática para o *A. aegypti*, o resultado deste estudo está em consonância com a literatura, que aponta que áreas mais adequadas climaticamente apresentam maior ocorrência do vetor da dengue (Scavuzzo *et al.*, 2018). No modelo dinâmico de distribuição de espécie, a umidade, a velocidade do vento, radiação e temperatura máxima foram as variáveis mais influentes (Apêndice B). As curvas de resposta indicam que maior umidade (Pressão de Evapotranspiração = 3,11 kPa), radiação elevada

(em torno de 25 MJ/m²/dia), temperaturas elevadas (Temperatura Máxima = 23°C), e menor velocidade do vento (2,67 m/s) aumentam a adequabilidade climática para o vetor da dengue (Apêndice C). A umidade relativa é crucial para a sobrevivência dos ovos e longevidade dos adultos, criando condições ideais para a proliferação do mosquito. Radiação e temperaturas altas aceleram o desenvolvimento larval e a atividade reprodutiva, enquanto ventos leves facilitam a permanência dos mosquitos em áreas específicas, aumentando a disseminação do vírus (Cruz *et al.*, 2020; Scavuzzo *et al.*, 2018).

5. CONCLUSÃO

A complexidade das problemáticas da dengue evoca a necessidade de uma abordagem holística como “One Health” a partir da transdisciplinaridade. Neste estudo, observamos a associação entre variáveis ambientais, socioeconômicas, de uso de solo e intrínsecas à ecologia do mosquito, influenciando a incidência de dengue no estado de Sergipe, que possui o terceiro maior coeficiente de letalidade dessa arbovirose no país. As duas variáveis mais importantes para o estado foi, o PIB que apresentou uma relação inversa com a incidência, que pode estar relacionada com a maior infraestrutura dos municípios com valores maiores desse índice. Além disso, observamos que as características próprias de cada município podem influenciar na epidemiologia da doença. Este trabalho enfatiza a integração e a importância de dados de sensoriamento remoto e uso de ferramentas de Machine Learning, para monitorar e compreender a dinâmica da dengue a fim de melhor orientar a tomada de decisões para o seu combate.

6. REFERÊNCIAS

ABADI, M.; BARHAM, P.; CHEN, J.; CHEN, Z.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; IRVING, G.; ISARD, M.; KUDLUR, M.; LEVENBERG, J.; MONGA, R.; MOORE, S.; MURRAY, D. G.; STEINER, B.; TUCKER, P.; VASUDEVAN, V.; WARDEN, P.; WICKE, M.; YU, Y.; ZHENG, X. TensorFlow: A system for large-scale machine learning. 2015.

AFRANE, Y. A.; LAWSON, B. W.; GITHEKO, A. K.; YAN, G. Effects of Microclimatic Changes Caused by Land Use and Land Cover on Duration of Gonotrophic Cycles of *Anopheles gambiae* (Diptera: Culicidae) in Western Kenya Highlands. **Journal of Medical Entomology**, v. 42, n. 6, p. 974–980, 1 nov. 2005. <https://doi.org/10.1093/jmedent/42.6.974>.

AIELLO-LAMMENS, M. E.; BORJA, R. A.; RADOSAVLJEVIC, A.; VILELA, B.; ANDERSON, R. P. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. **Ecography**, v. 38, n. 5, p. 541–545, maio 2015. <https://doi.org/10.1111/ecog.01132>.

AMUASI, J. H.; LUCAS, T.; HORTON, R.; WINKLER, A. S. Reconnecting for our future: The Lancet One Health Commission. **The Lancet**, v. 395, n. 10235, p. 1469–1471, maio 2020. [https://doi.org/10.1016/S0140-6736\(20\)31027-8](https://doi.org/10.1016/S0140-6736(20)31027-8).

ANDRADE, J. D. S. UNIVERSIDADE FEDERAL DE SERGIPE CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA DEPARTAMENTO ESTATÍSTICA E CIÊNCIAS ATUARIAS GRADUAÇÃO EM ESTATÍSTICA. 2017.

ANTWARG, L.; MILLER, R. M.; SHAPIRA, B.; ROKACH, L. Explaining anomalies detected by autoencoders using Shapley Additive Explanations. **Expert Systems with Applications**, v. 186, p. 115736, dez. 2021. <https://doi.org/10.1016/j.eswa.2021.115736>.

AZODI, C. B.; TANG, J.; SHIU, S.-H. Opening the Black Box: Interpretable Machine Learning for Geneticists. **Trends in Genetics**, v. 36, n. 6, p. 442–455, jun. 2020. <https://doi.org/10.1016/j.tig.2020.03.005>.

BHATT, S.; GETHING, P. W.; BRADY, O. J.; MESSINA, J. P.; FARLOW, A. W.; MOYES, C. L.; DRAKE, J. M.; BROWNSTEIN, J. S.; HOEN, A. G.; SANKOH, O.; MYERS, M. F.; GEORGE, D. B.; JAENISCH, T.; WINT, G. R. W.; SIMMONS, C. P.; SCOTT, T. W.; FARRAR, J. J.; HAY, S. I. The global distribution and burden of dengue. **Nature**, v. 496, n. 7446, p. 504–507, abr. 2013. <https://doi.org/10.1038/nature12060>.

BRUNKARD, J. M.; CIFUENTES, E.; ROTHENBERG, S. J. Assessing the roles of temperature, precipitation, and enso in dengue re-emergence on the Texas-Mexico border region. **Salud Pública de México**, v. 50, n. 3, jun. 2008. DOI 10.1590/S0036-36342008000300006. Disponível em: http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0036-36342008000300006&lng=en&nrm=iso&tlng=en. Acesso em: 13 ago. 2024.

CABRERA, M.; LEAKE, J.; NARANJO-TORRES, J.; VALERO, N.; CABRERA, J. C.; RODRÍGUEZ-MORALES, A. J. Dengue Prediction in Latin America Using Machine Learning and the One Health Perspective: A Literature Review. **Tropical Medicine and Infectious Disease**, v. 7, n. 10, p. 322, 21 out. 2022. <https://doi.org/10.3390/tropicalmed7100322>.

CAMPBELL, K. M.; HALDEMAN, K.; LEHNIG, C.; MUNAYCO, C. V.; HALSEY, E. S.; LAGUNA-TORRES, V. A.; YAGUI, M.; MORRISON, A. C.; LIN, C.-D.; SCOTT, T. W. Weather Regulates Location, Timing, and Intensity of Dengue Virus Transmission between Humans and Mosquitoes. **PLOS Neglected Tropical Diseases**, v. 9, n. 7, p. e0003957, 29 jul. 2015. <https://doi.org/10.1371/journal.pntd.0003957>.

CAO, Z.; LIU, T.; LI, X.; WANG, J.; LIN, H.; CHEN, L.; WU, Z.; MA, W. Individual and Interactive Effects of Socio-Ecological Factors on Dengue Fever at Fine Spatial Scale: A Geographical Detector-Based Analysis. **International Journal of Environmental Research and Public Health**, v. 14, n. 7, p. 795, 17 jul. 2017. <https://doi.org/10.3390/ijerph14070795>.

CARRINGTON, L. B.; ARMIJOS, M. V.; LAMBRECHTS, L.; SCOTT, T. W. Fluctuations at a Low Mean Temperature Accelerate Dengue Virus Transmission by *Aedes aegypti*. **PLoS Neglected Tropical Diseases**, v. 7, n. 4, p. e2190, 25 abr. 2013. <https://doi.org/10.1371/journal.pntd.0002190>.

CARVAJAL, T. M.; VIACRUSIS, K. M.; HERNANDEZ, L. F. T.; HO, H. T.; AMALIN, D. M.; WATANABE, K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. **BMC Infectious Diseases**, v. 18, n. 1, p. 183, dez. 2018. <https://doi.org/10.1186/s12879-018-3066-0>.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. *In*: KDD '16: THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 13 ago. 2016. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [...]**. San Francisco California USA: ACM, 13 ago. 2016. p. 785–794. DOI 10.1145/2939672.2939785. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>. Acesso em: 16 ago. 2024.

CHEN, Y.; XU, Y.; WANG, L.; LIANG, Y.; LI, N.; YA, Y.; ZHAO, H.; CAZELLES, B.; LIU, Z.; WANG, Z.; BRADY, O. J.; TIAN, H. Anomalias de temperatura no Oceano Índico prevêm tendências globais de dengue a longo prazo. 2024.

CRUZ, L. C. D. T. A. DA et al. Influence of climatic variables on the *Aedes aegypti* and *Culex quinquefasciatus* populations in Mato Grosso, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 53, p. e20190185, 16 mar. 2020.

DE CASTRO, A. F.; DE OLIVEIRA, A. G. Study on Dengue cases using Data Analysis techniques: A case study in the state of Pernambuco, Brazil. *In*: 2021 16TH IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES (CISTI), 23 jun.

2021. **2021 16th Iberian Conference on Information Systems and Technologies (CISTI)** [...]. Chaves, Portugal: IEEE, 23 jun. 2021. p. 1–6. DOI 10.23919/CISTI52073.2021.9476432. Disponível em: <https://ieeexplore.ieee.org/document/9476432/>. Acesso em: 13 ago. 2024.

DELMELLE, E.; HAGENLOCHER, M.; KIENBERGER, S.; CASAS, I. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. **Acta Tropica**, v. 164, p. 169–176, dez. 2016. <https://doi.org/10.1016/j.actatropica.2016.08.028>.

DOBSON, R.; CHALLINOR, A. J.; CHEKE, R. A.; JENNINGS, S.; WILLIS, S. G.; DALLIMER, M. DYNAMICSDM: An R package for species geographical distribution and abundance modelling at high spatiotemporal resolution. **Methods in Ecology and Evolution**, v. 14, n. 5, p. 1190–1199, maio 2023. <https://doi.org/10.1111/2041-210X.14101>.

DO VALLE VARELA, J. P. et al. THE INCREASE IN THE NUMBER OF DENGUE CASES IN BRAZIL AND THE CORRELATION WITH DEFORESTATION AND GLOBALIZATION: A SYSTEMATIC REVIEW. **Health and Society**, v. 4, n. 02, p. 23-34, 2024.

EKANAYAKE, I. U.; MEDDAGE, D. P. P.; RATHNAYAKE, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). **Case Studies in Construction Materials**, v. 16, p. e01059, jun. 2022. <https://doi.org/10.1016/j.cscm.2022.e01059>.

FAO; WHO; WOA. **Advancing Joint Risk Assessment using the One Health Approach in WHO South-East Asia Region. Report of the meeting. Colombo, Sri Lanka, 25 – 27 July 2023.** [S. l.]: WOA (World Organisation for Animal Health), 12 jun. 2024. DOI 10.20506/woah.3510. Disponível em: <https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloId=43913>. Acesso em: 17 ago. 2024.

GALEANA-PIZAÑA, J. M.; CRUZ-BELLO, G. M.; CAUDILLO-COS, C. A.; JIMÉNEZ-ORTEGA, A. D. Impact of deforestation and climate on spatio-temporal spread of dengue fever in Mexico. **Spatial and Spatio-temporal Epidemiology**, v. 50, p. 100679, ago. 2024. <https://doi.org/10.1016/j.sste.2024.100679>.

GIESEN, C. et al. The impact of climate change on mosquito-borne diseases in Africa. **Pathogens and Global Health**, v. 114, n. 6, p. 287–301, 2020.

GOMES, A. F.; NOBRE, A. A.; CRUZ, O. G. Temporal analysis of the relationship between dengue and meteorological variables in the city of Rio de Janeiro, Brazil, 2001-2009. **Cadernos de Saúde Pública**, v. 28, n. 11, p. 2189–2197, nov. 2012. <https://doi.org/10.1590/S0102-311X2012001100018>.

GUBLER, D. J. The Economic Burden of Dengue. **The American Society of Tropical Medicine and Hygiene**, v. 86, n. 5, p. 743–744, 1 maio 2012. <https://doi.org/10.4269/ajtmh.2012.12-0157>.

HALASA, Y. A.; SHEPARD, D. S.; ZENG, W. Economic Cost of Dengue in Puerto Rico. **The American Society of Tropical Medicine and Hygiene**, v. 86, n. 5, p. 745–752, 1 maio 2012. <https://doi.org/10.4269/ajtmh.2012.11-0784>.

JALIL, E. S.; SHIMIZU, Y. Neglected tropical diseases (NTDs) are a diverse set of 20 diseases and disease groups with a singular commonality: their impact on impoverished communities. Together they affect more than 1 billion people with devastating health, social and economic consequences.

JUNG, Y.; HU, J. A K -fold averaging cross-validation procedure. **Journal of Nonparametric Statistics**, v. 27, n. 2, p. 167–179, 3 abr. 2015. <https://doi.org/10.1080/10485252.2015.1010532>.

JUNIOR, J. B. S.; MASSAD, E.; LOBAO-NETO, A.; KASTNER, R.; OLIVER, L.; GALLAGHER, E. Epidemiology and costs of dengue in Brazil: a systematic literature review. **International Journal of Infectious Diseases**, v. 122, p. 521–528, set. 2022. <https://doi.org/10.1016/j.ijid.2022.06.050>.

LI, Y.; DOU, Q.; LU, Y.; XIANG, H.; YU, X.; LIU, S. Effects of ambient temperature and precipitation on the risk of dengue fever: A systematic review and updated meta-analysis, **Environmental Research**, v. 191, 2020, ISSN 0013-9351, <https://doi.org/10.1016/j.envres.2020.110043>.

LIMA, E. D. A.; FIRMINO, J. L. D. N.; GOMES FILHO, M. F. A relação da previsão da precipitação pluviométrica e casos de dengue nos estados de Alagoas e Paraíba nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 23, n. 3, p. 264–269, set. 2008. <https://doi.org/10.1590/S0102-77862008000300001>.

MACHAULT, V.; VIGNOLLES, C.; PAGÈS, F.; GADIAGA, L.; GAYE, A.; SOKHNA, C.; TRAPE, J.-F.; LACAUX, J.-P.; ROGIER, C. Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. **Malaria Journal**, v. 9, n. 1, p. 252, dez. 2010. <https://doi.org/10.1186/1475-2875-9-252>.

MACKENZIE, J. S.; JEGGO, M. The One Health Approach—Why Is It So Important? **Tropical Medicine and Infectious Disease**, v. 4, n. 2, p. 88, 31 maio 2019. <https://doi.org/10.3390/tropicalmed4020088>.

MCMICHAEL, A. J.; WORLD HEALTH ORGANIZATION (Orgs.). **Climate change and human health: risks and responses**. Geneva: World Health Organization, 2003.

MENDONÇA, F. de A. SAÚDE PÚBLICA, URBANIZAÇÃO E DENGUE NO BRASIL. 2009.

MINISTÉRIO DA SAÚDE. Ministério da Saúde Secretaria de Vigilância em Saúde e Ambiente. Boletim epidemiológico. v: 55, 4 de jul, 2014. Acesso em: 12 de jul. 2024.

MISSLIN, R.; TELLE, O.; DAUDÉ, E.; VAGUET, A.; PAUL, R. E. Urban climate versus global climate change—what makes the difference for dengue? **Annals of the New York Academy of Sciences**, v. 1382, n. 1, p. 56–72, out. 2016. <https://doi.org/10.1111/nyas.13084>.

MORAIS, M. M. A proporção de Mata Atlântica remanescente no entorno de áreas urbanas relaciona-se à incidência de dengue? 2011.

MULAKOLI, F.; GACHARA, G.; NDOMBI, E.; KHAMADI, S. Dengue Virus Surveillance and Blood Safety: A One Health Perspective. *In*: APARECIDA SPERANÇA, M. (org.). **Infectious Diseases**: IntechOpen, 2023. v. 22, . DOI 10.5772/intechopen.109413. Disponível em: <https://www.intechopen.com/chapters/85474>. Acesso em: 16 ago. 2024.

NOHARA, Y.; MATSUMOTO, K.; SOEJIMA, H.; NAKASHIMA, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. **Computer Methods and Programs in Biomedicine**, v. 214, p. 106584, fev. 2022. <https://doi.org/10.1016/j.cmpb.2021.106584>.

ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2024. Dengue. Disponível em: <https://www.paho.org/pt/topicos/dengue>. Acesso em: 5 jun. 2024

PAZ-BAILEY, G.; ADAMS, L. E.; DEEN, J.; ANDERSON, K. B.; KATZELNICK, L. C. Dengue. **The Lancet**, v. 403, n. 10427, p. 667–682, fev. 2024. [https://doi.org/10.1016/S0140-6736\(23\)02576-X](https://doi.org/10.1016/S0140-6736(23)02576-X).

PEARSON; HENRICI, F.R.S. VII. Mathematical contributions to the theory of evolution.— III. Regression, heredity, and panmixia. **Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character**, v. 187, p. 253–318, 31 dez. 1896. <https://doi.org/10.1098/rsta.1896.0007>.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISSEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D. Scikit-learn: Machine Learning in Python. **MACHINE LEARNING IN PYTHON**, 2011. .

PETCH, S.; DONG, B.; QUAIFFE, T.; KING, R.; HAINES, K. Water and Energy budgets over hydrological basins on short and long timescales. 15 dez. 2022. DOI 10.5194/egusphere-2022-1237. Disponível em: <https://egusphere.copernicus.org/preprints/2022/egusphere-2022-1237/>. Acesso em: 19 ago. 2024.

PHILLIPS, S. J.; ANDERSON, R. P.; SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecological Modelling**, v. 190, n. 3–4, p. 231–259, jan. 2006. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.

QI, X.; WANG, Y.; LI, Y.; MENG, Y.; CHEN, Q.; MA, J.; GAO, G. F. The Effects of Socioeconomic and Environmental Factors on the Incidence of Dengue Fever in the Pearl

River Delta, China, 2013. **PLOS Neglected Tropical Diseases**, v. 9, n. 10, p. e0004159, 27 out. 2015. <https://doi.org/10.1371/journal.pntd.0004159>.

R Core Team (2023). R: Uma linguagem e ambiente para computação estatística. R Foundation para Computação Estatística, Viena, Áustria. URL <https://www.R-project.org/>.

ROSA MEURER, I.; SOARES COIMBRA, E. Doenças tropicais negligenciadas e o seu contexto no Brasil. **HU Revista**, v. 48, p. 1–2, 9 jun. 2022. <https://doi.org/10.34019/1982-8047.2022.v48.37905>.

ROY, S. K.; BHATTACHARJEE, S. Dengue virus: epidemiology, biology, and disease aetiology. **Canadian Journal of Microbiology**, v. 67, n. 10, p. 687–702, out. 2021. <https://doi.org/10.1139/cjm-2020-0572>.

SCAVUZZO, J. M. et al. Modeling Dengue vector population using remotely sensed data and machine learning. *Acta Tropica*, v. 185, p. 167–175, 1 set. 2018.

SCHMIDT, K. A.; OSTFELD, R. S. BIODIVERSITY AND THE DILUTION EFFECT IN DISEASE ECOLOGY. **Ecology**, v. 82, n. 3, p. 609–619, mar. 2001. [https://doi.org/10.1890/0012-9658\(2001\)082\[0609:BATDEI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[0609:BATDEI]2.0.CO;2).

SEAH, A.; AIK, J.; NG, L.-C.; TAM, C. C. The effects of maximum ambient temperature and heatwaves on dengue infections in the tropical city-state of Singapore – A time series analysis. **Science of The Total Environment**, v. 775, p. 145117, jun. 2021. <https://doi.org/10.1016/j.scitotenv.2021.145117>.

SECRETARIA DE ESTADO DA SAÚDE DE SERGIPE. Boletim epidemiológico de arboviroses. 32^a ed. Secretaria de Estado da Saúde de Sergipe, 2024. Disponível em: https://saude.se.gov.br/wp-content/uploads/2024/08/Boletim-Arboviroses-SE_32.pdf. Acesso em: 16 ago. 2024.

SERGIPE. Lei complementar N°. 33, de 26 de dezembro de 1996. Institui o Código de Organização e de Procedimento da Administração Pública do Estado de Sergipe. Aracaju, SE: 175° da Independência e 108° da República. 1996

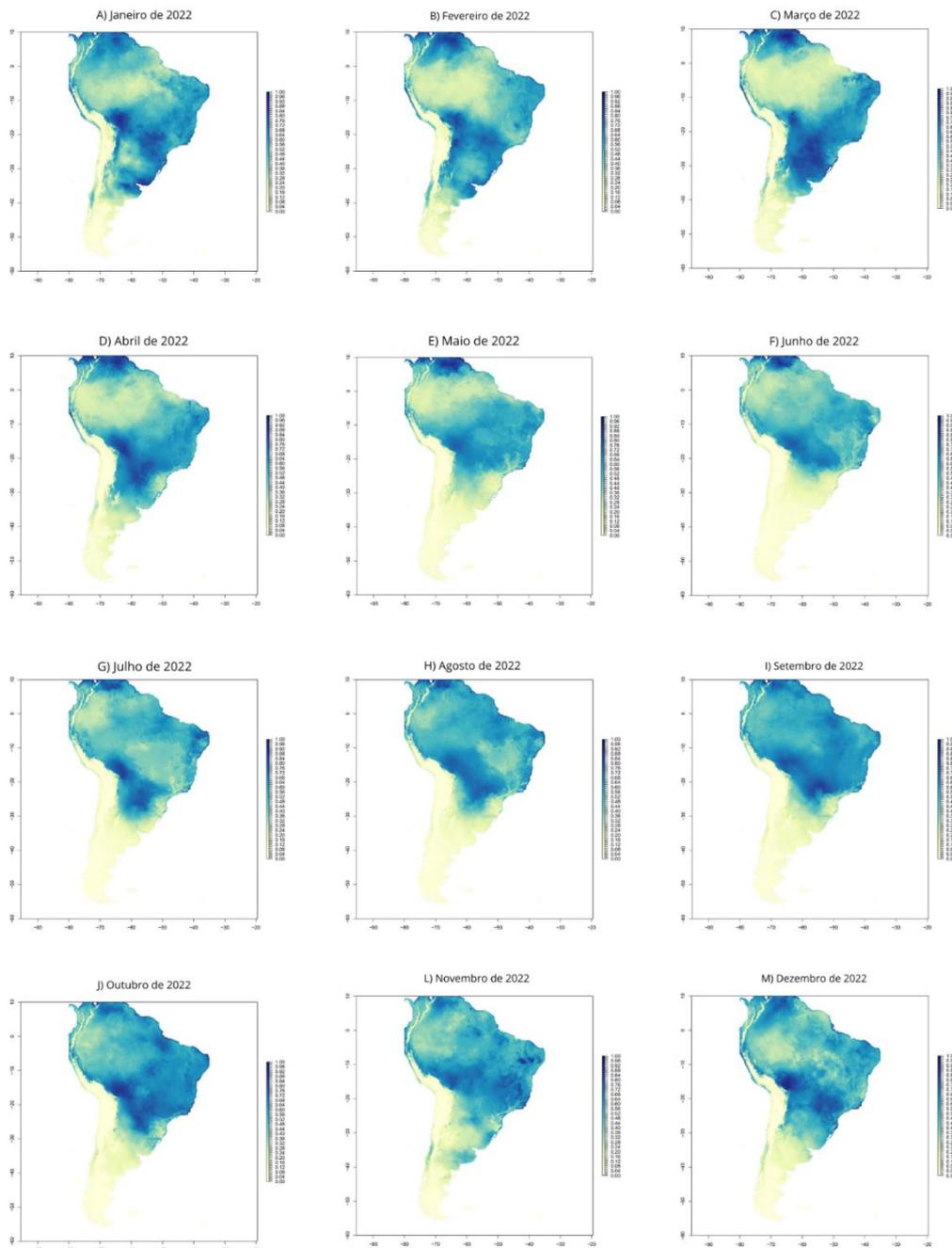
SILVA, F. D.; SANTOS, A. M. D.; CORRÊA, R. D. G. C. F.; CALDAS, A. D. J. M. Temporal relationship between rainfall, temperature and occurrence of dengue cases in São Luís, Maranhão, Brazil. **Ciência & Saúde Coletiva**, v. 21, n. 2, p. 641–646, fev. 2016. <https://doi.org/10.1590/1413-81232015212.09592015>.

SILVA, J. S.; MARIANO, Z. D. F.; SCOPEL, I. A DENGUE NO BRASIL E AS POLÍTICAS DE COMBATE AO Aedes Aegypti: DA TENTATIVA DE ERRADICAÇÃO ÀS POLÍTICAS DE CONTROLE. **Hygeia - Revista Brasileira de Geografia Médica e da Saúde**, v. 4, n. 6, p. 163–175, 25 jun. 2008. <https://doi.org/10.14393/Hygeia416906>.

- ŠTRUMBELJ, E.; KONONENKO, I. Explaining prediction models and individual predictions with feature contributions. **Knowledge and Information Systems**, v. 41, n. 3, p. 647–665, dez. 2014. <https://doi.org/10.1007/s10115-013-0679-x>.
- TAUIL, P. L. Aspectos críticos do controle do dengue no Brasil. **Cadernos de Saúde Pública**, v. 18, n. 3, p. 867–871, jun. 2002. <https://doi.org/10.1590/S0102-311X2002000300030>.
- VALVERDE, R. 2013. Fiocruz: Doenças Negligenciadas. Disponível em: <https://agencia.fiocruz.br/print/4740>. Acesso em: 1 jan. 2024
- VEGA GARCÍA, M.; AZNARTE, J. L. Shapley additive explanations for NO2 forecasting. **Ecological Informatics**, v. 56, p. 101039, mar. 2020. <https://doi.org/10.1016/j.ecoinf.2019.101039>.
- VIANA, D. V.; IGNOTTI, E. A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática. **Revista Brasileira de Epidemiologia**, v. 16, n. 2, p. 240–256, jun. 2013. <https://doi.org/10.1590/S1415-790X2013000200002>.
- VIEIRA, J. S.; ROCHA, R. D. M.; JUSTO, W. R. Evaluation of the effectiveness of LIRAA as an instrument to monitor Dengue. **Revista Caribeña de Ciencias Sociales**, v. 12, n. 8, p. 3615–3633, 22 dez. 2023. <https://doi.org/10.55905/rcssv12n8-012>.
- WATSON, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. **Nature Genetics**, v. 51, n. 8, p. 1207–1214, ago. 2019. <https://doi.org/10.1038/s41588-019-0439-2>.
- WHITMIRE, R. E.; BURKE, D. S.; NISALAK, A.; HARRISON, B. A.; WATTS, D. M. Effect of Temperature on the Vector Efficiency of *Aedes aegypti* for Dengue 2 Virus. **The American Journal of Tropical Medicine and Hygiene**, v. 36, n. 1, p. 143–152, 1 jan. 1987. <https://doi.org/10.4269/ajtmh.1987.36.143>.
- WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MULLER, K.; VAUGHAN, D. dplyr: Uma gramática de manipulação de dados. Pacote R versão 1.1.4, 2023 <https://github.com/tidyverse/dplyr>, <https://dplyr.tidyverse.org>.
- WONGKOON, S.; JAROENSUTASINEE, M.; JAROENSUTASINEE, K. Fatores climáticos que influenciam a ocorrência de dengue em Nakhon Si Thammarat, Tailândia. 2013.
- WORLD HEALTH ORGANIZATION. (2020). Ending the neglect to attain the sustainable development goals: a road map for neglected tropical diseases 2021–2030: overview. World Health Organization. <https://iris.who.int/handle/10665/332094>. Licença: CC BY-NC-SA 3.0 IGO

Apêndice A

Mapas mensais da adequabilidade climática de *Aedes aegypti* para o ano de 2022. As colorações mais claras mostram zonas menos adequadas climaticamente, enquanto que as regiões com coloração azul escuro mostram zonas mais adequadas climaticamente ao vetor da dengue.



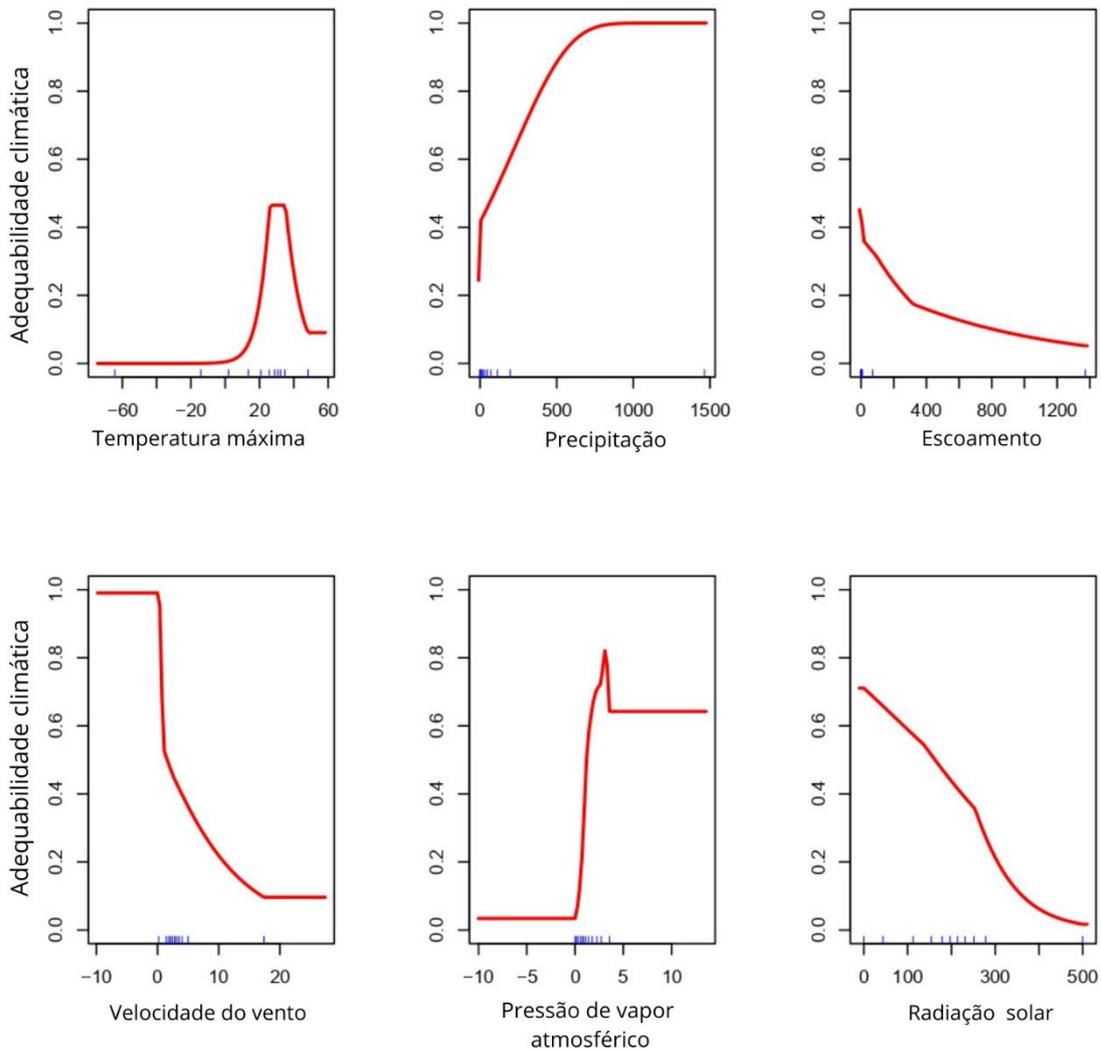
Apêndice B

A seguinte tabela mostra as variáveis utilizadas na construção dos modelos dinâmicos de nicho climático. Na primeira coluna é possível observar as variáveis, em seguida, vemos a importância de permutação e na última coluna vemos a porcentagem de contribuição das variáveis para o melhor modelo gerado.

| Variável | Importância de permutação | Porcentagem de contribuição |
|---------------------|---------------------------|-----------------------------|
| Pressão de Vapor | 29,9 | 80,7 |
| Velocidade do Vento | 11,1 | 8 |
| Radiação Solar | 7,2 | 4,3 |
| Temperatura Máxima | 40,7 | 4,1 |
| Escoamento | 3,6 | 2,4 |
| Precipitação | 7,5 | 0,5 |

Apêndice C

Curvas de resposta das variáveis climática utilizadas na construção dos modelos dinâmicos de distribuição de espécies. O eixo y possui valores de adequabilidade climática, o eixo x, possui valores das variáveis climáticas.



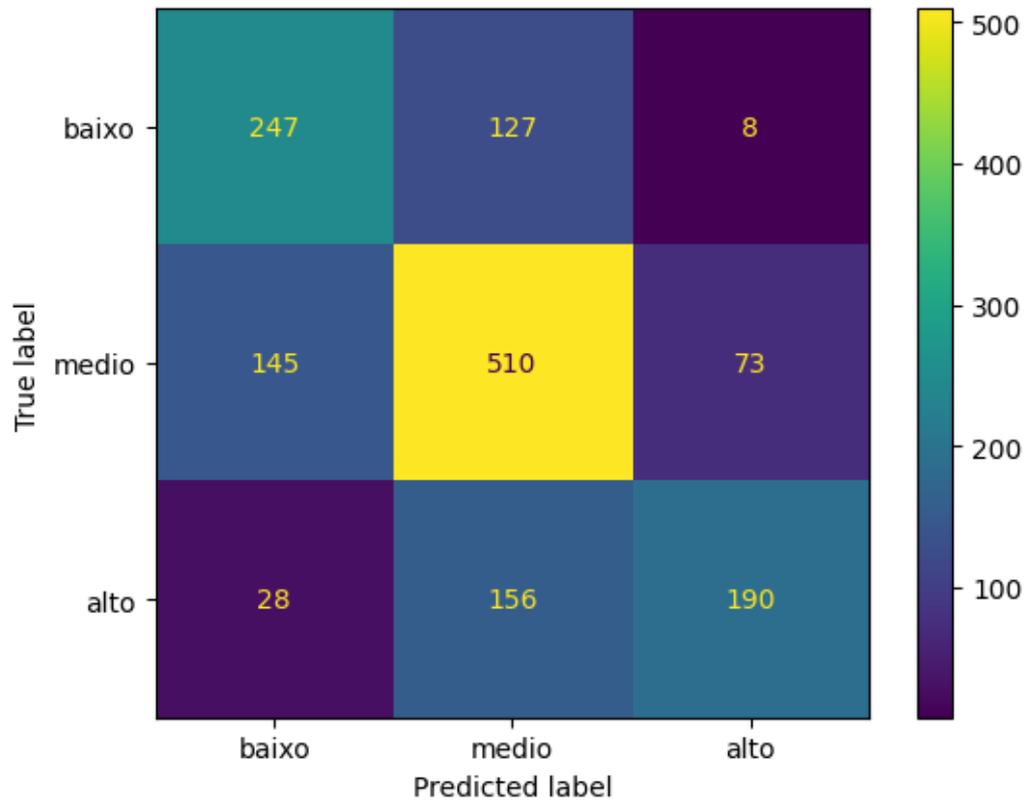
Apêndice D

Valores das métricas em cada modelo nos cenários de treinos e de testes.

| Métricas | Treino | | | | Teste | | | |
|--------------|--------|-------|-------|-------|-------|-------|-------|-------|
| | XGB | RF | DL | SMV | XGB | RF | DL | SMV |
| Precision | 0.644 | 0.653 | 0.584 | 0.551 | 0.621 | 0.617 | 0.628 | 0.564 |
| Recall | 0.618 | 0.598 | 0.572 | 0.504 | 0.584 | 0.56 | 0.577 | 0.512 |
| F1-score | 0.625 | 0.608 | 0.567 | 0.516 | 0.597 | 0.575 | 0.574 | 0.522 |
| Macro avg | 0.629 | 0.620 | 0.574 | 0.525 | 0.600 | 0.584 | 0.593 | 0.533 |
| Weighted avg | 0.639 | 0.637 | 0.586 | 0.554 | 0.613 | 0.600 | 0.604 | 0.555 |
| Accuracy | 0.638 | 0.637 | 0.586 | 0.561 | 0.613 | 0.597 | 0.599 | 0.559 |

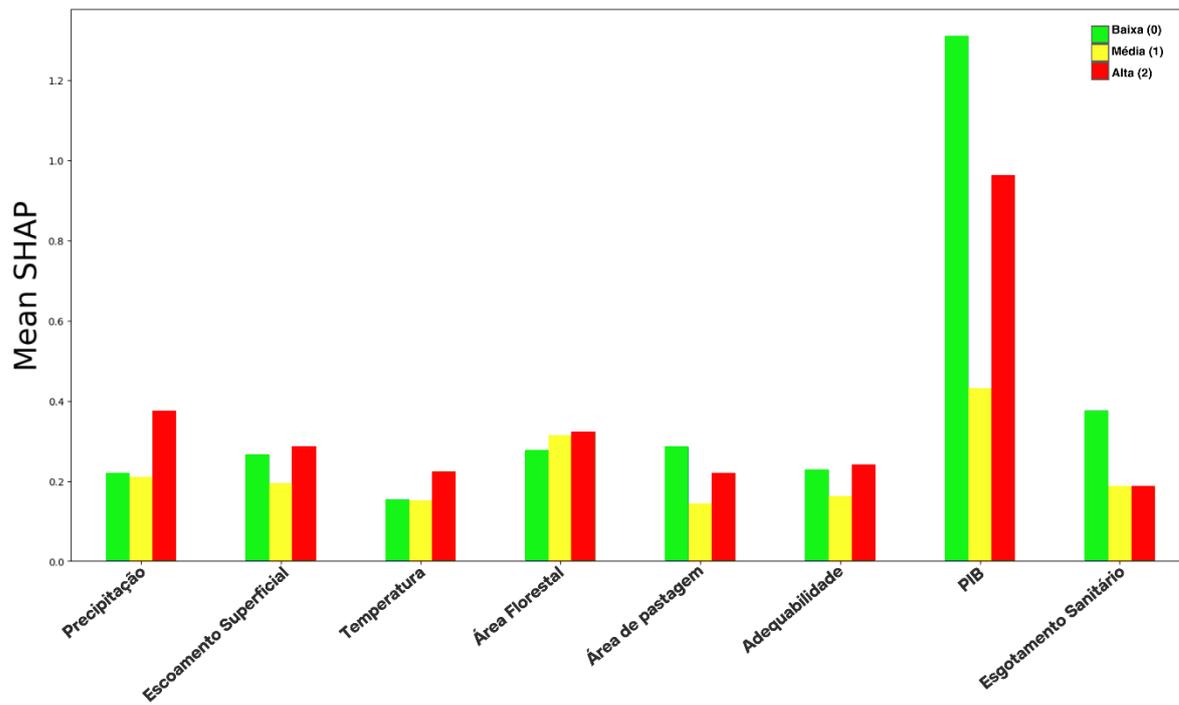
Apêndice E

Matriz de confusão gerada a partir do eXtreme Gradient Boosting Classifier no treino.



Apêndice F

Gráfico de barras mostrando os valores médios de SHAP de cada variável sobre as categorias (Baixa, Média e Alta).



Apêndice G

Este trabalho foi desenvolvido com a colaboração de Mariana Santos Rodrigues, Irene Barbosa da Fonseca Teixeira, Diene Oliveira Santos, Thais de Andrade Moura e Pablo Ariel Martinez.