

Descritor baseado em MFCCs para Detecção do Mosquito *Aedes Aegypti*

Antônio Rodrigues* Eurico Júnior** Bianca Soares**
Mayra Oliveira** Ricardo Rabelo* Deborah Magalhães*,**

* *Programa de Pós-Graduação em Engenharia Elétrica - PPGEE,
Universidade Federal do Piauí - UFPI*

** *Departamento de Sistemas de Informação,
Campus Senador Helvídio Nunes de Barros,
Universidade Federal do Piauí - UFPI*

Resumo: Fatores como urbanização descontrolada, aquecimento global e aumento da resistência dos mosquitos aos inseticidas têm tornado o combate a proliferação do *Aedes aegypti* desafiador. O processo de identificação manual dos focos através da identificação de locais com água parada e instalação de armadilhas para mosquitos envolve recursos financeiros, humanos e consome muito tempo. Nesse contexto, soluções que se baseiam em processamento de áudio e técnicas de aprendizado de máquina para detecção automática de mosquitos através dos tons produzidos durante seu voo têm sido investigadas. Soluções baseadas em aprendizado profundo têm apresentado resultados promissores, no entanto, exigem um grande volume de dados e possuem alto custo computacional, tornando sua ampla adoção desafiadora especialmente quando se trata de dispositivos de captura e processamento do áudio baixo poder computacional como, por exemplo, *smartphones*. Diante disso, este trabalho propõem um descritor compacto e eficaz para detectar a presença do *Aedes aegypti* baseado em coeficientes do domínio Mel-espectral (MFCCs). Os resultados apontam que o descritor composto por 40 MFCCs juntamente com o classificador XGBoost consegue detectar a presença do *Aedes aegypti* com valores acima de 95% para as métricas acurácia, Kappa e F1-score. Conseqüentemente, acreditamos que nossas descobertas podem apoiar a implantação de um sistema de detecção automática do *Aedes aegypti* escalável e de baixo custo.

Abstract: Factors such as urbanization, global warming, and increased resistance of mosquitoes to insecticides have made combating the proliferation of *Aedes aegypti* challenging. The manual identification of outbreaks throughout mosquito traps involves financial and time-consuming resources. In this context, solutions based on audio processing and machine learning techniques for the automatic detection of mosquitoes through the tones produced during their flight have been investigated. Solutions based on deep learning have shown promising results; however, they require a large volume of data and have a high computational cost, making their wide adoption challenging, especially when it comes to audio capture and processing devices with low computational power, such as, for example, *smartphones*. Therefore, this work proposes a compact descriptor capable of and effective in detecting the presence of *Aedes aegypti* based on Mel-Frequency Cepstral Coefficients (MFCCs). The results show that the descriptor composed of 40 MFCCs combined with the XGBoost classifier can detect the presence of *Aedes aegypti* with values above 95% of accuracy, Kappa and F1-score. Consequently, we believe that our findings can support the implementation of a scalable and low-cost automatic *Aedes aegypti* detection system.

Keywords: Audio processing; Dengue; Vector control; Bioacoustics; Machine Learning

Palavras-chaves: Processamento de áudio; Dengue; Detecção de mosquitos; Aprendizado de Máquina; Bioacústica

1. INTRODUÇÃO

A dengue é a infecção viral mais prevalente e promove cerca de 40.000 mortes anualmente ¹. Durante 2019, a maioria dos países endêmicos enfrentou surtos massivos de dengue com grandes impactos econômicos e humanos. Além disso, os países ocidentais também estão sob risco crescente da doença devido ao aquecimento global ². A doença possui o mosquito *Aedes aegypti* como principal vetor de transmissão. Tal mosquito também é vetor do vírus da febre amarela, Chikungunya e Zika. Além do aquecimento global, outro aspecto desafiador consiste nos mosquitos estarem se tornando mais resistentes aos inseticidas (Şengül Demirak and Canpolat, 2022). Nesse contexto, compreender o número e a localização de potenciais vetores é imprescindível para ajudar na redução da transmissão (Kiskin et al., 2020). Diferentes estratégias têm sido empregadas para combater a proliferação do *Aedes aegypti* como, por exemplo, identificação de locais com água parada, que servem para a reprodução do mosquito, instalação de armadilhas para mosquitos e utilização de mosquitos geneticamente modificados que não se reproduzem (Fernandes et al., 2021).

A conscientização da comunidade local é fundamental para combater a incidência do mosquito em uma região. No entanto, há uma escassez de dados devido ao processo de identificação manual dos focos que envolve recursos financeiros, humanos e tempo. Assim, são necessários métodos de identificação de focos do mosquito de baixo custo e de fácil adoção. Nesse contexto, soluções que se baseiam em processamento de áudio e técnicas de aprendizado de máquina para detecção automática de mosquitos através dos tons produzidos durante seu vôo têm sido investigadas na literatura, entre elas, destacamos: Fernandes et al. (2021) utilizaram áudios de batidas de asas do *Aedes aegypti* para treinar uma rede neural convolucional (CNN) a fim de detectar a presença do *Aedes aegypti* e obtiveram uma acurácia de 97,65% ($\pm 0,55$). Kiskin et al. (2020) treinaram uma rede neural convolucional com espectrogramas extraídos dos áudios de 2s de duração do vôo de mosquitos tanto criados em laboratório quanto presentes em cenários urbanos para identificar seis espécies de mosquitos transmissores de doenças, incluindo o *Aedes aegypti*. Mulchandani et al. (2019) utilizaram uma base de dados com 6 espécies de mosquito, incluindo *Aedes aegypti*, e treinaram uma arquitetura de CNN própria com espectrogramas retirados a partir dos áudios e alcançaram uma acurácia de classificação de 86%.

Apesar das soluções baseadas em aprendizado profundo apresentarem resultados promissores para a tarefa de classificação de mosquitos, existem desafios, entre eles: o ruído de fundo presente nos áudios captados, diferentes espécies

de mosquitos possuem distribuições de frequência sobrepostas, tornando difícil diferenciá-las, informações temporais, como a hora e data pode ajudar a melhorar a precisão da classificação, mas raramente estão disponíveis (Joshi and Miller, 2021). Além disso, tais soluções exigem um grande volume de dados e possuem alto custo computacional, tornando sua ampla adoção desafiadora especialmente quando se trata de um smartphone utilizado para captura, processamento e classificação do sinal de áudio. Diante disso, este trabalho propôs um descritor compacto baseado em coeficientes do domínio espectral (MFCCs) para caracterizar a presença do *Aedes aegypti*. Nós escolhemos o *Aedes aegypti* por ser o principal vetor de diversas doenças, como dengue, febre amarela, chikungunya e Zika vírus. Para isso, aplicamos uma etapa de pré-processamento que inclui aumento de dados para garantir o equilíbrio entre as classes. Em sequência, extraímos características do domínio cepstral dos sinais de áudio. Na etapa de classificação, empregamos classificadores amplamente utilizados no contexto de reconhecimento de som: máquina de vetores de suporte (SVM), floresta aleatória (RF), XGBoost (XGB). Por fim, avaliamos o desempenho da classificação por meio de acurácia, kappa e F1-score.

As contribuições deste trabalho são: (1) propomos um descritor eficaz para caracterizar a presença do *Aedes aegypti* baseado em coeficientes do domínio cepstral que supera o estado da arte na detecção do *Aedes aegypti* através de sons, incluindo modelos de aprendizado profundo; (2) empregamos o aumento de dados para contornar conjuntos de dados pequenos e desequilibrados; O descritor proposto tem uma dimensão baixa e é baseado em apenas uma característica, o que reduz o número de dependências de cálculo e biblioteca. Diante disso, acreditamos que nossas descobertas podem apoiar a implantação em dispositivos de baixa computação para detecção de focos do mosquito *Aedes aegypti* de maneira acessível, escalável e em tempo real.

A restante da estrutura deste artigo está organizada da seguinte forma. A seção 2 discute trabalhos relevantes relacionados ao processamento de áudio para detecção de mosquitos. A seção 3 descreve a metodologia seguida para detectar a presença do *Aedes Aegypti*. Os resultados dos experimentos e sua discussão são apresentados na Seção 4, incluindo a comparação do desempenho da abordagem proposta com o estado da arte na classificação de mosquitos. Na Seção 5, são apresentadas as conclusões e perspectivas de trabalhos futuros.

2. TRABALHOS RELACIONADOS

Esta seção apresenta um levantamento do estado da arte sobre o emprego da detecção automática do mosquito *Aedes aegypti* por extração de características e classificação dos sons produzidos por estes, utilizando técnicas clássicas de aprendizado de máquina e aprendizado profundo. As principais características dos trabalhos apresentados estão resumidas na Tabela 1.

¹ <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>

² <https://www.thesynergist.org/annual-report-2019/break-dengue.html>

Tabela 1. Resumo dos trabalhos relacionados de nosso levantamento no atual estado da arte

Trabalho	Qt. Classe	DP	PP	EC	ALG	Métricas
Kiskin et al. (2021)	2 classes	Sim	Não	log-mel spec	BNNs	ROC: 0,927 PR: 0,716
Alar and Fernandez (2021)	2 classes	Sim	Sim	FFT	IQR	ACC: 92,8%
Fernandes et al. (2021)	2 classes	Sim	Sim	FFT	CNN	ACC: 97,65%
Okawa et al. (2019)	2 classes	Sim	Sim	MFCCs	CNN	ACC: 88,4% %
Kiskin et al. (2017)	2 classes	Sim	Sim	CWT	CNN	ROC: 0,970

DP: Dataset Público; PP: Pré-processamento; EC: Extração de Características;
 FFT: Transformada Rápida de Fourier; CWT: transformada wavelet contínua;
 ALG: Algoritmo; ACC: Acurácia/Precisão.

Kiskin et al. (2021) utilizaram redes neurais convolucionais bayesianas aplicadas ao contexto de detecção de mosquitos. Eles realizaram a coleta automatizada e processamento para identificar assinaturas acústicas em mais 1.500 h de áudios não rotulados. Desse modo, os autores extraíram log-mel espectrogramas dos áudios e utilizaram as redes neurais Bayesianas (BNNs, do inglês: textit bayesian neural networks) para diferenciar o som produzido pelos mosquitos do ruído de fundo e para diferenciar espécies de mosquitos. Os experimentos foram realizados utilizando áudios capturados em diferentes condições. No Teste A, os áudios foram capturados por smartphones de baixo custo na Tanzânia, em um ambiente que simula a construção das moradias locais. Já os áudios do Teste B foram capturados de jaulas no departamento de Zoologia de Oxford. Os resultados alcançados na detecção da presença de mosquitos com os áudios do Teste A foram ROC AUC acima de 0,93 e PR AUC acima de 0,90. Já com os áudios do Teste B, os autores obtiveram um ROC AUC de 0,770 e um PR AUC de 0,749. Na classificação multiespecie, os autores utilizaram apenas os áudios do Teste A, formado por 36 espécies, mas apenas as 8 mais frequentes foram empregadas na classificação, incluindo *Aedes Aegypti*. Os resultados alcançados foram um ROC AUC of 0,927 and PR AUC de 0,716.

Alar and Fernandez (2021) apresentaram um modelo de classificação de baixa dimensionalidade baseado na mediana e no intervalo entre quartis (IQR, do inglês: *interquartile range*) de cada um dos 26-value filter bank extraídos dos espectrogramas das instâncias de áudio. Tal descritor foi utilizado na detecção de mosquitos dos seguintes gêneros: *Aedes aegypti*, *Anopheles gambiae* e *Culex quinquefasciatus*. Além disso, uma classe adicional foi considerada: *No-Mosquito*. O conjunto de 365 instâncias de gravações de áudio também foi utilizado no treinamento de uma rede neural convolucional (CNN, do inglês: *convolutional neural networks*). Os resultados foram avaliados em termos de precisão da classificação, tempo de treinamento e tempo de previsão da resposta. O desempenho de classificação com o descritor proposto foi de 92,8% de acurácia, sendo superior à acurácia de 86,6% obtida pela CNN. Além disso, o tempo médio de treinamento do modelo de classificação com o descritor proposto foi de 71,16 segundos, enquanto o tempo médio para o treinamento da CNN foi de 151,036. Essa diferença foi reduzida na etapa de predição, onde o descritor proposto realizou a predição em 45,68 segundos e o tempo de predição da CNN foi de 55 segundos.

Fernandes et al. (2021) empregam técnicas de aprendizagem de máquina e análise de áudios capturados por meio de smartphones. Os autores utilizaram amostras de diferentes espécies de mosquitos, incluindo o *Aedes aegypti*, para treinar uma rede neural convolucional (CNN, do inglês: *convolutional neural networks*). Eles utilizaram espectrogramas para representação visual da frequência

do batimento das asas dos mosquitos. Os espectrogramas foram adotados no treinamento da rede considerando três cenários de classificação: binário, multiclasse e um comitê de classificadores binários. A avaliação da classificação binária e comitê obtiveram acurácia de 97,65% ($\pm 0,55$) e 94,56% ($\pm 0,77$). A classificação multiclasse obteve acurácia de 78,12% ($\pm 2,09$). Assim pode-se perceber que o comitê obteve melhor desempenho na acurácia de (96,82% $\pm 1,62$), seguido da multiclasse específico para o *Aedes aegypti* com (90,23% $\pm 3,83$) e do binário com (88,49% $\pm 6,68$). A classificação binária e multiclasse apresentam um melhor equilíbrio e precisão na classificação dos mosquitos. Já o comitê obteve uma menor precisão, no entanto, foi considerado pelos autores o mais poderoso para detectar *Aedes aegypti* em seu conjunto de dados.

Os autores de Okawa et al. (2019) trabalham num novo método de representação de dados para classificação dos áudios dos mosquitos usando uma rede neural profunda. A proposta é transformar uma onda de áudio bruta em sequência de valor inteiro em ondas de pulso multicanal baseada nos bits ou padrão de imagem de bits, isto dependendo dos requisitos da entrada na rede neural. Os experimentos das tarefas mostram que a representação de bits de uma onda de áudio obteve desempenho favorável em comparação aos demais métodos da literatura, onde os resultados alcançados no primeiro teste na representação de bits da onda de áudio tiveram taxa de acurácia de 88,4% em relação a outros métodos comparados. Já no segundo teste, o que era binário na representação para classificação, estava sujeito a efeitos de ruído, no entanto, apesar da representação de bits ter sido robusta no ambiente ruidoso a mesma teve suas taxas prejudicadas. Ainda assim, os autores defendem que a representação de bits para forma de onda de áudio é essencial para as atividades de classificação de áudio dos mosquitos em condições de treinamento de rede neural.

Em Kiskin et al. (2017) os autores propuseram apresentar uma aplicação de aprendizagem profunda para detecção de eventos de áudios de mosquitos. No qual desenvolveram redes neurais convolucionais (CNNs, do inglês: *convolutional neural networks*) operando em transformações com wavelet nas gravações de áudio dos mosquitos. A rede neural é treinada em um espectrograma de wavelet bruto. Além disso, realizaram algumas comparações com classificadores convencionais, condicionando os recursos e ajustando manualmente e genérico, com intuito de enfatizar a precisão da aprendizagem automática de recursos profundos. De acordo com essa proposta, a detecção é alcançada com métricas de desempenho que superam principalmente os métodos de algoritmos existentes na literatura, além de favorecer as métricas de desempenho para alcançar a detecção automática dos áudios dos mosquitos. A filtragem mediada das previsões da (CNN) condicionada a wavelet tem um impacto considerável nas métricas de desempenho, permitindo que a proposta supere os especialistas humanos. A precisão de rótulos humanos foram fornecidos como absolutos ($y_i = 1$, $y_i = 0$), pois, um rótulo incorreto pode acarretar grandes penalidades nas áreas sobre a curva (ROC, do inglês: textit area under the curve) e precisão recall. Isso a torna capaz de realizar classificações com F1 score, curva da área de precisão recall (PR, do inglês: textit precision-recall curve areas) e áreas do ROC,

muito superior de 0,970 para proposta (CNN-wavelet), em comparação com 0,873, 0,901 e 0,874 os três modelos dos especialistas. Outras melhorias levantadas pelos autores estão relacionadas a exemplo de agregação de classificador e agrupamento temporal.

Logo, nos inspiramos nesses trabalhos para investigar a utilização de características cepstrais para produzir descritores de baixa dimensionalidade. Ainda, para contornar o problema de base de dados com classes pequenas e desbalanceadas, o trabalho visa um pré-processamento que inclui aumento dos dados baseado na variação de semitons para balancear as classes.

3. MATERIAIS E MÉTODOS

A metodologia proposta para detectar a presença do *Aedes aegypti*, está ilustrada na Figura 1. A primeira etapa consiste na aquisição de dados; o segundo estágio é o processamento de amostras de áudio para garantir a uniformidade em suas propriedades e o equilíbrio entre as classes. A terceira etapa corresponde à extração manual dos MFCCs. Por fim, na quarta e quinta etapas, as características extraídas são classificadas e diferentes métricas são usadas para avaliar o desempenho dos modelos de classificação.

3.1 Aquisição dos áudios

Nós adotamos o dataset disponibilizado em Fernandes et al. (2021), cuja origem é o Abuzz project Mukundarajan et al. (2017). O dataset possui 1285 amostras, contendo um total de 23 espécies de mosquitos. Os arquivos de áudio foram obtidos por meio de 8 modelos de celulares diferentes, como iPhone 4S e Xperia Z3 Compact. De modo que, eles foram capazes de capturar sons da batida de asas de até 50 mm de distância. Os sons foram registrados de mosquitos em diferentes circunstâncias, desde ao ar livre, como também capturados em copos/sacos Mukundarajan et al. (2017). Tendo em vista os diferentes celulares utilizados, assim como, os diferentes formatos de arquivos, é notável uma variação na frequência e na duração dos áudios, uma vez que cada espécie foi registrada por dois ou mais dispositivos. No entanto, tal variabilidade traz um cenário mais próximo do real.

3.2 Pré-processamento

Inicialmente, foi realizada a divisão das amostras do conjunto de dados em áudios de 15 segundos. Os áudios restantes, insuficientes para compor novas amostras, foram descartados. Ao fim, temos um novo conjunto de dados composto por 861 amostras, a distribuição por classe é apresentada na Tabela 2 antes e depois do processo de aumento dos dados. Com o auxílio da biblioteca LibROSA McFee et al. (2015b), as amostras foram normalizadas dentro do intervalo $[-1,1]$ e a taxa de amostragem foi alterada para 22,050 Hz, assim, essas propriedades foram padronizadas, o que é essencial para a etapa de extração de características.

A distribuição das amostras estava desbalanceada, tendo 114 amostras de *Aedes aegypti* e 747 amostras que pertenciam às demais classes. Como o design experimental prevê testes para detectar a presença do *Aedes aegypti*,



Figura 1. Etapas do método proposto. A 1ª etapa envolve a aquisição dos áudios, a 2ª etapa refere-se ao pré-processamento, incluindo uma etapa de aumento de dados, na 3ª etapa ocorre a extração de características, a 4ª etapa envolve o uso de diferentes modelos de classificação. E, a 5ª etapa compreende a validação.

foi necessário balancear as classes para esse cenário. Uma técnica de aumento foi aplicada somente nos dados de treinamento da classe *Aedes aegypti*. Tal aumento consiste na variação dos semitons $[-1.0, -1.5, -2.0, 1.0, 1.5, 2.5]$ com o auxílio do pacote MUDA McFee et al. (2015a). Desse modo, nós temos as amostras originais e mais 6 variações sintéticas, resultando em 798 amostras. Algumas amostras sintéticas foram aleatoriamente removidas e, finalmente, a classe *Aedes aegypti* ficou com 747 amostras. Comparamos os resultados de classificação binária (*Aedes Aegypti* vs. Não *Aedes Aegypti*). Tal cenário foi avaliado com e sem a abordagem de balanceamento das classes a fim de avaliar seu impacto no desempenho de classificação.

Tabela 2. Balanceamento

	Antes do balanceamento	Depois do balanceamento
Aedes	114	747
Não Aedes	747	747
Total	861	1494

3.3 Extração das características

Inspirados por Soares et al. (2022) e Silva et al. (2013), após o pré-processamento, foi realizada a extração de uma característica do domínio cepstral. Tais trabalhos

apontaram que os coeficientes cepstrais de frequência Mel (MFCCs) foram promissores no contexto de classificação de abelhas e mosquitos, respectivamente.

Os MFCCs descrevem de forma concisa a forma geral de um envelope espectral, ou seja, o limite em que o espectro do sinal está contido (Virtanen et al. (2018)). Eles são obtidos através da transformada do cosseno inversa do logaritmo da energia das bandas de frequência na escala Mel, representado pela seguinte equação:

$$mfcc(t, c) = \sqrt{\frac{2}{M_{mfcc}}} \sum_{m=1}^{M_{mfcc}} \log(\tilde{X}_m(t)) \cos\left(\frac{c(m - \frac{1}{2})\pi}{M_{mfcc}}\right) \quad (1)$$

onde M_{mfcc} é o número de bandas de frequência mel, m o índice de banda de frequência, $\tilde{X}_m(t)$ é a energia na m ésima banda de frequência mel e c é o índice do coeficiente cepstral. Neste estudo, nós extraímos esses coeficientes, compondo um descritor com 40 MFCCs. O número de 20 e 60 coeficientes também foram avaliados, no entanto, 40 MFCCs ofereceram o melhor *tradeoff* entre tamanho do descritor e desempenho.

3.4 Classificação

Para analisar o desempenho dos decritor, nós utilizamos três classificadores com diferentes características: Random Forest (RF), Support Vector Machine (SVM), XGBoost (XGB).

O Random Forest é um algoritmo baseado em árvores de decisão e menos sensível a overfitting. O SVM, geralmente apresenta bons resultados quando há poucas amostras para o treinamento em uma classificação. O XGBoost é um algoritmo gradient boosting, que combina varios modelos básicos formados por árvores de decisão em um modelo composto mais eficiente, o XGBoost destaca-se por melhorar o desempenho do gradient boosting ao otimizar a utilização do hardware.

Os parâmetros dos classificadores RF(max_depth=default) e SVM foram definidos com o algoritmo GridSearch (Bergstra and Bengio (2012)). A Tabela 3 apresenta o espaço de busca para cada parâmetro e, em negrito, o parâmetro selecionado. Já com o XGB, nós utilizamos os parâmetros default (n_estimators=100, max_depth [default=6]). Também utilizamos validação cruzada com 10 pastas nos testes.

Tabela 3. Parametros definidos pelo Grid Search

Classificadores	Intervalo dos parâmetros
RF	n_estimators = [90, 100, 110 , 120]
SVM	C = [0.1 , 0.5, 0.7, 1.0, 1.2] kernel = [linear , rbf, sigmoid, poly]

3.5 Validação

Para a validação dos resultados utilizou-se as métricas: acurácia (Equação 2), kappa (Equação 3), e F1-score (Equação 4).

O cálculo da acurácia (Acc) (Baratloo et al. (2015)) é definido pela seguinte fórmula 2:

$$Acc = \frac{VP + VN}{VP + VN + FP + FN}, \quad (2)$$

onde VP , VN , FP , e FN representam Verdadeiro Positivo, Verdadeiro Negativo, Falso Positivo e Falso Negativo, respectivamente. Quanto mais o valor se aproxima a 1, melhor é o resultado da acurácia.

A métrica kappa de Cohen (1960) representa como os classificadores selecionados se saíram melhor que o classificador que simplesmente adivinha de forma aleatória conforme a frequência de cada uma das classes. Os valores de seu índice são categorizados como: ruim ($\kappa \leq 0.2$), razoável ($0.21 \leq \kappa \leq 0.4$), bom ($0.41 \leq \kappa \leq 0.6$), muito bom ($0.61 \leq \kappa \leq 0.8$), e excelente ($\kappa \geq 0.81$). Ele é definido como:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad (3)$$

onde p_o é a probabilidade empírica de concordância no rótulo atribuído a qualquer amostra (concordância observada), e p_e é a concordância esperada quando ambos os anotadores atribuem rótulos aleatoriamente (concordância esperada).

O F1-score se trata da média harmônica entre precisão e sensibilidade (Olson and Delen (2008)). Tendo 1 como seu valor mais alto, de maneira a indicar precisão e sensibilidade perfeitos.

$$F1 - Score = \frac{VP}{VP + \frac{1}{2}(FP + FN)}. \quad (4)$$

4. RESULTADOS

As Tabelas 4 e 5 apresentam os resultados da classificação binária com 40 MFCCs como descritor de entrada nos cenários sem e com balanceamento, respectivamente. Comparamos os resultados desses cenários utilizando as três métricas de avaliação para os três classificadores adotados neste trabalho. Considerando os desvios-padrão apresentados nas tabelas, há uma interseção entre os resultados obtidos pelos diferentes classificadores, portanto, não houve diferença estatística entre eles. Porém, foi notada uma melhoria em relação aos valores médios de kappa e F1-score para ambos classificadores no cenário com balanceamento. Isso ocorre pois métricas sensíveis ao desbalanceamento são diretamente impactadas pela forma que os classificadores acertam de maneira mais equilibrada entre as classes. Dito isso, o RF foi o classificador que mais errou em relação a classe minoritária no cenário sem o balanceamento, de modo a justificar a melhoria do kappa no cenário com balanceamento. Por fim, utilizando os 40 MFCCs juntamente com o classificador XGB no cenário com balanceamento, foi possível descrever a presença do Aedes Aegypti com valores acima de 95% em todas as métricas.

Tabela 4. Resultados da classificação binária (40 MFCCs)

Classificação binária (40)	Sem balanceamento		
	RF	SVM	XGB
ACC	0,981 (+/- 0,01)	0,984 (+/- 0,01)	0,994 (+/- 0,00)
Kappa	0,912 (+/- 0,08)	0,935 (+/- 0,04)	0,978 (+/- 0,01)
F1-score	0,923 (+/- 0,06)	0,944 (+/- 0,04)	0,957 (+/- 0,04)

Tabela 5. Resultados da classificação binária (40 MFCCs)

Classificação binária (40)	Com balanceamento		
	RF	SVM	XGB
ACC	0,977 (+/- 0,02)	0,968 (+/- 0,02)	0,986 (+/- 0,00)
Kappa	0,954 (+/- 0,04)	0,937 (+/- 0,04)	0,973 (+/- 0,01)
F1-score	0,976 (+/- 0,02)	0,968 (+/- 0,02)	0,986 (+/- 0,00)

4.1 Comparação com a literatura

Fernandes et al. (2021) foi o único dos citados a utilizar da mesma base que este trabalho. Em relação a classificação binária, o mesmo treinou uma CNN com os áudios da base e realizou a extração de características, gerando um descritor de dimensão igual a 256. Enquanto que, neste trabalho, utilizamos MFCCs como descritor promovendo um vetor de dimensão 40. Considerando o melhor cenário promovido pelo classificador XGB onde o mesmo alcançou valores superiores em termos de acurácia (99%) e F1-score (95%) com um descritor de tamanho 6 vezes menor superamos uma abordagem com maior custo computacional. O tempo de extração dos 40 MFCCs foi de 17,8 ms para a base sem aumento dos dados e 22,3 ms para a base com aumento dos dados. Fernandes et al. (2021) não forneceram dados referentes ao tempo de extração. Sendo assim oferecemos um descritor mais compacto e preciso na identificação do *Aedes Aegypti*.

5. CONCLUSÃO

Neste trabalho, nós investigamos o uso de MFCCs para descrever a presença do *Aedes Aegypti*. O objetivo principal foi alcançado através da proposição de um descritor compacto, eficiente e de baixo custo computacional que auxilia na detecção de focos do mosquito. Tal descritor supera abordagens de identificação de mosquitos baseadas em aprendizado profundo.

Concluimos que a aplicação de uma etapa de aumento de dados adequada permite a utilização de classes subamostradas bem como o balanceamento entre classes, resultando em uma melhoria no desempenho de classificação. Além disso, uma descrição compacta é adequada para aplicações móveis de detecção de mosquitos, o que é importante para uma solução de monitoramento de mosquitos escalável e na produção de conjuntos de dados de alta qualidade.

Em perspectivas futuras, nós investigaremos como outras características podem enriquecer o descritor sem comprometer sua baixa dimensionalidade. Além disso, não consideramos informações de tempo e geolocalização. Por fim, nós implantaremos o descritor proposto em um dispositivo de baixo custo com poder computacional limitado e avaliaremos seu desempenho em termos de tempo computacional, precisão e tamanho dos dados.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq via Concessão No Processo: 407274/2021-9, Chamada CNPq/MCTI/FNDCT N^o 18/2021 - Faixa A - Grupos Emergentes e do CIATEN - www.ciaten.org.br.

REFERÊNCIAS

- Alar, H.S. and Fernandez, P.L. (2021). Accurate and efficient mosquito genus classification algorithm using candidate-elimination and nearest centroid on extracted features of wingbeat acoustic properties. *Computers in Biology and Medicine*, 139, 104973.
- Baratloo, A., Hosseini, M., Negida, A., and El Ashal, G. (2015). Evidence based emergency medicine; part 1: Simple definition and calculation of accuracy, sensitivity and specificity. *Emergency*, 3, 48–49.
- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37–46.
- Fernandes, M., C., W., and R.M., M. (2021). Detecting aedes aegypti mosquitoes through audio classification with convolutional neural networks. *Computers in Biology and Medicine*, 129, 104152.
- Joshi, A. and Miller, C. (2021). Review of machine learning techniques for mosquito control in urban environments. *Ecological Informatics*, 61, 101241.
- Kiskin, I., Cobb, A.D., Wang, L., and Roberts, S. (2020). Humbug zooniverse: A crowd-sourced acoustic mosquito dataset. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 916–920. IEEE.
- Kiskin, I., Orozco, B.P., Windebank, T., Zilli, D., Sinka, M., Willis, K., and Roberts, S. (2017). Mosquito detection with neural networks: the buzz of deep learning. *arXiv preprint arXiv:1705.05180*.
- Kiskin, I., Wang, L., Sinka, M., Cobb, A.D., Gutteridge, B., Zilli, D., Rafique, W., Dam, R., Marinos, T., Li, Y., Killeen, G., Msaky, D., Kaindoa, E., Willis, K., and Roberts, S.J. (2021). Humbugdb: a large-scale acoustic mosquito dataset.
- McFee, B., Humphrey, E.J., and Bello, J.P. (2015a). A software framework for musical data augmentation. In *ISMIR*, volume 2015, 248–254.
- McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., and Nieto, O. (2015b). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8.
- Mukundarajan, H., Hol, F.J.H., Castillo, E.A., Newby, C., and Prakash, M. (2017). Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. *elife*, 6, e27854.
- Mulchandani, P., Siddiqui, M.U., Kanani, P., et al. (2019). Real-time mosquito species identification using deep learning techniques. *International Journal of Engineering and Advanced Technology*, 2249–8958.
- Okawa, M., Saito, T., Sawada, N., and Nishizaki, H. (2019). Audio classification of bit-representation waveform. *arXiv preprint arXiv:1904.04364*.
- Olson, D.L. and Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.
- Sengül Demirak, M.Ş. and Canpolat, E. (2022). Plant-based bioinsecticides for mosquito control: Impact on insecticide resistance and disease transmission. *Insects*, 13(2), 162.
- Silva, D.F., De Souza, V.M., Batista, G.E., Keogh, E., and Ellis, D.P. (2013). Applying machine learning

and audio analysis techniques to insect recognition in intelligent traps. In *2013 12th International Conference on Machine Learning and Applications*, volume 1, 99–104. IEEE.

Soares, B.S., Luz, J.S., de Macêdo, V.F., e Silva, R.R.V., de Araújo, F.H.D., and Magalhães, D.M.V. (2022). Mfcc-based descriptor for bee queen presence detection. *Expert Systems with Applications*, 117104.

Virtanen, T., Plumbley, M.D., and Ellis, D. (2018). *Computational analysis of sound scenes and events*. Springer.