

Universidade de São Paulo

Faculdade de Medicina

VINICIO RODRIGUES DE LIMA

**Inteligência artificial para classificação de espécies de culicídeos baseada em
morfometria de asa**

São Paulo

2023

VINICIO RODRIGUES DE LIMA

**Inteligência artificial para classificação de espécies de culicídeos baseada em
morfometria de asa**

Dissertação apresentada à
Faculdade de Medicina da
Universidade de São Paulo para
obtenção do título de Mestre em
Ciências

Programa Medicina Tropical

Área de Concentração: Doenças
Tropicais e Saúde Internacional

Orientadora: Profa. Dra. Karin
Kirchgatter

Coorientador: Dr. Mauro César
Cafundó de Morais

(Versão Corrigida. Resolução CoPGr 6018, de 13 de outubro de 2011.

A versão original está disponível na Biblioteca da FMUSP)

São Paulo

2023

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Lima, Vinicio Rodrigues de
Inteligência artificial para classificação de
espécies de culicídeos baseada em morfometria de asa
/ Vinicio Rodrigues de Lima. -- São Paulo, 2023.
Dissertação (mestrado)--Faculdade de Medicina da
Universidade de São Paulo.
Programa de Medicina Tropical. Área de
Concentração: Doenças Tropicais e Saúde Internacional.
Orientadora: Karin Kirchgatter.
Coorientador: Mauro César Cafundó de Moraes.

Descritores: 1.Morfometria alar 2.Processamento
de imagem assistida por computador. 3.Aprendizado
de máquina 4.Inteligência artificial 5.Culicidae
6.Biologia computacional

USP/FM/DBD-459/23

Responsável: Erinalva da Conceição Batista, CRB-8 6755

**Aos meus pais, que sempre
apoiaram meus sonhos,
e ao meu avô, que me
ensinou a sonhar.**

AGRADECIMENTOS

A realização deste trabalho contou com o suporte de diversos profissionais, colegas, familiares e amigos para que pudesse ser concluído.

Agradeço principalmente à orientadora deste Mestrado, a Profa. Dra. Karin Kirchgatter por não apenas aceitar e apoiar as minhas propostas, quase nunca muito convencionais, e sempre muito mais difíceis e trabalhosas do que a ingenuidade de um aluno consegue compreender, e mesmo assim acreditar no meu potencial de completar essas tarefas, mesmo que meus processos não façam muito sentido no início e pareçam um tanto caóticos. É valioso que em tempos de hiper-produção acadêmica, o processo criativo-científico ainda seja estimulado por uma pesquisadora cujo ela mesma e seus alunos sejam referência em seus campos de atuação no Brasil e no Mundo.

Agradecimentos à minha família e amigos:

- Agradeço à minha mãe e ao meu pai, Aparecida Maria Rodrigues de Lima e Waldemir de Lima, por sempre apoiarem meu sonho de me tornar um cientista.
- Agradeço aos meus avós Ana Gimenes de Lima†, Araci Maria da Silva Rodrigues e Fortunato Martins Rodrigues† pelo carinho e apoio, principalmente ao meu avô, que aprendeu todos os seus vários ofícios sozinho, e mesmo não-alfabetizado de forma tradicional, sempre me ensinou o valor de ser curioso e aprender sempre.
- Agradeço à minha namorada, Larissa Tiburcio Correia por sempre acreditar em mim, por todo amor e alegrias que me proporciona, e pela paciência nas várias ocasiões em que passou ao meu lado, absorto em livros, artigos e códigos de programação.
- Agradeço a todos os meus tios e primos, por compreender as minhas ausências nos momentos necessários, e principalmente ao meu primo MSc. Itamar, por dividir as angústias e alegrias da vida acadêmica, e pelo compartilhamento de histórias, experiências e ideias.
- Agradeço aos meus amigos Vitor, Arthur, Débora, Wellington, Everton, Thais Rangel, Eduardo, Tais Pereira e Giorgio, pelos momentos de leveza e sorrisos.
- Agradeço também aos meus amigos e colegas de trabalho MSc. Bruno Mathias da Silva, MSc. Carolina Clares dos Anjos e Med. Vet. Gabriella Ricomini Caetano Clemente pela ajuda com a montagem das asas envolvidas nesse projetos e as outras milhares que ainda serão utilizadas, e por todo apoio dentro e principalmente fora do laboratório.

Agradecimento aos colegas de trabalho:

- À Dra. Lilian de Oliveira Guimarães e à Dra. Eliana Ferreira Monteiro pelos ensinamentos desde o primeiro dia no grupo de pesquisas.
- Aos “estagiários” Bruna Antonetti Damato, Jesus Maia e Gabriel de Freitas Ferrari Rodrigues por toda prestatividade que alguém pode desejar.
- À Biol. Vanessa, Dra. Juliana, Dra. Regiane e Dra. Rosa Tubaki, que me ensinaram tudo que eu sei sobre coleta, dissecação e identificação de mosquitos.
- À Biomed. Vanessa dos Santos Morais pela presença sempre alegre.
- Ao Dr. Adriano Pinter por sempre manter a leveza e funcionalidade dos laboratórios, e ao Dr. Thiago pela presteza e genialidade em ensinar sempre da melhor forma possível.

Agradecimento aos Colaboradores:

- Ao Dr. Mauro César Cafundó de Morais, coorientador deste projeto, por compartilhar dos seus conhecimentos em processamento de imagem, e esforço em aprender o lado biológico do trabalho.
- Ao Dr. André Bruno Barreto Wilke, que me ensinou tudo que eu sei sobre Morfometria Geométrica Alar. Bem como pela contribuição com imagens de asas de mosquitos.
- Ao Prof. Dr. Lincoln e ao Prof. Dr. Roberto pelas críticas e sugestões de extrema importância no exame de qualificação, que possibilitaram o avanço e desenvolvimento melhor direcionado da última fase do projeto.
- Ao Prof. Dr. Mauro Marrelli, por ceder imagens de asas de mosquitos de projetos seus, e dos seus alunos André Bruno Barreto Wilke e Ana Letícia.
- Ao MSc. Matheus Barros Pereira por ter me ajudado a corrigir um código em Python crucial para a conclusão deste projeto, e por toda paciência e prestatividade que teve em nosso breve contato.

Fortis Fortuna Adiuvat

RESUMO

Lima VR. Inteligência artificial para classificação de espécies de culicídeos baseada em morfometria de asa [dissertação]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2023.

Os culicídeos formam um grupo numeroso e cosmopolita, presentes em toda a região tropical e temperada do planeta. Os gêneros *Aedes*, *Anopheles* e *Culex* se destacam dentro da família dos culicídeos pela sua importância médica, sendo responsáveis pela transmissão de vários patógenos que causam doenças como febre do Nilo, dengue e malária. A forma mais comum de identificação desses insetos é feita com base nos seus caracteres morfológicos de exemplares adultos, junto de chaves taxonômicas. Entretanto, a sistemática é limitada a poucas diferenças anatômicas entre algumas espécies e gêneros, sendo que em algumas espécies, a diferenciação dos indivíduos adultos se restringe a identificação do órgão sexual masculino. Este conjunto de características podem se perder até chegar às mãos de um taxonomista qualificado para realizar essa identificação da forma precisa. Uma vez que se obtém êxito em identificar corretamente espécies de mosquitos, podemos aplicar a Morfometria Geométrica Alar (WGM) como técnica complementar. Uma inovação que tem crescido dentro do campo da WGM é o uso de inteligência artificial. Podemos citar a aplicação de algoritmos de aprendizado de máquina, tais como redes neurais artificiais (ANN - Artificial Neural Network) capazes de realizar o reconhecimento de padrões em imagens e classificar as amostras baseando-se nas suas características morfométricas. Isso demonstra que a área de WGM pode ser beneficiada pelo uso de ferramentas de processamento de imagens digitais e inteligência artificial. Assim, esse trabalho teve como objetivo desenvolver um modelo de inteligência artificial para classificar espécies de culicídeos através de morfometria de asa. Um total de 180 imagens de asas de 6 espécies de culicídeos dos gêneros *Culex*, *Aedes*, *Wyeomyia* e *Anopheles* foram utilizadas. Nessas imagens anotamos manualmente as coordenadas dos pontos anatômicos usados na morfometria para a identificação de espécies. Para a obtenção automática dos pontos anatômicos, elaboramos um algoritmo na linguagem Python para o processamento das imagens digitais. Em linhas gerais, as veias alares foram segmentadas e as coordenadas dos pontos presentes nas linhas foram obtidas após “afinar” as veias a um pixel de espessura. Os marcos anatômicos previamente marcados foram integrados ao processamento com Procrustes gerados a partir do pacote Geomorph, em R. Utilizando filtros morfológicos para filtrar o número de pontos previstos, e distância de Mahalanobis, comparando com os TPS anotados dos espécimes e Procrustes de cada espécie, mostramos ser possível a determinação da similaridade de um espécime com as espécies presentes no banco.

Palavras-chave: Morfometria alar, Processamento de imagem assistida por computador, Aprendizado de máquina, Inteligência artificial, Culicidae, Biologia computacional.

ABSTRACT

Lima VR. Artificial intelligence for classifying culicidae species based on wing morphometry [dissertation]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2023.

Culicids form a numerous and cosmopolitan group, present in tropical and temperate regions worldwide. The genera *Aedes*, *Anopheles*, and *Culex* stand out within the Culicidae family due to their medical importance, as they are responsible for transmitting various pathogens responsible for diseases such as West Nile fever, dengue, and malaria. The most common method of identifying these insects is based on the morphological characteristics of adult specimens, along with taxonomic keys. However, systematic identification is limited to a few anatomical differences between some species and genera. In some species, differentiation of adult individuals is restricted to the identification of the male genital organ. This set of characteristics can be lost until it reaches the hands of a qualified taxonomist capable of making accurate identifications. Once successful in correctly identifying mosquito species, we can apply Wing Geometric Morphometry (WGM) as a complementary technique. An innovation that has grown within the field of WGM is the use of artificial intelligence. We can mention the application of machine learning algorithms, such as artificial neural networks (ANN), capable of pattern recognition in images and classifying samples based on their morphometric characteristics. This demonstrates that the field of WGM can benefit from the use of digital image processing tools and artificial intelligence. Therefore, this study aimed to develop an artificial intelligence model to classify Culicidae species through wing morphometry. A total of 180 wing images from 6 species of Culicidae from the genera *Culex*, *Aedes*, *Wyeomyia* and *Anopheles* were used. In these images, we manually annotated the coordinates of the anatomical points used in morphometry for species identification. To automatically obtain anatomical points, we developed an algorithm in Python for processing digital images. In general, wing veins were segmented, and the coordinates of the points present on the veins were obtained after "thinning" the veins to one-pixel thickness. The previously marked anatomical *landmarks* were integrated into the processing with Procrustes generated from the Geomorph package in R. Using morphological filters to filter the number of predicted points and Mahalanobis distance, compared to the annotated TPS of specimens and Procrustes of each species, we demonstrated the possibility of determining the similarity of a specimen to the species in the database.

Keywords: Wing morphometry. Computer-assisted image processing. Machine learning. Artificial intelligence. Culicidae. Computational biology.

LISTA DE ILUSTRAÇÕES

Figura 1. Fluxograma simplificado das três principais técnicas usadas para identificação de espécies de mosquitos.....	16
Figura 2. Asa direita de <i>Culex quinquefasciatus</i> com os 18 landmarks mais utilizados em WGM, dispostos numericamente.....	20
Figura 3. Relações entre os vários níveis presentes em Inteligência Artificial.....	21
Figura 4. Trecho de código responsável pela padronização inicial das imagens.....	31
Figura 5. Processo de segmentação, seguido de binarização.....	32
Figura 6. trecho de código responsável pelo treinamento do segmentador em Random Forest.....	32
Figura 7. Transformação de uma imagem binarizada em uma imagem esqueletonizada.....	33
Figura 8. Exemplo das imagens capturadas via microscopia, de cada uma das espécies que passaram pelo processamento, e foram analisadas neste trabalho...	36
Figura 9. Asas de <i>Culex quinquefasciatus</i> e asas de <i>Culex nigripalpus</i> em lâmina para microscopia, segmentação e esqueletonização.....	37
Figura 10. Média de pontos preditos por espécie na esqueletonização das imagens segmentadas das asas.....	38
Figura 11. Porcentagem da redução no nº de pontos por filtro morfológico.....	39
Figura 12. Asa de <i>Culex quinquefasciatus</i> ao lado da esqueletonização da imagem, tratada previamente com os 4 filtros morfológicos testados, e sobreposição na imagem pré-processada da asa, contendo <i>Ground-truth</i> , e os pontos preditos nos dois tratamentos mais discrepantes.....	41
Figura 13. Asa direita de <i>Culex quinquefasciatus</i> e sobreposição dos pontos detectados na venação sobre a imagem da asa.....	42
Figura 14. Foto de asa de <i>Aedes aegypti</i> sobreposta pelos pontos preditos e pontos ground-truth.....	45
Figura 15. Morfoespaço gerado pelo cálculo da Variação Canônica entre as amostragens selecionadas para o estudo.....	46

LISTA DE TABELAS

Tabela 1. Espécies utilizadas neste estudo, e quantidade de asas de cada uma.... **29**

Tabela 2. Distância Média Euclidiana entre os pontos mais próximos detectados no esqueleto da asa da amostra L 37_D0071, comparada com o banco das espécies selecionadas..... **43**

Glossário

Script	Sequência de comandos dada ao computador, via linguagens de programação, como R e Python.
KAARMA	Do inglês <i>kernel adaptive autoregressive-moving average</i> . Tipo de algoritmo que ajuda o computador a reconhecer padrões em sons, baseado em modelos auto-regressivos de médias móveis.
Redes Neurais Convolucionais	CNN, do inglês, <i>Convolutional Neural Networks</i> , é uma arquitetura de redes neurais projetadas para processar e analisar dados de grade, como imagens. Elas são eficazes na extração de características hierárquicas por meio de camadas convolucionais, permitindo o reconhecimento de padrões complexos.
GoogleNet	Arquitetura de rede neural profunda desenvolvida pela Google, notável por sua complexidade e eficácia em tarefas de reconhecimento de padrões em imagens.
ResNet50	Arquitetura de rede neural profunda com 50 camadas, normalmente mais eficaz em comparação com arquiteturas mais rasas.
VGG16	Arquitetura de rede neural convolucional que se destaca por ter 16 camadas, incluindo camadas convolucionais e totalmente conectadas.
VGG19	Arquitetura de rede neural convolucional que se assemelha à VGG16, mas possui 19 camadas, incluindo camadas convolucionais e totalmente conectadas.
AlexNet	Arquitetura pioneira de rede neural convolucional, projetada para competir no Desafio de Reconhecimento de Imagens do ImageNet em 2012. Desenvolvida por Alex Krizhevsky, ela foi uma das primeiras CNNs profundas a obter sucesso em tarefas de visão computacional em grande escala.
YOLO	Do inglês, <i>You Only Look Once</i> . Abordagem de detecção de objetos em imagens que permite a identificação de múltiplos objetos em uma única passagem pela rede neural. Essa técnica é conhecida por sua eficiência e rapidez, sendo amplamente utilizada em tarefas de visão computacional, como reconhecimento de objetos em tempo real.
Key Point Detection	Refere-se à identificação e localização de pontos-chave em uma imagem. Esses pontos representam características específicas e distintas, e a detecção deles é fundamental em diversas aplicações, como reconhecimento de objetos, rastreamento de movimentos e análise de imagem.
Inteligência Artificial	AI, do inglês, <i>Artificial Intelligence</i> , refere-se à simulação da inteligência humana em máquinas.
Aprendizado de Máquina	ML, do inglês, <i>Machine Learning</i> , é um ramo da AI que permite que computadores aprendam e tomem decisões sem programação explícita, através de algoritmos e modelos estatísticos.
Redes Neurais Artificiais	ANN, do inglês, <i>Artificial Neural Networks</i> , é um componente fundamental do aprendizado profundo, inspirado na estrutura e função do cérebro humano.
Deep Learning	DL, do inglês, <i>Deep Learning</i> , DL é um ramo especializado do ML que usa redes neurais artificiais para modelar e resolver problemas complexos.

Sumário

1. INTRODUÇÃO.....	14
1.1 Morfometria Geométrica.....	17
1.1.1 Morfometria Geométrica Alar.....	18
1.2 Inteligência Artificial e Aprendizado de Máquina.....	21
1.2.1 Testes supervisionados e não supervisionados.....	22
1.2.2 Interpretabilidade e Explicabilidade.....	23
1.2.3 Aplicabilidade dos modelos.....	23
1.3 WGM e Inteligência Artificial.....	24
2. OBJETIVOS.....	28
2.1 Objetivo Geral:.....	28
2.2 Objetivos Específicos:.....	28
3. MATERIAL E MÉTODOS.....	29
3.1 Origem das imagens.....	29
3.2 Padronização da aquisição de imagens digitais das asas.....	29
3.2.1 Morfometria geométrica alar.....	30
3.3 Extração dos pontos anatômicos para morfometria alar.....	30
3.3.1 Processamento de imagens.....	31
3.3.2 Esqueletonização dos segmentos das veias alares.....	33
3.3.3 Obtenção dos pontos para WGM.....	34
4. RESULTADOS.....	36
4.1 Modelo de Segmentação e Esqueletonização.....	36
4.1.1. Filtros morfológicos.....	38
4.2 Definindo os pontos mais próximos.....	40
4.3 Classificação pela Menor Distância Média.....	42
5. DISCUSSÃO.....	44
6. CONCLUSÃO.....	48
REFERÊNCIAS.....	49
ANEXOS.....	54
Anexo 1. Autorização CEUA.....	54
Anexo 2. Ficha do Aluno Janus.....	54
Anexo 3. Primeira página do artigo publicado.....	57

1. INTRODUÇÃO

Os mosquitos (ordem Diptera e família Culicidae) configuram um grupo com mais de 3.500 espécies descritas (Orlandin *et al.*, 2016; Harbach 2023), presentes em toda a região tropical e temperada do planeta. São encontrados por toda estratificação vertical dentro dos ambientes que habitam, coabitando com populações humanas em meio rural e urbano. Entretanto, seu desenvolvimento e maior diversidade são proeminentes em regiões de floresta tropical (Harbach 2007).

A estratificação vertical de mosquitos em ambiente natural depende diretamente dos hábitos alimentares de cada espécie e da disponibilidade de alimentos. Os machos de algumas espécies se alimentam de fontes líquidas vegetais como seivas, néctar e frutas. Entretanto, na maior parte das espécies, as fêmeas necessitam de aporte sanguíneo proveniente de vertebrados hoemotérmicos. Na indisponibilidade de sangue quente, outros animais de sangue frio como cobras, anfíbios e até mesmo helmintos e anelídeos podem servir como fonte de repasto para completar a maturação de seus ovos (Harbach 2007). Esse hábito hematófago torna os culicídeos importantes vetores de arbovírus e outros patógenos transmitidos pela picada desses insetos (Harbach 2023).

A identificação de espécies de relevância médica contribui para uma maior compreensão das características biológicas desses insetos vetores de doenças. Esses insetos frequentemente enfrentam várias formas de pressão seletiva, incluindo infecções parasitárias, mudanças no ambiente e paisagem, uso urbano e agrícola de inseticidas sobre as populações de insetos, entre outros (Dujardin, 2008). Muitas espécies de mosquitos, apesar de semelhantes, possuem biologia única, e interpretar como determinada população interage com o ambiente onde está inserida, avaliar o estresse que sofre durante seu desenvolvimento e mapear ao longo do tempo é fundamental (Mpho *et al.*, 2002).

A identificação dos mosquitos é feita majoritariamente com a utilização de duas técnicas, sendo elas a taxonomia clássica e taxonomia molecular (Camargo-Neves *et al.*, 2021).

A identificação morfológica dos mosquitos é feita com base em chaves dicotômicas, utilizando como referência caracteres previamente descritos como exclusivos de determinados grupos ou espécies, até que se chegue a um gênero,

ou mesmo espécie. Também é possível identificar sub-espécies desses animais, quando há preservação das amostras e a *expertise* do profissional responsável pela identificação permitem (Consoli & Lourenço de Oliveira, 1994).

As técnicas de biologia molecular, por outro lado, comparam sequências de DNA, podendo ser utilizadas em diferentes estágios de vida do inseto e com pequenas partes do mesmo. A principal abordagem molecular para esse propósito é o DNA barcode (Hebert *et al.*, 2003), envolvendo a utilização de uma sequência do gene citocromo c oxidase subunidade I (COI), também conhecido como COX1, presente no genoma mitocondrial (mtDNA) de vertebrados e invertebrados. Em mosquitos e em muitos outros organismos, o fragmento mais comumente empregado consiste em 658 pares de bases, que é a região delimitada pelos oligonucleotídeos LCO1490 e HCO2198 (Folmer *et al.*, 1994). A correlação genética entre as sequências de DNA dos indivíduos estudados e aquelas dos indivíduos presentes nos bancos de dados, tais como o GenBank e a aplicação BOLD, são usadas para validar a identificação (Hajibabaei *et al.*, 2006; Renaud *et al.*, 2012; Muhammad Tahir & Akhtar, 2015).

As duas abordagens para identificação de espécies (taxonomia morfológica e taxonomia molecular) têm sua importância, sendo usadas dentro da academia e da saúde pública, mas possuem as suas próprias desvantagens. A taxonomia clássica não exige um aparato laboratorial caro e complexo como o necessário para a identificação molecular (**Figuras 1A e 1B**), mas exige profissionais com *expertise*, que atualmente são raros na entomologia médica no nosso país e com pouca disponibilidade (Krieger *et al.*, 2002).

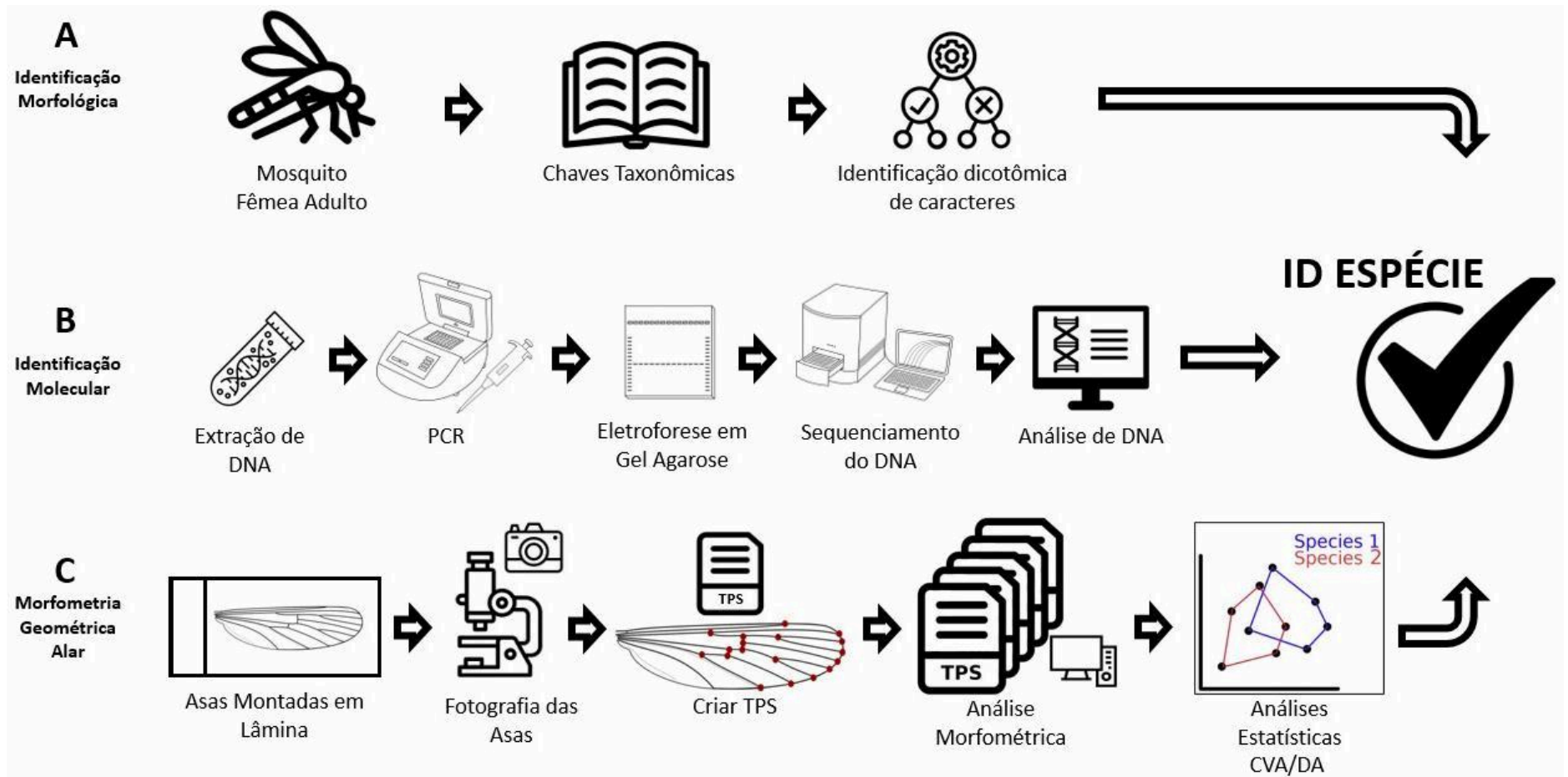


Figura 1. Fluxograma simplificado das três principais técnicas usadas para identificação de espécies de mosquitos. **A.** Identificação morfológica utilizando chaves taxonômicas. **B.** Identificação molecular com amplificação e sequenciamento de genes alvos. **C.** Morfometria Geométrica Alar.

Nos anos recentes, as técnicas de morfometria estão sendo cada vez mais aplicadas como forma de identificação complementar, visando contornar as dificuldades envolvidas nas abordagens descritas acima. Por se tratar de um conjunto de técnicas matemáticas, fica cada vez mais evidente que ferramentas computacionais sejam agregadas aos processos de morfometria, com a finalidade de automatizar esses processos. Neste trabalho abordamos como a Inteligência Artificial (AI, do inglês *Artificial Intelligence*) pode auxiliar na identificação de espécies de mosquitos juntamente com a morfometria geométrica alar, e como essa combinação tem sido utilizada em alguns insetos. Exploramos como a rápida identificação de espécies usando AI pode contribuir no monitoramento e, finalmente, controle das populações de mosquitos. Por fim, apresentamos perspectivas futuras onde aplicações desse campo de pesquisa e desenvolvimento poderão minimizar a crescente deficiência de taxonomistas em entomologia.

1.1 Morfometria Geométrica

Morfometria se refere ao estudo da forma e tamanho, e relação entre essas características (Moraes, 2003). A morfometria geométrica combina princípios geométricos com análise estatística multivariada, garantindo a caracterização da forma de uma estrutura (Klingenberg & Marugán-Lóbon, 2013). Este método tem sido empregado tanto para diferenciar espécies quanto para padronizar comparativos microevolutivos e diferenças populacionais, devido às suas características herdáveis que podem ter relevância em processos adaptativos e evolutivos (Dujardin, 2008). Assim, a análise da forma possibilita a comparação de fenótipos entre indivíduos, populações ou espécies, contanto que exista correspondência entre os indivíduos e os pontos de referência analisados em uma determinada estrutura anatômica (Dujardin, 2011).

A morfometria se mostra extremamente útil, quando necessitamos de ferramentas objetivas para discriminação da forma, e reconhecimento de padrões, como por exemplo, em artefatos arqueológicos (Cardillo, 2010), onde a técnica transforma drasticamente a análise dos artefatos mais pragmáticos da arqueologia, os projéteis bifaciais (pontas de flechas, dardos entre outros, em pedra e outros minerais), que revelam informações cruciais sobre a cultura dos povos que as produziram, e das técnicas empregadas na sua confecção.

Na paleontologia, uma das aplicações mais conhecidas pelo público geral, é a distinção de ossadas masculinas e femininas pelo tamanho e formato dos ossos formadores da pélvis, possibilitando inclusive a comparação da pélvis feminina do período neolítico, com as mulheres modernas (Raia-Barjat *et al.*, 2011).

Ainda na paleontologia, a Morfometria Geométrica também é uma ferramenta extremamente útil, combinada com outras técnicas, para a análise anatômica comparada de animais extintos, como por exemplo o estudo sobre *Brachiosaurus* (Bonnan 2007), permitindo analisar as qualidades motoras de grandes animais e a relação entre ossos longos e características musculares.

Dessa forma, podemos assumir que a Morfometria Geométrica é uma ferramenta poderosa para auxiliar a análise e comparação de estruturas homólogas e análogas, utilizando a matemática e estatística para correlacioná-las entre si.

A aplicação computadorizada da análise dos dados, com a definição de modelos matemáticos apropriados para a interpretação de pontos anatômicos levou ao surgimento da morfometria geométrica multivariada redundante. A medição da forma aplicada à estatística é uma técnica há muito utilizada em vários campos das ciências médicas e biológicas para o estudo da variedade morfológica dos mais diversos objetos de interesse (Mitteroecker & Gunz, 2009; Lorenz *et al.* 2017).

1.1.1 Morfometria Geométrica Alar

A Morfometria Geométrica Alar (WGM, do inglês *Wing Geometric Morphometrics*) é a aplicação da técnica de Morfometria Geométrica nas estruturas alares geralmente de insetos dípteros, como mosquitos. A abordagem utiliza imagens digitais de uma única asa por indivíduo (Lorenz *et al.*, 2017). Nas últimas décadas, o número de estudos realizados com base em WGM tem crescido e consolidado o uso dessa ferramenta versátil e de baixo custo, além de amigável para usuários menos experientes (Lorenz *et al.*, 2017). A WGM demonstrou eficácia na identificação precisa de gêneros e espécies de mosquitos, equiparando-se à taxonomia clássica e à biologia molecular (Lorenz *et al.*, 2012; Wilke *et al.*, 2016; Simões *et al.*, 2020). Além da identificação de mosquitos com a análise de variações interespecíficas, a WGM tem se mostrado uma ferramenta capaz de possibilitar a distinção entre variações intraespecíficas, como diferentes populações de uma mesma espécie e processos microevolutivos, pela característica hereditária das

venações alares (Vidal & Suesdek, 2012; Lorenz & Suesdek, 2020). Além disso, a WGM permite avaliações comparativas do dimorfismo sexual nas asas de culicídeos de relevância médica e veterinária (Virginio *et al.*, 2015; Christe *et al.*, 2016). A análise da variação na forma das asas também se mostrou capaz de detectar variações no fenótipo de populações pressionadas pelo avanço descontrolado da urbanização em áreas de oviposição (Multini *et al.*, 2019).

Além da identificação de espécimes por meio da Morfometria da Asa, WGM fornece informações suficientes para detecção de diversas interferências no desenvolvimento do inseto por meio da medida da assimetria bilateral das asas (Assimetria Flutuante), como stress gerado pelo contato com agentes químicos, flutuações de temperatura, e até mesmo a virulência de certos parasitas (Agnew 1997; Mpho *et al.*, 2002).

A Morfometria Geométrica Alar representa uma abordagem para estudar insetos de interesse médico, funcionando como complemento às técnicas já utilizadas, com a vantagem de depender apenas das asas do inseto, estruturas menos complexas, que costumam se preservar nas coletas (Calle *et al.*, 2002; Wilke *et al.*, 2016). Nas asas, os pontos anatômicos de interesse são denominados *landmarks* e estão presentes nas intersecções entre as venações que ocorrem dentro e na borda das asas, representando 24 intersecções, homólogas nas asas de todos os mosquitos (Diptera: Culicidae). Visando a universalização e padronização dos experimentos, são comumente utilizados 18 *landmarks* (**Figura 2**) uma vez que a localização de 6 dos 24 pontos é trabalhosa, e, frequentemente se perde durante a retirada da asa do animal. Além disso, os 18 *landmarks* (ou até menos) já foram provados pela literatura, suficientes para os procedimentos estatísticos englobados pela morfometria (Lorenz *et al.*, 2017).

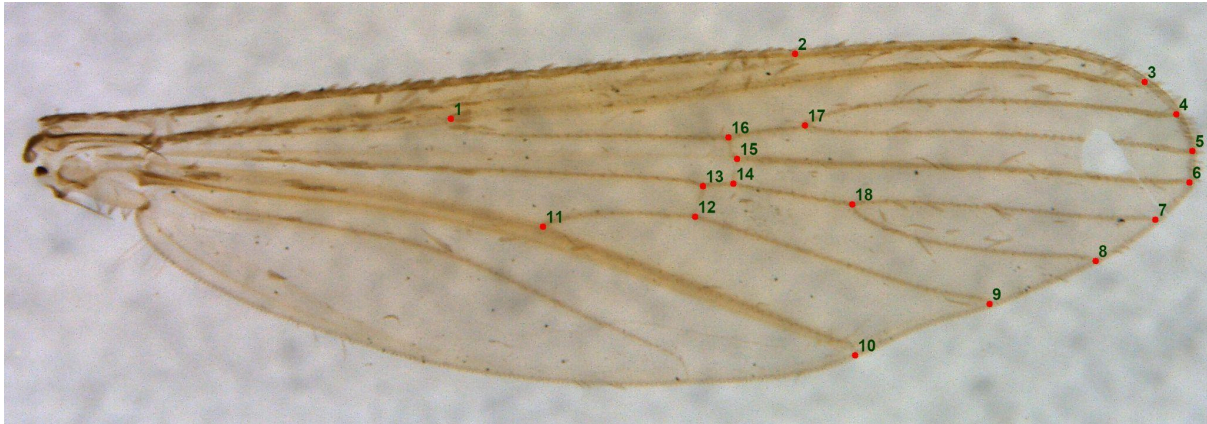


Figura 2. Asa direita de *Culex quinquefasciatus* com os 18 *landmarks* mais utilizados em WGM, dispostos numericamente. Fonte: Vinício Rodrigues Lima.

Para a realização da WGM, as asas dos espécimes adultos são montadas em lâmina e lamínula, então fotografadas e os *landmarks* são marcados utilizando, por exemplo, o software TPSDig (Rohlf, 2006; Wilke *et al.*, 2016; Simões *et al.*, 2020), gerando arquivos do tipo *thin-plate spline* (sigla: TPS) que apresentam uma matriz com os valores numéricos das coordenadas dos *landmarks* na imagem. Outros *softwares* e ferramentas como ImageJ (Schneider *et al.*, 2012) também servem a esse propósito, enquanto ferramentas como o MorphoJ (Klingenberg, 2011) e o Geomorph (Adams & Otárola-Castillo, 2013) são utilizadas para a realização de todos os cálculos estatísticos envolvendo os *landmarks* (**Figura 1C**).

Além das ferramentas computacionais amplamente utilizadas em WGM citadas anteriormente, o estudo da morfologia dos mosquitos abre possibilidades para criação de novas ferramentas que visem a automatização e padronização dos processos que ainda hoje são manuais. Isto porque algumas barreiras e limitações metodológicas ainda impedem que WGM se difunda mais rápido como uma ferramenta padrão na identificação de mosquitos. Por exemplo, quando se trata da marcação dos *landmarks*, ainda não existe consenso na comunidade científica. Alguns autores sugerem que mesmo diferentes operadores leigos são capazes de localizar os *landmarks* com alta precisão (Lorenz & Suesdek, 2013), enquanto outros sustentam que a marcação dos *landmarks* em todas as asas do banco de dados deve ser realizada por uma única pessoa, como medida de precaução para reduzir qualquer viés na marcação (Simões *et al.*, 2020). No entanto, essa prática entra em conflito com a necessidade de construir bancos de dados globais, públicos, acessíveis a múltiplos usuários, continuamente atualizados e

colaborativos, conforme proposto pela criação e implementação do WingBank (Virginio *et al.*, 2021). Portanto, vários esforços têm sido feitos ao longo dos últimos anos e cada vez um número maior de técnicas de AI se mostram como uma possibilidade na criação de alternativas para que essas limitações sejam superadas.

1.2 Inteligência Artificial e Aprendizado de Máquina

De acordo com Chollet (2021), pode-se afirmar que a Inteligência Artificial é um amplo campo do conhecimento que se originou na década de 1950. Foi resultado dos esforços no âmbito das Ciências da Computação em busca de respostas para a questão se os computadores têm ou não a capacidade de pensar. Conforme Alan Turing (1950), os computadores podem ser capazes de emular qualquer aspecto da inteligência humana.

Com a percepção de que os computadores não apenas podem realizar tarefas intelectuais, mas também têm a capacidade de aprender tarefas e padrões por si mesmos, mediante treinamento e programação adequados, outras técnicas no amplo campo do conhecimento da Inteligência Artificial começaram a surgir, incluindo o que foi denominado na década de 1980 como Aprendizado de Máquina (ML, do inglês *Machine Learning*) (Chollet, 2021). Dentro do campo de ML, surgem técnicas e ferramentas que se destacam, como as Redes Neurais Artificiais (ANNs) e o Aprendizado Profundo (Deep Learning - DL) (**Figura 3**).

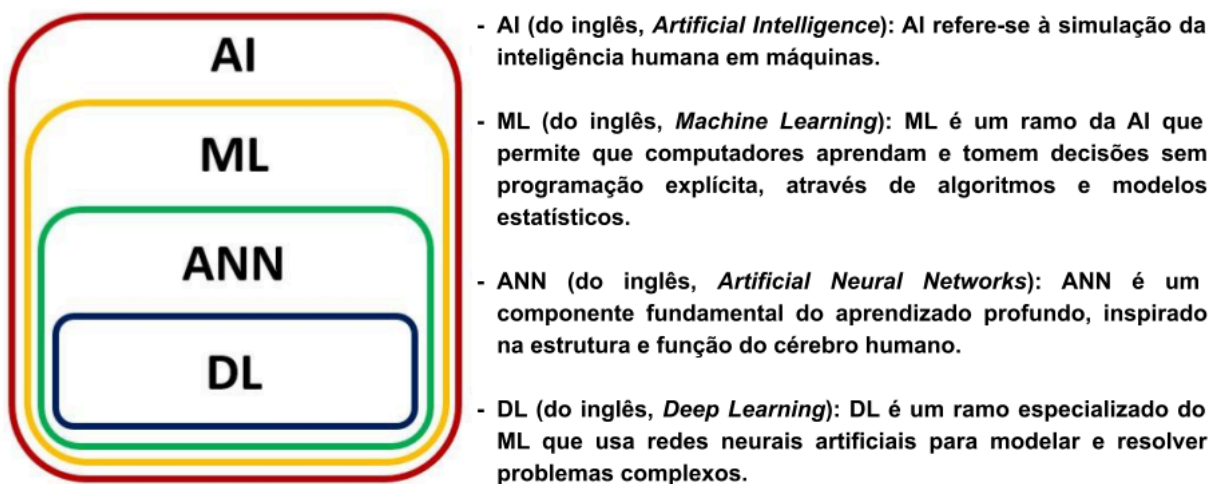


Figura 3. Relações entre os vários níveis e arquiteturas presentes em Inteligência Artificial.

O Aprendizado Profundo é uma subcategoria do Aprendizado de Máquina, caracterizada por envolver dezenas ou até mesmo centenas de camadas hierárquicas de parâmetros no processo de aprendizado. Isso permite o reconhecimento de padrões complexos nos dados, e as 'respostas' são obtidas através de Redes Neurais Artificiais (Krizhevsky *et al.*, 2017).

As ANNs são sistemas computacionais que se inspiram no funcionamento e organização da estrutura neural encontrada em organismos inteligentes. Elas têm a capacidade de aprender a partir da experiência. Uma Rede Neural é composta por diversas unidades de processamento organizadas em camadas: a camada de entrada, as camadas intermediárias e a camada de saída. Cada unidade de processamento em uma camada se conecta com todas as unidades da camada subsequente, e essas conexões individuais possuem pesos associados, que funcionam como valores limitantes para a validação de cada característica (feature) durante o processamento (LeCun *et al.*, 2015).

Quando implementamos um modelo de ANN como parte de um algoritmo de ML, o resultado é um modelo treinado para executar uma tarefa específica, o que caracteriza o Aprendizado Profundo (*Deep Learning* - DL). Semelhante ao ML tradicional, o DL se beneficia de grandes volumes de dados de entrada e requer ajustes adequados nas camadas intermediárias, incluindo limites para os pesos, o número de camadas e as unidades de processamento em cada camada. Uma das principais vantagens das ANNs é que, mesmo em caso de falhas em determinadas etapas, devido à natureza paralela de seus processos, o processamento continua sem interrupções (Goodfellow *et al.*, 2016).

Depois do algoritmo finalizado e treinado, as ML precisam ser validadas. É necessário saber se o algoritmo está apenas reproduzindo o que foi ensinado, ou se ele realmente gerou informações novas, se de fato “aprendeu” a realizar a tarefa proposta. A validação é realizada por meio de um *dataset* preparado especificamente para esta tarefa, sendo executado em comparação aos resultados simulados. É comum vermos trabalhos que separam algo em torno de 70% dos dados para treinamento do algoritmo, e 30% como *dataset* para validação do mesmo.

1.2.1 Testes supervisionados e não supervisionados

Dentro da grande área da AI, e mais especificamente da ML, existem duas abordagens relativas à metodologia aplicada no treinamento dos algoritmos, e essas abordagens são os testes supervisionados e os não supervisionados. A diferença fundamental entre as duas abordagens está no tratamento dos dados de entrada, quando são rotulados (supervisionados) e não rotulados (não-supervisionados). Explicando de forma individual, significa dizer que os modelos supervisionados possuem maior previsibilidade quanto ao *output* dos seus dados, uma vez que os dados de *input* são previamente categorizados

1.2.2 Interpretabilidade e Explicabilidade

Além da previsibilidade do *output* do algoritmo, ou seja, da predição ou processamento realizados pelo modelo treinado de ML, a interpretabilidade desses resultados é essencial. A interpretabilidade não é algo fácil de se explicar matematicamente, sendo um conceito subjetivo o quanto um ser humano é capaz de entender o que causou determinada decisão pelo algoritmo (Molnar 2023). A interpretabilidade é o quanto um ser humano pode, consistentemente, determinar o resultado de um modelo treinado. Assim, quanto maior a interpretabilidade de um modelo, mais fácil para uma pessoa compreender as decisões e os resultados do algoritmo (Molnar 2023). A importância da interpretabilidade está além do quanto é possível confiar nos resultados, pois tanto o modelo quanto o *dataset* de treinamentos devem ser adequados para determinada tarefa, possibilitando a realização dos ajustes mais delicados de todos os parâmetros envolvidos e refinando assim os resultados.

A interpretabilidade nem sempre é necessária, principalmente quando se trata de modelos que já foram exaustivamente estudados. Porém a explicabilidade é essencial, pois todo resultado deve responder a um por quê, como “por quê obtivemos este resultado, e não outro resultado qualquer?”. A qualidade e eficiência de determinado modelo não exclui a necessidade de ser possível explicar os resultados, como também a possibilidade do aprimoramento e adequação do modelo ao *dataset* e ao estudo a ser realizado.

1.2.3 Aplicabilidade dos modelos

Dentro da saúde humana, modelos de AI mostram seu potencial ao fazer o que parece ser impossível para humanos programarem. A criação de um código tradicional, onde o programador descreve literalmente e explicitamente as decisões que o computador deve tomar baseando-se nas variáveis presentes no *script*, tende a ser impossível frente ao volume e complexidade das variáveis envolvidas no funcionamento do corpo humano. Tal complexidade aumenta exponencialmente quando pensamos em sistemas de saúde onde são consideradas as informações de toda uma população (Rajkomar *et al.*, 2019).

A capacidade de aprender com exemplos faz com que ML possa entender as nuances e correlações entre as *features* e *labels*, criando uma espécie de mapa generalizado das variáveis e um modelo que pode ser utilizado para analisar casos “nunca antes vistos” pelo algoritmo, cruzando dados de *input* desses novos dados com os dados previamente conhecidos do seu banco e gerando uma resposta confiável (Rajkomar *et al.*, 2019).

A evolução na qualidade das imagens médicas e o avanço das técnicas de AI desde o início dos anos 1970 até o final dos anos 2010 contribuíram para o aumento significativo ano-a-ano das pesquisas médicas e dos diagnósticos auxiliados por computação, envolvendo ML e, mais recentemente, DL (Giger 2018). Em várias áreas a AI mostra potencial como cálculo de risco, detecção, diagnóstico e prognóstico. Cabe ressaltar o uso da AI na avaliação do sucesso terapêutico em diversos setores das imagens radiológicas, como na detecção de câncer de mama, interpretando em imagens de mamografia características que descrevem a densidade das mamas e padrões parenquimais, como também potencial para identificar fatores de risco no câncer de mamas. Além da detecção, existe grande potencial da AI para propor e avaliar a progressão de tratamentos (Giger 2018).

Na farmacologia, os estudos também empurram a fronteira do que ferramentas de AI podem fazer no campo da saúde. Algoritmos de AI podem ser úteis na redução e prevenção de eventos adversos por pacientes ao fazer uso de alguma droga, prevenir a interação droga-droga e até mesmo indicar uma opção de troca por uma droga que seja igualmente capaz de resolver o problema, sem causar maiores danos à saúde do paciente (Lee & Chen, 2021).

1.3 WGM e Inteligência Artificial

Na última década ocorreram demonstrações sólidas do potencial de aplicação de modelos de AI na identificação de gêneros e espécies de culicídeos. A arquitetura dos algoritmos e as ferramentas utilizadas para aquisição dos dados variam bastante, de sensores ópticos, acústicos, vídeos a imagens estáticas. Com sensores ópticos para captar a frequência pseudo-acústica da vibração das asas é possível criar uma identidade com precisão média de 93% na identificação de moscas e mosquitos de interesse médico, utilizando algoritmos “kernel adaptive autoregressive-moving average” (KAARMA) (Li & Principe, 2017). Similarmente, com o uso de Redes Neurais Convolucionais (CNN) , obteve-se precisão média em torno de 84% com espécies de *Aedes*, *Anopheles* e *Culex* (Fanioudakis *et al.*, 2018).

Outros estudos têm se concentrado na identificação de larvas de *Aedes*. Por exemplo, Arista-Jalife *et al.*, (2020) empregaram uma Deep Neural Network, enquanto Asmai *et al.*, (2019) utilizaram CNN implementada nos modelos GoogleNET, ResNet50, VGG16 e VGG19. O primeiro estudo visou distinguir as larvas de *Aedes aegypti* das amostras de *Aedes albopictus*, enquanto o segundo se concentrou em diferenciar as larvas de mosquitos "não-*Aedes*". Ambos os estudos alcançaram taxas de acurácia, atingindo 94% e 92%, respectivamente, na identificação das larvas.

Ainda com modelos de CNN, diferentes estudos exploraram a identificação de mosquitos inteiros. Um estudo comparou o método de extração de características, que envolve a definição manual de parâmetros a serem comparados, como cor, textura e forma, com métodos de ML, incluindo SVM, ResNet, VGGs e AlexNet. Os resultados mais promissores foram alcançados com ResNet, que atingiu uma precisão de até 95% (Okayasu *et al.*, 2019). Outro estudo recente, abrangendo 1585 espécimes de culicídeos e 90 espécimes de outros insetos alados (Kittichai *et al.*, 2021). Usando algoritmos de DL, os autores alcançaram uma precisão média de 99% e uma sensibilidade de 92% ao identificar gêneros de culicídeos. Um terceiro estudo com mosquitos do gênero *Aedes* alcançou taxas de acerto de até 95,19% na identificação (Siddiqua *et al.*, 2021). O estudo de Park *et al.* (2020) utilizou 3.600 imagens de 8 espécies, em diversas posturas e condições de deformação. Os modelos ResNet50 e VGG16 foram

aplicados na identificação de mosquitos. O modelo VGG16 foi superior na identificação dos mosquitos e apresentou uma acurácia de 97%. Utilizando uma armadilha luminosa, acoplada a um sensor de câmera fotográfica e sistema Raspberry Pi, que alimentam um modelo CNN, o estudo de Huang *et al.*, (2018), conseguiu 98% de acurácia na validação dos dados.

Em uma outra abordagem, algoritmos de CNN foram utilizados para criar um mapa com a localização geográfica do *upload* de dados referentes a espécies de mosquitos, com a finalidade de gerar previsões de futura ocorrência e abundância (Pataki *et al.*, 2021). Foi utilizado um banco de dados de fotos de mosquitos validadas por especialistas do Mosquito Alert, um sistema de ciência cidadã com um aplicativo para celular para coleta de imagens georreferenciadas. Como este sistema pode monitorar a propagação de várias espécies de mosquitos em todo o mundo, os autores treinaram nele um modelo de aprendizado profundo para encontrar mosquitos *Aedes albopictus* (Pataki *et al.*, 2021).

Essas abordagens diferem daquelas que se baseiam em WGM, uma vez que utiliza características de todo o mosquito, destacando a capacidade do ML em selecionar características e classificar mosquitos com base nelas. Além das demonstrações de como a AI é capaz de no futuro realizar a identificação visual automática de mosquitos em larga escala, utilizando as características do corpo inteiro, existem exemplos de como a utilização apenas da asa pode ser uma opção, principalmente quando falamos de espécimes identificados fora do seu habitat. A WGM pode fazer uso das ferramentas de AI de duas formas: processamento de arquivos TPS já marcados manualmente ou a detecção automática dos mesmos. Utilizando WGM como base, 388 asas pertencentes aos gêneros *Aedes*, *Culex* e *Anopheles* foram utilizadas para treinar um algoritmo multi-camadas (ANN), e identificar os mosquitos com base unicamente na morfometria dos 18 *landmarks* mais utilizados e previamente marcados (Lorenz *et al.*, 2015). Comparando com os resultados obtidos com Análise Discriminante, a métrica mais utilizada em WGM para classificação de espécimes, 5 das 17 espécies tiveram a mesma taxa de acerto na identificação dos mosquitos, enquanto nas outras 12 espécies, a taxa de acerto utilizando ANN foi superior, demonstrando maior confiabilidade no uso deste tipo de metodologia para a classificação utilizando WGM.

Em estudos de detecção automática de pontos, o escopo se estende para além dos mosquitos. Um programa desenvolvido em C chamado 'Findwing',

implementado em um dispositivo denominado 'Wingmachine', alcançou uma repetibilidade de mais de 90% na detecção de *landmarks* em asas de *Drosophildae* (Houle *et al.*, 2023). Em outra pesquisa focada nas asas posteriores de besouros, o par de asas primário utilizado para identificação, algoritmos de CNN (ResNet) foram empregados para mapear *landmarks* (Yang *et al.*, 2023).

Recentemente, a empresa Datamarkin (2023) lançou um protótipo de identificação automática dos *landmarks* nas asas de insetos de importância médica e ambiental, como mosquitos, flebotomíneos e abelhas. Embora o algoritmo que desenvolveram, conhecido como '*key point detection*', não seja capaz de identificar os mosquitos, a realização automática da detecção dos *landmarks* pela Datamarkin representa um avanço significativo para a área. Isso porque supera uma conhecida barreira dos estudos em WGM.

2. OBJETIVOS

2.1 Objetivo Geral:

Esse trabalho teve como objetivo desenvolver e avaliar um modelo de inteligência artificial para classificação de espécies de culicídeos através de morfometria de asa.

2.2 Objetivos Específicos:

- Automatizar a extração dos pontos a partir de imagens digitais;
- Classificar as espécies através de imagens de asa.

3. MATERIAL E MÉTODOS

3.1 Origem das imagens

Neste estudo foram utilizadas imagens de asas de culicídeos fêmeas previamente armazenadas provenientes de coletas realizadas no Brasil (Parques Municipais de São Paulo e Manaus) entre os anos de 2013 e 2016, utilizadas nos estudos de Wilke et al., 2016, como descrito da Tabela 1.

Tabela 1. Espécies utilizadas neste estudo, e quantidade de asas de cada uma.

Espécie	Número de Asas	Origem	Ano
<i>Culex nigripalpus</i>	30	PMSP, São Paulo	2015
<i>Culex quinquefasciatus</i>	30	PMSP, São Paulo	2015
<i>Aedes aegypti</i>	30	São Paulo, SP	2013
<i>Aedes serratus</i>	30	Marsilac, São Paulo - SP Horto Florestal, São Paulo, SP	2015/2016
<i>Anopheles darlingi</i>	30	Manaus, AM	2015
<i>Wyeomyia confusa</i>	30	Horto Florestal, São Paulo, SP	2015/2016
Total	180		

PMSP = Parques Municipais de São Paulo.

3.2 Padronização da aquisição de imagens digitais das asas

Para padronizar a aquisição das imagens após a identificação, a asa direita foi destacada com auxílio de pinça de ponta fina e acondicionada em microtubo, imersa em álcool etílico 70%, rotulada e armazenada. Posteriormente foi montada em lâmina e lamínula para microscopia, fixada com Bálsamo do Canadá. As imagens foram capturadas através de câmera fotográfica digital da marca Leica, modelo DFC320, acoplada em lupa estereoscópica marca Leica, modelo S6, no aumento de 40x. Para a marcação dos 18 pontos anatômicos foi utilizado o programa “tpsDig 1.4” (Rohlf, 2006), onde os 18 pontos anatômicos selecionados se encontram nas intersecções das veias alares. Dessa forma foi criado um arquivo

TPS, que consiste na sequência de coordenadas onde os marcos anatômicos estão localizados na imagem.

3.2.1 Morfometria geométrica alar

Podemos definir a alometria como o cálculo da regressão multivariada das coordenadas de Procrustes contra o tamanho do centróide usando um teste de permutação com 10.000 randomizações (Wilke *et al.*, 2016). Em seguida foi realizada a análise discriminante para explorar o grau de similaridade da forma das asas dos indivíduos de cada espécie no morfoespaço produzido pela análise de variáveis canônicas (CVA). As distâncias de Mahalanobis foram calculadas por meio da CVA, levando em consideração a variação e correlação entre as variáveis dos grupos pré-definidos e as distâncias entre os pontos. Dessa forma, cada indivíduo pode então ser classificado usando o teste de reclassificação cruzada e agrupado utilizando a distância de Mahalanobis (Lorenz & Suesdek, 2020; Vidal *et al.*, 2011). Para essas análises o software MorphoJ foi utilizado.

3.3 Extração dos pontos anatômicos para morfometria alar

A extração dos pontos anatômicos para realização do estudo foi feita utilizando scripts criados nas linguagens de programação R e Python. O processamento das imagens consistiu em preparar as imagens para a segmentação. A segmentação é parte crucial do protocolo, dependendo portanto de uma imagem pré-processada de maneira eficiente. Para iniciar o desenvolvimento do modelo, selecionamos uma asa direita de *Culex (Culex) quinquefasciatus* pertencente ao conjunto de imagens inicialmente disponível para este projeto.

Inicialmente, as imagens são transformadas para a resolução 1024x768 pixels², por se tratar de uma proporção comum em câmeras fotográficas de lupas e microscópios. Após este processo, a imagem é transformada para tons de cinza, convertida em 8-bit, evitando assim que diferenças de intensidades, brilho e contraste interfiram no resultado final, como descrito no trecho de código da **Figura 4**.

```
%% Cell 2
# Grayscale, Resize and Equalize histogram
img_gray = rgb2gray(img)
img_resize = resize(img_gray, (768, 1024), anti_aliasing=True)
# Histogram equalization
p2, p98 = np.percentile(img_resize, (2, 98))
img_rescale = rescale_intensity(img_resize, in_range=(p2, p98))
```

Figura 4. Trecho de código responsável pela padronização inicial das imagens.

Os passos listados anteriormente têm como principal objetivo equalizar ao máximo toda a variedade de cores provenientes não apenas da qualidade da iluminação e captura das imagens pelos mais diversos equipamentos ópticos, como também os meios de fixação, onde as mais diversas resinas utilizadas podem interferir nas características das imagens

3.3.1 Processamento de imagens

Com essa imagem, foi realizada a anotação manual das *labels* por meio de um visualizador multidimensional de imagens para Python (Napari, v0.4.18), sendo então as anotações que delimitam, sobre uma imagem, as regiões sob uma determinada categoria. Nas imagens de asas, três categorias foram criadas que posteriormente serviram para o algoritmo segmentar as imagens. As categorias são as venações, fundo e a área não marcada. Na área não marcada foram utilizadas tanto partes do fundo quanto venações, ou ruídos da imagem como escamas e cerdas possivelmente presentes na imagem, que não são destacadas manualmente (**Figura 5B**).

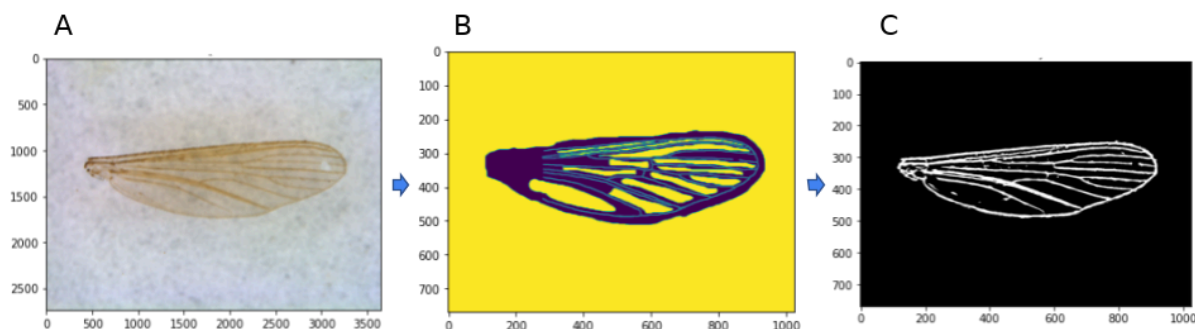


Figura 5. Processo de segmentação, seguido de binarização. **A.** Imagem original da asa do mosquito. **B.** *Labels* de marcação, onde temos em amarelo, o fundo, em azul claro, venações, e em azul escuro, a área não especificada pelas *labels*. **C.** Imagem binarizada.

Posteriormente é realizada a segmentação das veias alares utilizando classificador baseado em Random Forest (RF) (Smith, 2010). Foram considerados filtros de intensidade, textura e borda como características da imagem para classificar os pixels da imagem na região de interesse que representa as veias alares, como demonstrado no trecho de código da **Figura 6**. Assim, os marcos para a morfometria geométrica alar estão contidos dentro dessa região de interesse. Desta forma foi obtida a imagem da asa segmentada, ou seja, contendo apenas valores binários para o próximo passo, a esqueletonização.

```

%% Cell 3
# Segmentation
sigma_min = 1
sigma_max = 16
features_func = partial(feature.multiscale_basic_features,
                        intensity=True, edges=True, texture=True,
                        sigma_min=sigma_min, sigma_max=sigma_max)
features = features_func(img_rescale)
clf = RandomForestClassifier(n_estimators=50, n_jobs=8, #n_jobs=-1,
                           max_depth=10, max_samples=0.05)
clf = future.fit_segementer(labels, features, clf)

img_segmented = future.predict_segementer(features, clf)

```

Figura 6. Trecho de código responsável pelo treinamento do segmentador em *Random Forest*.

Em seguida, aplicam-se técnicas de morfologia matemática (Costa & Cesar, Jr., 2018) para afinar as estruturas e formar um “esqueleto” correspondente às veias

alares (Lee *et al.*, 1994). Então, através da ferramenta de análise de esqueleto (Nunez-Iglesias *et al.*, 2018), os pontos de intersecção das ramificações das veias alares foram obtidos; estes pontos foram usados como referência para a detecção dos marcos para morfometria geométrica alar.

3.3.2 Esqueletonização dos segmentos das veias alares

Para extração das informações do “esqueleto” gerado, foi utilizada a biblioteca Skan sobre a asa esqueletonizada. Utilizando a função *skeletonize* do Scikit-image, realizou-se o processo de afilar a imagem previamente segmentada ao ponto de que cada segmento demarcado e segmentado como venação, tenha apenas um pixel de espessura (**Figura 7**) e possa então ser mapeado. Esse processo pode ser chamado de esqueletonização.

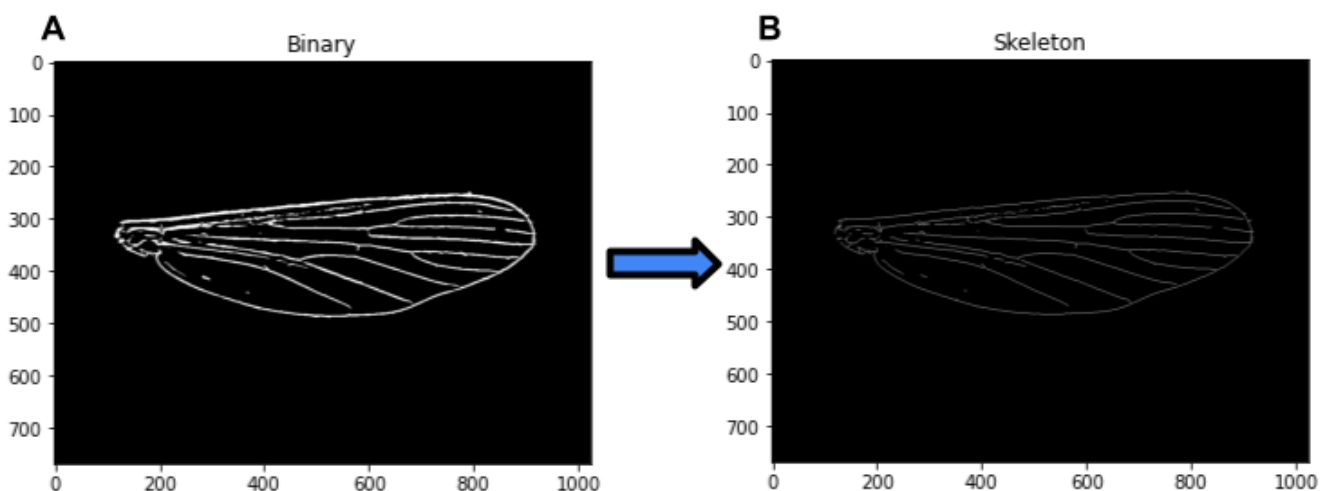


Figura 7. Transformação de uma imagem binarizada (A) em uma imagem esqueletonizada (B).

Experimentos com os esqueletos gerados também foram realizados, a fim de reduzir o número total de pontos gerados, tendo como objetivo a simplificação do processamento, e diminuindo também a chance de detecção de “falsos positivos”, ou seja, a detecção de pontos que não representem de fato os pontos de interesse

para a realização da Morfometria da Asa. Foram utilizadas as seguintes funções de filtros morfológicos da biblioteca de Python *scikit-image.morphology*: *binary_opening*, *binary_closing*, *binary_erosion* e *binary_dilation*. Os filtros morfológicos foram aplicados à imagem segmentada, antes de gerar os respectivos esqueletos em cada uma das análises, adicionalmente ao controle.

3.3.3 Obtenção dos pontos para WGM

Utilizando as funções *summarize* e *Skeleton* da biblioteca Skan, extraímos as variáveis dos segmentos das veias presentes na imagem do “esqueleto” da asa. Dentro dessas variáveis encontram-se as coordenadas relacionadas à origem e destino de todas as linhas, bem como as suas intersecções com outras linhas como ilustrado na **Figura 7**.

Com as ferramentas morfológicas presentes na biblioteca Scikit-image, foi realizada a redução no número de pontos gerados, simplificando assim a obtenção dos pontos de interesse. Também foi realizada a avaliação da qualidade em cada um dos tratamentos, para gerar o melhor processamento possível.

Pode ser observado que o Skan nos disponibiliza informações muito além das coordenadas necessárias para a criação de um arquivo no formato do TPS. Todas essas informações podem então ser organizadas em uma tabela, para que seja possível a manipulação das mesmas.

A tabela foi organizada então em duas colunas, correspondentes a localização de todos os pontos que compõem os segmentos das linhas identificadas. Utilizando o ferramental de manipulação de tabelas da biblioteca Pandas, essas informações foram organizadas, de modo que apenas o necessário para a geração de um arquivo com a estrutura do TPS com as coordenadas de todos os pontos de intersecção presentes no esqueleto seja usado.

Para a padronização da obtenção dos pontos de forma automatizada, foi criado um algoritmo para automatizar todos os passos anteriores em uma seleção de 30 asas pertencentes às espécies *Culex nigripalpus*, *Culex quinquefasciatus*, *Aedes aegypti*, *Aedes serratus*, *Anopheles darlingi* e *Wyeomyia confusa*, extraíndo o TPS de todos os pontos presentes no esqueleto de todas as asas, nos 4

tratamentos morfológicos previamente citados, juntamente do controle (nenhum tratamento aplicado).

Para a identificação dos pontos nas asas, outro algoritmo foi criado para realizar a sobreposição dos TPS “ground truth” e cada um dos TPS gerados para as asas, e dessa forma ser possível relacionar o ponto estimado mais próximo do manualmente anotado.

4. RESULTADOS

4.1 Modelo de Segmentação e Esqueletonização

O Modelo de Segmentação, apesar de desenhado em cima apenas das venações, e para não incluir na seleção o fundo e outros elementos (como demonstrado na **Figura 5**, evidenciou que a qualidade do objeto de entrada, ou seja, a qualidade da montagem da lâmina da asa, é um grande fator para os passos seguintes.

Além da qualidade da montagem da asa, um fator importante para o processamento das imagens, é a forma em que essas imagens são capturadas. Diferentes configurações na obtenção das imagens, características de iluminação, foco e nitidez, como também alterações nas características colorimétricas podem ocorrer, como mostra a **Figura 8**.

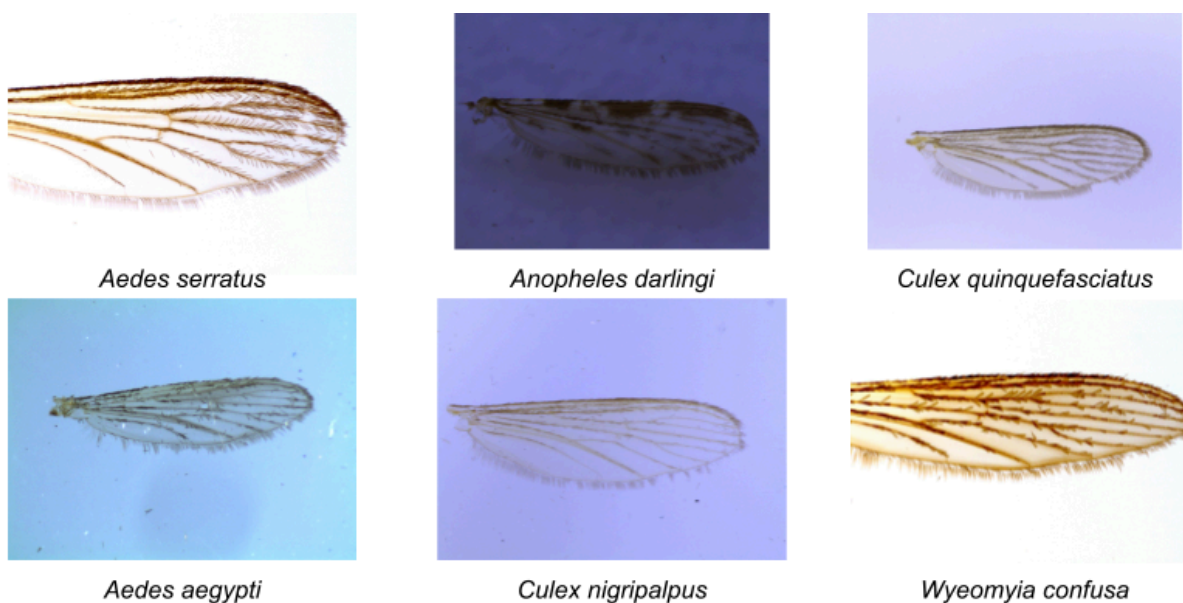


Figura 8. Exemplo das imagens capturadas via microscopia, de cada uma das espécies que passaram pelo processamento, e foram analisadas neste trabalho.

Nas **Figuras 9A e B** dois exemplos são apresentados, comparando duas asas de duas espécies diferentes, com características semelhantes, porém com montagens diferentes, submetidas ao mesmo algoritmo *Random Forest*. Foram obtidas segmentações muito diferentes, onde em A, o algoritmo de Random Forest

não foi capaz de distinguir as venações de interesse, dos artefatos decorrentes do preparo da amostra “L 13_D0025”. Em B, o excesso de escamas presentes ao redor da asa (amostra A5_Dx30.0008), decorrentes do processo de retirada das mesmas, não pôde ser diferenciado como ruído, na segmentação e esqueletonização.

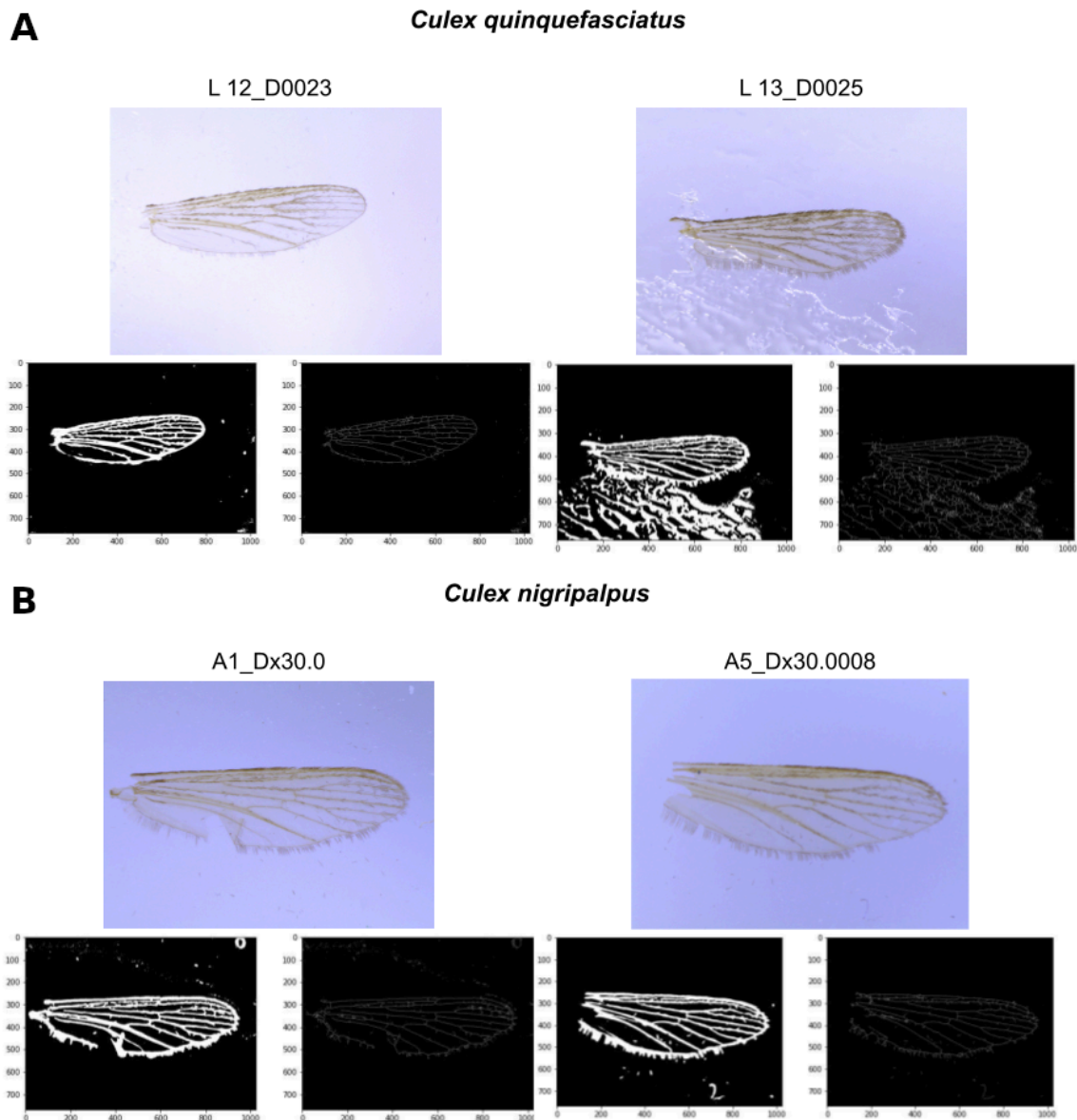


Figura 9. A. Asas de *Culex quinquefasciatus* em lâmina para microscopia, segmentação e esqueletonização. **B.** Asas de *Culex nigripalpus* em lâmina para microscopia, segmentação e esqueletonização.

Mesmo com imprecisões, as linhas gerais das venações foram obtidas, e com maior ou menor ruído, traduzidas nas imagens esqueletonizadas. Mesmo com maior número de pontos fora das asas, os pontos de interesse foram preservados na esqueletonização.

4.1.1. Filtros morfológicos

Para encontrar o filtro morfológico mais adequado, é necessário avaliar o impacto real de cada um, no número total de pontos gerados pelo algoritmo, sem a interferência de filtros, além da própria segmentação, anterior à eskeletonização.

Devido às particularidades anatômicas de cada espécie, como apresentar mais ou menos escamas na asa, bem como a facilidade em retirá-las durante o preparo das lâminas, a média dos pontos encontrados varia como demonstrado na **Figura 10**.

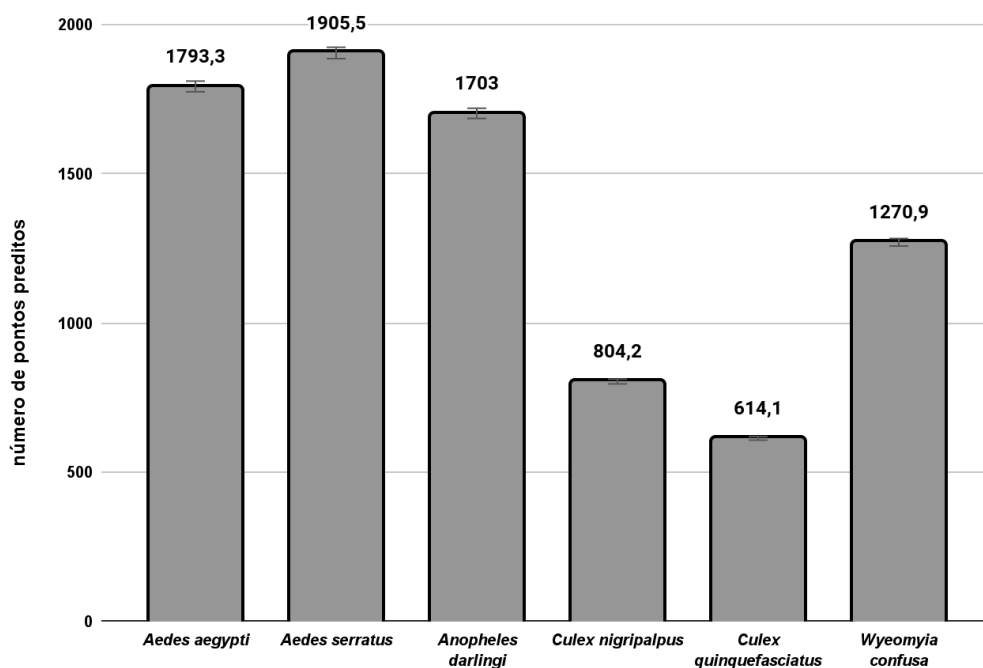


Figura 10. Média de pontos preditos por espécie na eskeletonização das imagens segmentadas das asas.

Em números absolutos, a média de pontos obtida pela espécie com a menor média (*Culex quinquefasciatus*), foi inferior a 33% do valor da média obtida pela espécie com maior número de pontos detectados (*Aedes serratus*) (**Figura 10**). Além das características anatômicas de cada espécie, precisamos levar em conta as variações nos procedimentos de montagem e aquisição das imagens. Boa parte destas variações são minimizadas no pré-processamento das imagens, mas ainda assim, as variações devem ser levadas em consideração.

Contudo, deve-se observar o sucesso dos filtros morfológicos em reduzir o número de pontos total das imagens. Os quatro filtros morfológicos obtiveram sucesso em reduzir por volta de 50% dos pontos detectados no esqueleto utilizando o *skan*, sendo a maior diferença obtida pelo filtro de erosão (59%), e a menor diferença obtida pelo filtro de dilatação (46,5%), como mostrado na **Figura 11**.

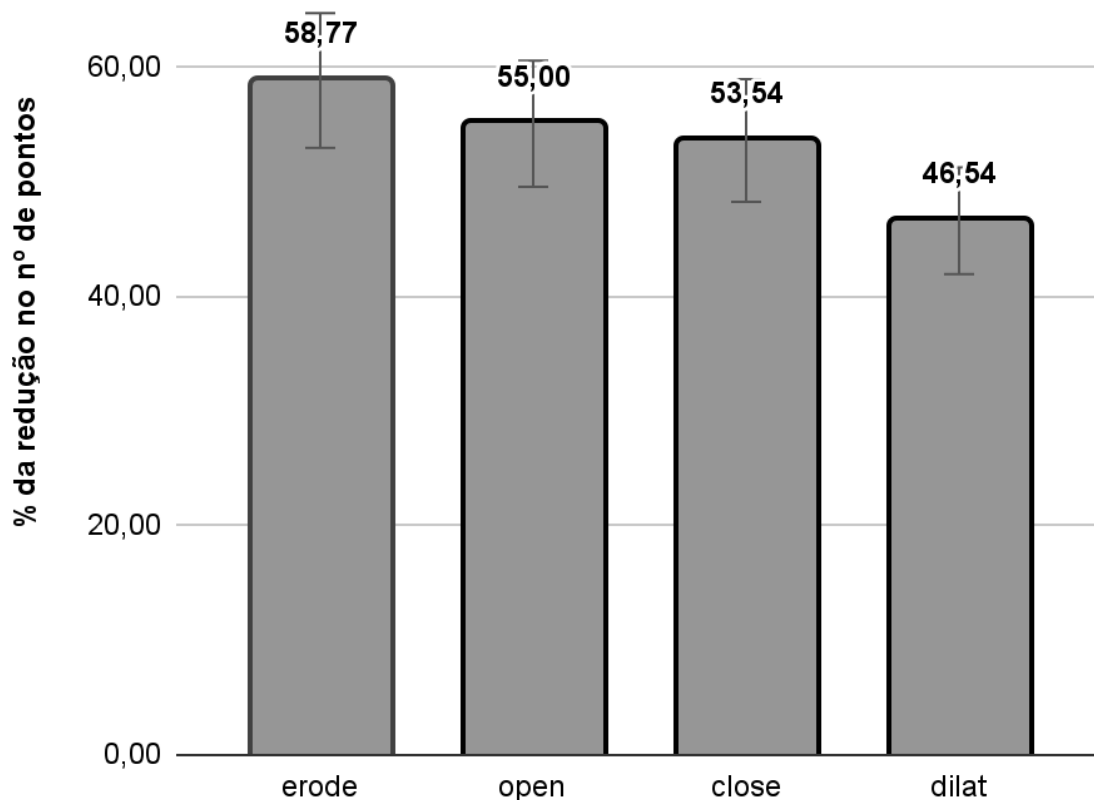


Figura 11. Porcentagem da redução no nº de pontos por filtro morfológico, utilizados (*scikit-image.morphology*) reduziram da média total de pontos obtidos em todas as amostras.

Analisando as informações obtidas pelos tratamentos, os resultados indicam que a diferença de aproximadamente 18% entre os filtros mais discrepantes, não mostra apenas a redução do número detectado de pontos, mas também afetou a localização dos pontos preditos mais próximos dos pontos reais, os pontos *ground-truth* marcados manualmente.

4.2 Definindo os pontos mais próximos

A sobreposição dos pontos *ground-truth* pelos pontos preditos no processamento das imagens, mostra que apesar da redução drástica no número de pontos preditos com a utilização dos filtros morfológicos, como evidenciado na **Figura 12A**, a estrutura básica do esqueleto não foi drasticamente alterada. O desenho geral da asa permanece inalterado com as reduções de pontos detectados.

Os pontos preditos mais próximos dos pontos *ground-truth* não sofreram grandes alterações. Alguns pontos preditos mais próximos mudam de um tratamento para o outro (**Figura 12B**), mas a média das distâncias de Mahalanobis entre cada uma das 18 correspondências se manteve em 1,75 em todos os tratamentos e no controle. A consistência entre Controle, Erosão e Dilatação (os dois tratamentos mais discrepantes), no entanto, não elimina a importância da realização de um ou mais tratamentos morfológicos na imagem, uma vez que falsos positivos podem representar um viés importante em bancos de dados mais volumosos.

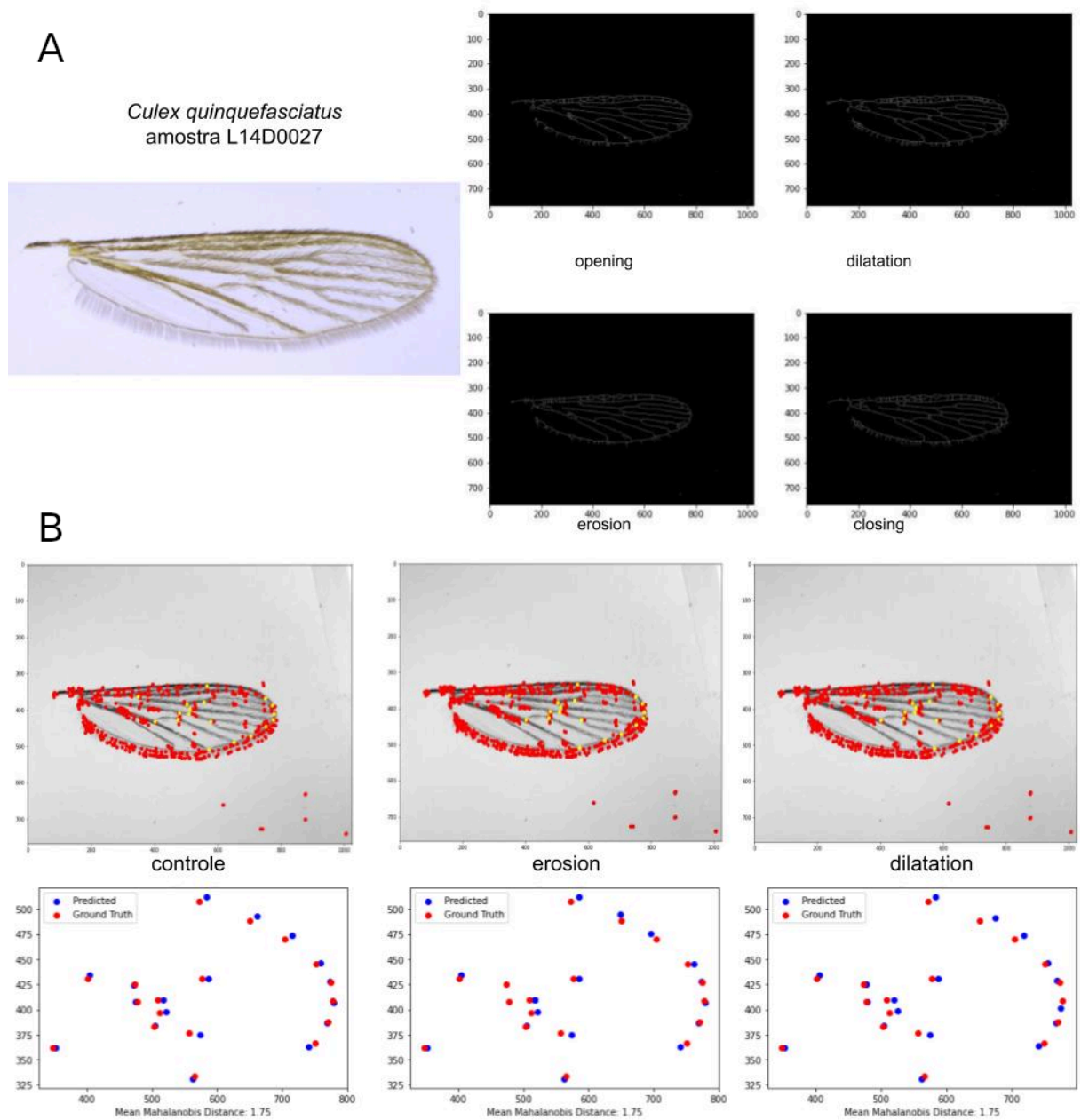


Figura 12. A. Asa de *Culex quinquefasciatus* ao lado da esqueletonização da imagem, tratada previamente com os 4 filtros morfológicos testados neste trabalho. **B.** Sobreposição na imagem pré-processada da asa, contendo em amarelo, *Ground-truth* e em vermelho, os pontos preditos nos dois tratamentos mais discrepantes; pontos preditos mais próximos dos pontos *Ground-truth*, com base na distância de Mahalanobis.

4.3 Classificação pela Menor Distância Média

Para encontrar a menor distância média entre os pontos preditos em uma asa não pertencente ao banco utilizado de base, e as espécies, uma asa teste (*Culex quinquefasciatus*, amostra L 37_D0071, (Figura 13A), passou por todos os passos do processamento, e teve os pontos do seu esqueleto (Figura 13B) anotados e exportados para um arquivo TPS.

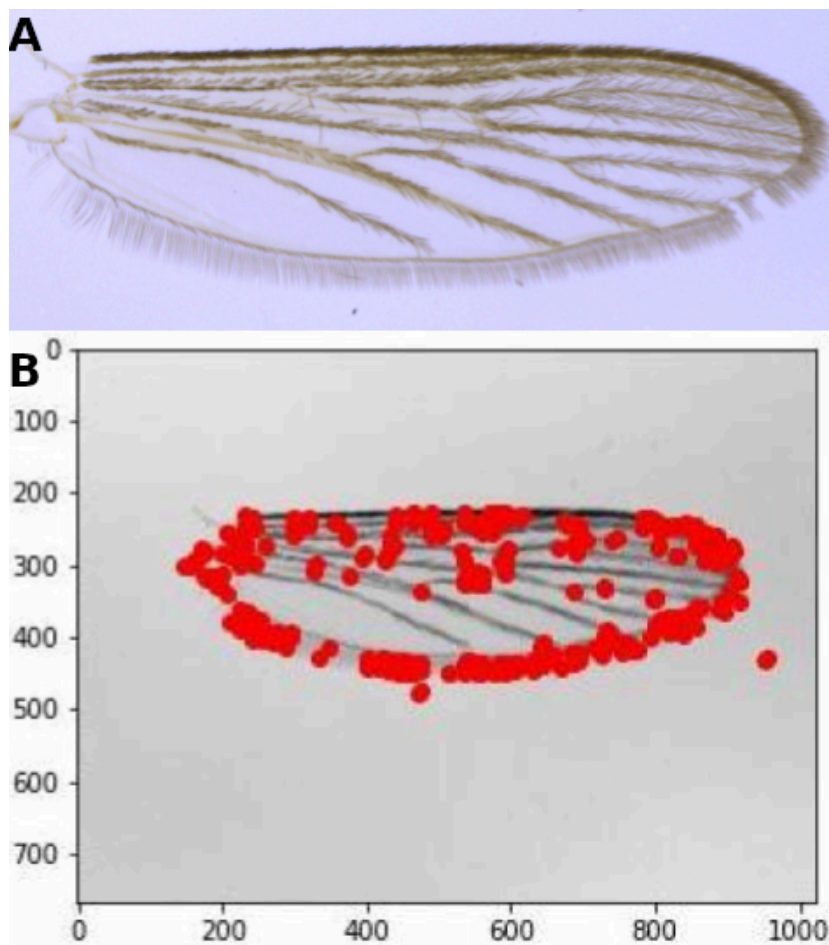


Figura 13. A. Asa direita de *Culex quinquefasciatus*. B. Sobreposição dos pontos (vermelho) detectados na venação sobre a imagem da asa.

Este arquivo de TPS gerado pelo algoritmo foi então comparado a cada amostra de cada espécie do banco (Tabela 1), de forma semelhante ao processo realizado na Figura 12B, com a diferença que após o alinhamento utilizando o cálculo do centróide, foram medidas as distâncias euclidianas entre os pontos mais

próximos, e então calculado a média entre as distâncias dos pontos, e enfim, a distância média entre a amostra de *Cx. quinquefasciatus* e toda a amostragem de cada espécie (**Tabela 2**).

Tabela 2. Distância Média Euclidiana entre os pontos mais próximos detectados no esqueleto da asa da amostra L 37_D0071, comparada com o banco das espécies selecionadas.

Espécie	Distância Média Euclidiana
<i>Culex quinquefasciatus</i>	175,1841292
<i>Culex nigripalpus</i>	222,7152369
<i>Aedes serratus</i>	295,6915567
<i>Aedes aegypti</i>	178,2622054
<i>Anopheles darlingi</i>	226,4491463
<i>Wyeomyia confusa</i>	325,4379989

Desta forma o algoritmo foi capaz de calcular a menor distância média da asa entre as espécies, chegando à menor distância com a amostragem de *Culex quinquefasciatus* do nosso banco de imagens. Assim, a metodologia empregada neste trabalho foi capaz de cruzar os dados gerados automaticamente, com dados de referência gerados manualmente, e apontar a espécie mais provável de uma asa de fora do banco, neste exemplo, de forma precisa.

5. DISCUSSÃO

A forma da asa dos mosquitos e outros insetos tem sido objeto de estudo quanto à sua capacidade de fornecer informações necessárias para a sua identificação. Durante as últimas décadas as metodologias de extração dessas informações passaram por grandes avanços, em especial após a integração de elementos computacionais nesse processo. Os métodos morfométricos demonstram capacidade semelhante a outros métodos de identificação, porém com suas próprias vantagens (Moraes, 2003; Rohlf, 2006; Vidal & Suesdek, 2012; Klingenberg & Marugán-Lóbon, 2013; Wilke *et al.*, 2016; Lorenz *et al.*, 2017; Simões *et al.*, 2020; Lorenz & Suesdek, 2020).

Ferramentas de Inteligência Artificial aliadas a ferramentas matemáticas e estatísticas, deram espaço para que diferentes métodos de automatizar a extração dessas informações pudessem surgir. A utilização de um algoritmo baseado em *Random Forest* (Smith, 2010) possibilitou que neste trabalho, o processo de segmentação das imagens de asas, em uma representação binária na região das venações, ocorresse de forma relativamente uniforme e bastante consistente. Mesmo nos casos onde há maior interferência na imagem, alheia às asas, os esqueletos foram extraídos da imagem original.

A importância do pré processamento e de uma segmentação consistentes ficou evidente quando se observa que, entre as imagens de asa com as melhores condições e as imagens com pior condição, todos os pontos presentes nas venações puderam ser preditos.

Os tratamentos morfológicos das imagens binarizadas (Costa & Cesar, Jr. 2018), previamente à eskeletonização, dispuseram de dados importantes. Reduzindo por volta de 50% o número de pontos preditos, mantendo todos os pontos de interesse, é possível avaliar que as escamas presentes nas asas dos mosquitos, principalmente nas venações das bordas da asa (veia Costal e toda região dorsal e apical), inflam o número de pontos preditos, correndo o risco de detecção de falsos positivos nessas áreas. Isso pode servir como indício de que a retirada de cerdas e escamas previamente à montagem das asas, é procedimento importante para a análise automatizada.

Outro ponto que ficou bastante evidente é a importância da qualidade da montagem das lâminas, pois, ainda que os *ground-truth* estejam localizados junto a pontos preditos. o excesso de pontos, e a presença exacerbada de pontos fora da asa, impossibilitam a identificação do espécime utilizando as ferramentas desenvolvidas neste trabalho.

Essa impossibilidade se dá pelo fato de que a principal ferramenta utilizada para alinhar os pontos preditos aos *ground-truth*, e por consequência em TPS de referência, é justamente o cálculo do centróide. Mesmo que asas como a mostrada na **Figura 14** passem por todo o processamento e gerem um dado de saída, esse dado de saída não será confiável, pois o cálculo centróide dos pontos previstos em uma imagem como essa sofrerá com a interferência de inúmeros pontos que não estão presentes nas asas dos mosquitos, sendo desviados, impedindo a confiabilidade e a previsibilidade do processamento.

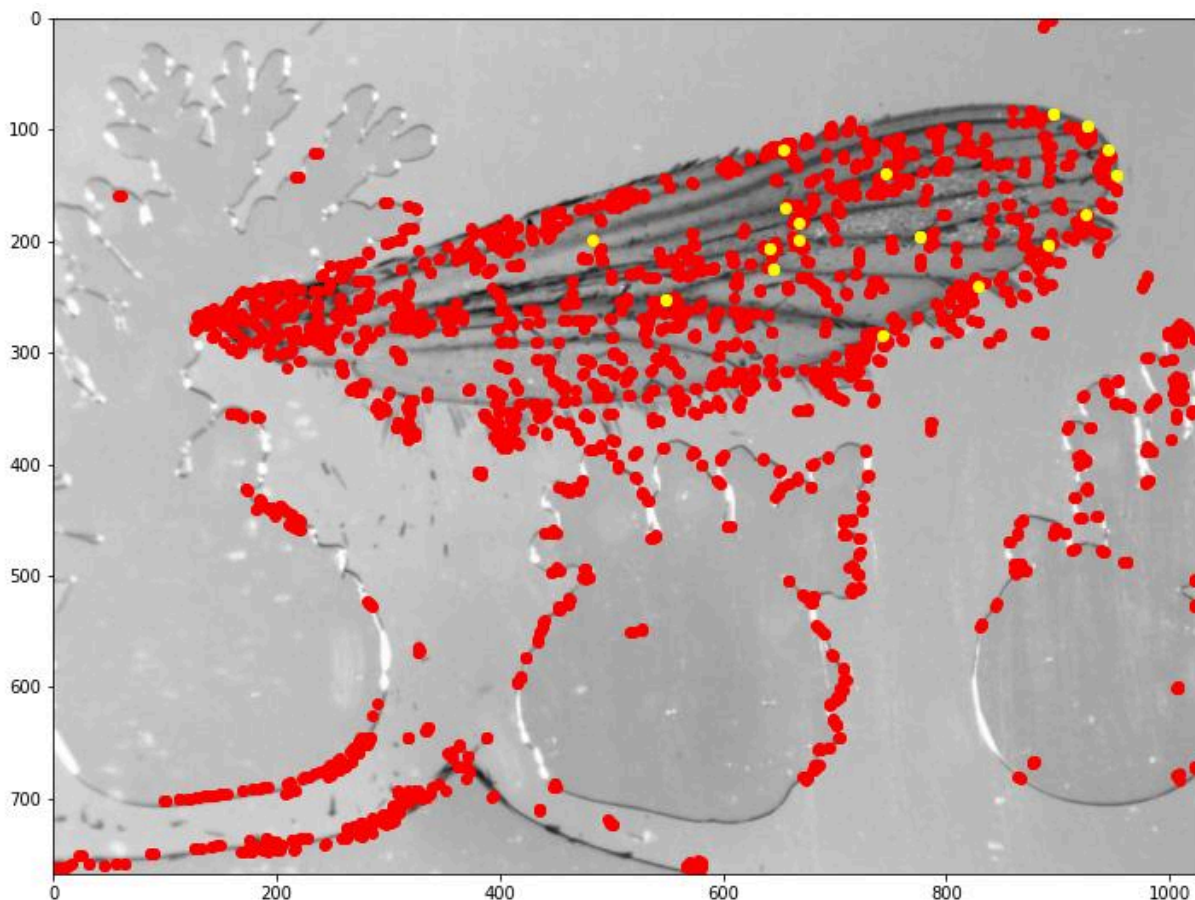


Figura 14. Foto de asa de *Aedes aegypti* sobreposta pelos pontos preditos (vermelho) e pontos *ground-truth* (amarelos).

A análise foi limitada ao número de espécies presentes no banco, assim como a identificação de mosquitos provenientes de fora do banco. Por este motivo existe a necessidade de manter os bancos de imagem e TPS ativos, atualizados e diversos, garantindo uma ampla variedade de espécies, e evitando identificações errôneas, pela falta de representatividade de amostras.

Neste trabalho, uma asa de *Culex quinquefasciatus*, onde as informações morfométricas foram geradas unicamente pelos algoritmos produzidos ao longo do projeto, foi classificada corretamente. Porém é necessário ressaltar que, como apresentado na **Tabela 2**, a segunda espécie que mais se aproximou da asa analisada foi *Aedes aegypti*, e não *Culex nigripalpus* como esperado por pertencerem ao mesmo gênero. No entanto, podemos observar na **Figura 15**, que as espécies *Ae. aegypti* e *Cx. quinquefasciatus* não apresentam nenhuma sobreposição na forma da asa.

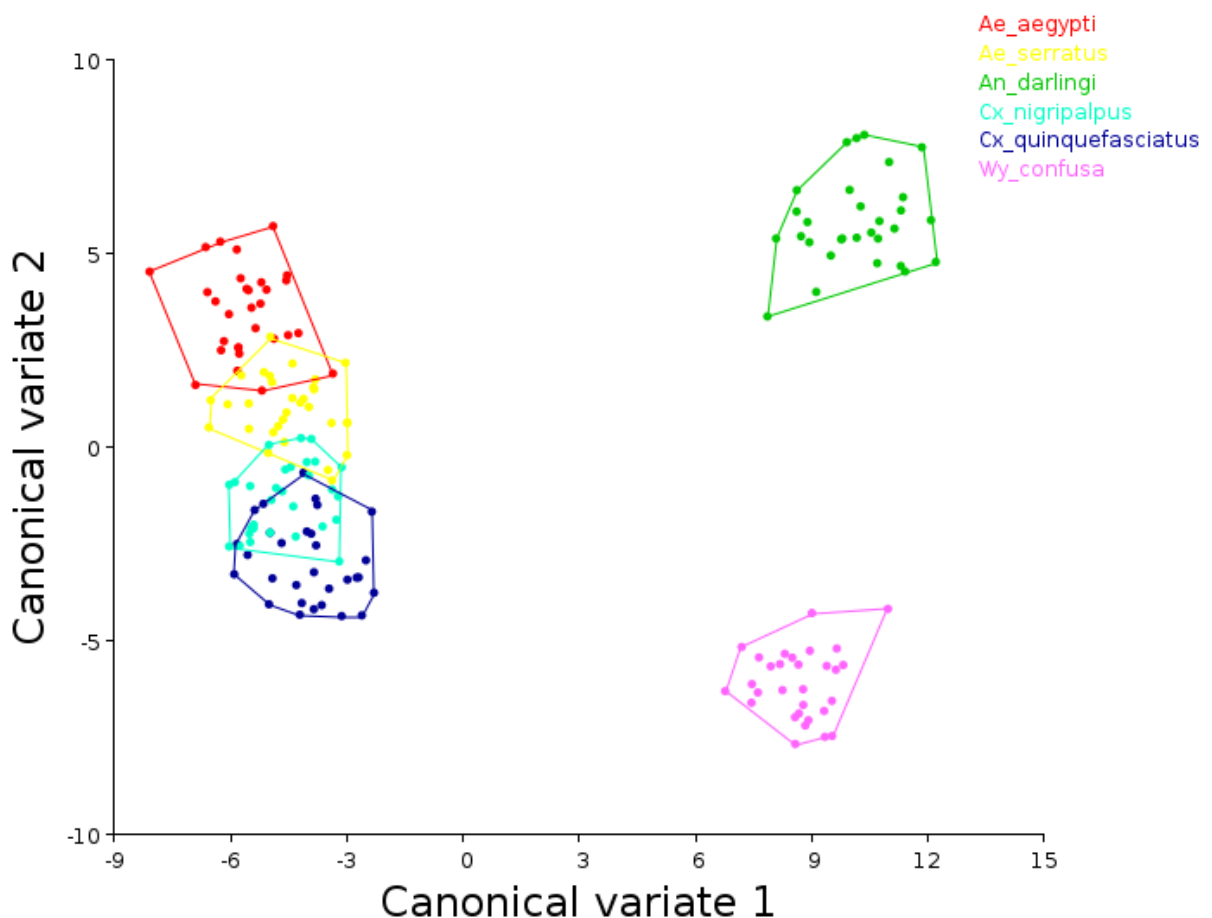


Figura 15. Morfoespaço gerado pelo cálculo da Variação Canônica entre as amostragens selecionadas para o estudo.

Estes resultados mostram que apesar do sucesso das metodologias desenvolvidas e empregadas ao longo deste projeto, existe margem para aprimoramentos, correções e implementações de novas técnicas.

Pode ser mencionado com destaque as duas metodologias que mais se aproximam do realizado neste trabalho. A primeira realizada por Lorenz *et al.*,(Lorenz *et al.*, 2015), que fez a utilização de Redes Neurais Artificiais para a classificação de mosquitos baseado na morfometria das asas, com o diferencial que neste estudo, foram utilizados como dados de entrada, arquivos de TPS previamente marcados, assim como para o treinamento do modelo de AI. A utilização de metodologia semelhante tem grande potencial de aprimorar as identificações geradas pelos pontos detectados neste projeto, como também no próprio processo de detecção dos pontos.

A outra metodologia que se assemelha a empregada por este projeto, é a detecção automática dos *landmarks* promovida pelo protótipo disponibilizado online pela empresa Datamarkin, que ao implementar um algoritmo de *key-points-detection*, conseguiu detectar os pontos de interesse em imagens de asa.

Ambas as estratégias de identificação tem grande valor para a comunidade científica, e por certo devem ser avaliadas como possíveis adições ao modelo gerado no presente trabalho, a fim de desenvolver uma solução rápida, prática e de baixo custo, com a finalidade de prover para a sociedade, uma ferramenta precisa e eficaz na identificação de insetos de grande importância médica.

6. CONCLUSÃO

- Foi desenvolvido um modelo de inteligência artificial para classificação de espécies de culicídeos através de morfometria de asa;
- A extração dos pontos para cálculo da alometria a partir de imagens digitais foi automatizada através da criação de um código em Python, com o auxílio de algoritmo de *Random Forest*;
- O modelo foi avaliado por meio do cálculo das Distâncias de Mahalanobis entre pontos preditos e referência;
- Foi avaliado o impacto de diferentes filtros morfológicos sobre a geração dos esqueletos e pontos preditos, com base nas distâncias de Mahalanobis;
- Uma imagem de asa de *Culex quinquefasciatus* foi classificada corretamente com base no cálculo das Distâncias Euclidianas entre banco de referências em TPS, gerando pontos preditos no esqueleto gerado automaticamente.

REFERÊNCIAS

- Adams, D. C.; Otárola-Castillo, E. Geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*. v. 4, p. 393-399, 2013. <https://doi.org/10.1111/2041-210X.12035>
- Agnew P, Koella JC. Virulence, parasite mode of transmission, and host fluctuating asymmetry. *Proc Biol Sci*. 1997 Jan 22;264(1378):9-15. doi: 10.1098/rspb.1997.0002.
- Arista-Jalife A, Nakano M, Garcia-Nonoal Z, Robles-Camarillo D, Perez-Meana H, Arista-Viveros HA (2020) Aedes mosquito detection in its larval stage using deep neural networks. *Knowl Based Syst* 189:104841. <https://doi.org/10.1016/j.knsys.2019.07.012>
- Asmai S, Zukhairin MNM, Jaya A, Rahman AFNA, Abas Z (2019) Mosquito larvae detection using deep learning. *Int J Innovative Technol Exploring Eng* 8(12):804–809. <https://doi.org/10.35940/ijitee.L3213.1081219>
- Bonnan MF. Linear and geometric morphometric analysis of long bone scaling patterns in Jurassic neosauropod dinosaurs: their functional and paleobiological implications. *Anat Rec (Hoboken)*. 2007 Sep;290(9):1089-111. doi: 10.1002/ar.20578.
- Calle L DA, Quiñones ML, Erazo HF, Jaramillo ON. Morphometric discrimination of females of five species of Anopheles of the subgenus Nyssorhynchus from Southern and Northwest Colombia. *Mem Inst Oswaldo Cruz*. 2002 Dec;97(8):1191-5. doi: 10.1590/s0074-02762002000800021.
- Cardillo M. 2010. Some applications of Geometric Morphometrics to archaeology. *Morphometrics to Nonmorphometricians*. (Lecture Notes in Earth Sciences). Edited by Ashraf M. T. Elewa. Springer, 325-341.
- Camargo-Neves VLF, Bergo ES, Guimarães LO, Clemente GRC, Lima VR, Kirchgatter K (2021) Atualidades em Medicina Tropical na América do Sul: Vetores, Atualizações no Métodos para Identificação de Espécies de Culicídeos. *Stricto Sensu*, p.197-228
- Consoli, R.A.G.B, Lourenço-De-Oliveira R. Principais Mosquitos de Importância Sanitária no Brasil. Rio de Janeiro: Fiocruz, 1994.
- Christe RO, Wilke ABB, Vidal PO, Marrelli MT. Wing sexual dimorphism in *Aedes fluviatilis* (Diptera: Culicidae). *Infect Genet Evol*. 2016 Nov;45:434-436. doi: 10.1016/j.meegid.2016.10.007.
- Chollet F (2021) *Deep Learning with Python*. 2. ed. Shelter Island, NY 11964: Manning Publications Co., 2021. 814 p.
- Datamarkin. Disponível em <<https://datamarkin.com/science>> Acessado em 10/07/2021.

Dujardin, J.P. Modern morphometrics of medically important insects. In: Tibayrenc, M. *Genetics and Evolution of Infectious Diseases*. Amsterdam: Elsevier, 2011.

Dujardin JP. Morphometrics applied to medical entomology. *Infect Genet Evol*. 2008 Dec;8(6):875-90. doi: 10.1016/j.meegid.2008.07.011.

Fanioudakis E, Geismar M, Potamitis I (2018) Mosquito wingbeat analysis and classification using deep learning, in: 2018 26th European Signal Processing Conference (EUSIPCO). Presented at the 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, Rome, pp. 2410–2414. Disponível em <<https://doi.org/10.23919/EUSIPCO.2018.8553542>> Acessado em 10/07/2021.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*. 1994 Oct;3(5):294-9.

Giger ML. Machine Learning in Medical Imaging. *J Am Coll Radiol*. 2018 Mar;15(3 Pt B):512-520. doi: 10.1016/j.jacr.2017.12.028.

Hajibabaei, M.; Smith, M.A.; Janzen, D.H.; Rodriguez, J.J.; Whitfield, J.B.; Hebert, P.D.N. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol Ecol Notes*, v. 6, p. 959–64, 2006.

Harbach, R.E. Mosquito Taxonomic Inventory. 2023. Disponível em <<https://mosquito-taxonomic-inventory.myspecies.info/>>. Acessado em: 10/07/2021.

Harbach, R.E. The Culicidae (Diptera): review of taxonomy, classification and phylogeny. *Zootaxa*, v. 1668, p. 591-638, 2007.

Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc Biol Sci*. 2003 Feb 7;270(1512):313-21. doi: 10.1098/rspb.2002.2218.

Houle D, Mezey J, Galpern P, Carter A. Automated measurement of *Drosophila* wings. *BMC Evol Biol*. 2003 Dec 11;3:25. doi: 10.1186/1471-2148-3-25.

Huang LP, Hong MH, Luo CH, Mahajan S, Chen LJ (2018) A vector mosquitoes classification system based on edge computing and deep learning. Conference on Technologies and Applications of Artificial Intelligence (TAAI), IEEE, pp. 24–27. Disponível em <<https://ieeexplore.ieee.org/document/8588471>>. Acessado em 10/09/2021.

Kittichai V, Pengsakul T, Chumchuen K, Samung Y, Sriwichai P, Phatthamolrat N, Tongloy T, Jaksukam K, Chuwongin S, Boonsang S. Deep learning approaches for challenging species and gender identification of mosquito vectors. *Sci Rep*. 2021 Mar 1;11(1):4838. doi: 10.1038/s41598-021-84219-4.

Klingenberg CP. MorphoJ: an integrated software package for geometric morphometrics. *Mol Ecol Resour*. 2011 Mar;11(2):353-7. doi: 10.1111/j.1755-0998.2010.02924.x.

Klingenberg CP, Marugán-Lobón J. Evolutionary covariation in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. *Syst Biol*. 2013 Jul;62(4):591-610. doi: 10.1093/sysbio/syt025.

Krieger, H.; Beiguelman, B.; Camargo E.P.; Krieger, M.; Vanin, S.A. Área de Ciências Biológicas. Parcerias estratégicas. Ministério da Ciência e Tecnologia, Centro de Estudos Estratégicos, Academia Brasileira de Ciências, Brasília, MCT, CGEE, ABC, Ed. Especial, v. 4, n. 14, 2002. Disponível em <<http://www.anped11.uerj.br/abc.pdf>>. Acessado em 10/07/2021.

Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60:84-90. <https://doi.org/10.1145/3065386>

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>

Lee CY, Chen YP. Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Brief Bioinform.* 2021 Mar 22;22(2):1884-1901. doi: 10.1093/bib/bbaa040.

Li K, Principe JC (2017) Automatic insect recognition using optical flight dynamics modeled by kernel adaptive ARMA network, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, New Orleans, LA, pp. 2726–2730. Disponível em <https://doi.org/10.1109/ICASSP.2017.7952652> Acessado em 10/07/2021.

Lorenz C, Ferraudo AS, Suesdek L. Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Trop* 2015;152:165–9. <https://doi.org/10.1016/j.actatropica.2015.09.011>.

Lorenz C, Almeida F, Almeida-Lopes F, Louise C, Pereira SN, Petersen V, Vidal PO, Virginio F, Suesdek L. Geometric morphometrics in mosquitoes: What has been measured? *Infect Genet Evol.* 2017 Oct;54:205-215. doi: 10.1016/j.meegid.2017.06.029.

Lorenz C, Suesdek L. Evaluation of chemical preparation on insect wing shape for geometric morphometrics. *Am J Trop Med Hyg.* 2013 Nov;89(5):928-931. doi: 10.4269/ajtmh.13-0359.

Lorenz C, Marques TC, Sallum MA, Suesdek L. Morphometrical diagnosis of the malaria vectors *Anopheles cruzii*, *An. homunculus* and *An. bellator*. *Parasit Vectors.* 2012 Nov 13;5:257. doi: 10.1186/1756-3305-5-257.

Lorenz C, Suesdek L. The use of wing shape for characterising macroevolution in mosquitoes (Diptera: Culicidae). *Infect Genet Evol.* 2020 Jan;77:104052. doi: 10.1016/j.meegid.2019.104052.

Mitteroecker, P.; Gunz, P. Advances in Geometric Morphometrics. *Evol Biol*, v. 36, p. 235–247, 2009.

Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. 2023. Disponível em <<https://christophm.github.io/interpretable-ml-book/>>. Acessado em 10/09/2023.

Moraes, D.A. A morfometria geométrica e a “Revolução na morfometria” localizando e visualizando mudanças nas formas dos organismos. *Bioetím*, v. 3, p. 1-5, 2003.

Mpho M, Holloway GJ, Callaghan A. A comparison of the effects of organophosphate insecticide exposure and temperature stress on fluctuating asymmetry and life history traits in *Culex quinquefasciatus*. *Chemosphere*. 2001 Nov;45(6-7):713-20. doi: 10.1016/s0045-6535(01)00140-0.

Mpho M, Callaghan A, Holloway GJ. Temperature and genotypic effects on life history and fluctuating asymmetry in a field strain of *Culex pipiens*. *Heredity (Edinb)*. 2002 Apr;88(4):307-12. doi: 10.1038/sj.hdy.6800045.

Muhammad Tahir H, Akhtar S. Services of DNA barcoding in different fields. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016 Nov;27(6):4463-4474. doi: 10.3109/19401736.2015.1089572.

Multini LC, Wilke ABB, Marrelli MT. Urbanization as a driver for temporal wing-shape variation in *Anopheles cruzii* (Diptera: Culicidae). *Acta Trop*. 2019 Feb;190:30-36. doi: 10.1016/j.actatropica.2018.10.009.

Orlandin E, Santos EB, Piovesan M, Favretto MA, Schneeberger AH, Souza VO, Muller GA, Wagner G (2017) Mosquitoes (Diptera: Culicidae) from crepuscular period in an Atlantic Forest area in Southern Brazil. *Braz. J. Biol.* 77 (01). <https://doi.org/10.1590/1519-6984.09815>

Okayasu K, Yoshida K, Fuchida M, Nakamura A (2019) Vision-based Classification of Mosquito Species: comparison of conventional and deep learning methods. *Appl Sci* 9:3935. <https://doi.org/10.3390/app9183935>

Park J, Kim DI, Choi B, Kang W, Kwon HW. Classification and Morphological Analysis of Vector Mosquitoes using Deep Convolutional Neural Networks. *Sci Rep*. 2020 Jan 23;10(1):1012. doi: 10.1038/s41598-020-57875-1.

Pataki BA, Garriga J, Eritja R, Palmer JRB, Bartumeus F, Csabai I. Deep learning identification for citizen science surveillance of tiger mosquitoes. *Sci Rep*. 2021 Feb 25;11(1):4718. doi: 10.1038/s41598-021-83657-4.

Raia-Barjat T, Tardieu AS, Amouzougan A, Trombert B, Chauleur C, Varlet MN, Patural H, Seffert P, Chêne G. Anthropometric analysis of obstetrical pelvis from Neolithic area: obstetrical consequences. Preliminary study. *Gynecol Obstet Fertil*. 2011 Nov;39(11):614-9. French. doi: 10.1016/j.gyobfe.2011.07.028.

Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019 Apr 4;380(14):1347-1358. doi: 10.1056/NEJMra1814259.

Renaud AK, Savage J, Adamowicz SJ. DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. *BMC Ecol*. 2012 Nov 23;12:24. doi: 10.1186/1472-6785-12-24.

Rohlf, F. J. (2006). *tpsDig, Digitize Landmarks and Outlines, Version 2.05*. Stony Brook, NY: Department of Ecology and Evolution, State University of New York.

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012 Jul;9(7):671-5. doi: 10.1038/nmeth.2089.

Siddiqua R, Rahman S, Uddin J (2021) A Deep Learning-based Dengue Mosquito Detection Method Using Faster R-CNN and Image Processing Techniques. *AETiC*, 5:3, p. 11-23. <https://doi.org/10.33166/AETiC.2021.03.002>

Simões RF, Wilke ABB, Chagas CRF, Menezes RMT, Suesdek L, Multini LC, Silva FS, Grech MG, Marrelli MT, Kirchgatter K. Wing Geometric Morphometrics as a Tool for the Identification of *Culex* Subgenus Mosquitoes of *Culex* (Diptera: Culicidae). *Insects*. 2020 Aug 25;11(9):567. doi: 10.3390/insects11090567.

A. Smith, B. Sterba-Boatwright, J. Mott (2010) Novel application of a statistical technique, Random Forests, in a bacterial source tracking study *Water Res.*, 44, pp. 4067-4076. doi: 10.1016/j.jes.2023.03.012

Turing AMI (1950) Computing Machinery and Intelligence. *Mind* LIX:433-460. <https://doi.org/10.1093/mind/LIX.236.433>

Vidal PO, Suesdek L. Comparison of wing geometry data and genetic data for assessing the population structure of *Aedes aegypti*. *Infect Genet Evol*. 2012 Apr;12(3):591-6. doi: 10.1016/j.meegid.2011.11.013.

Virginio F, Domingues V, da Silva LCG, Andrade L, Braghetto KR, Suesdek L. WingBank: A Wing Image Database of Mosquitoes. *Front. Ecol. Evol.*, v. 9, 2021. doi: 10.3389/fevo.2021.660941

Wilke AB, Christie Rde O, Multini LC, Vidal PO, Wilk-da-Silva R, de Carvalho GC, Marrelli MT. Morphometric Wing Characters as a Tool for Mosquito Identification. *PLoS One*. 2016 Aug 23;11(8):e0161643. doi: 10.1371/journal.pone.0161643.

Yang Y, Liu X, Li W, Li C, Ma G, Yang G, Ren J, Ge S. Detection of Hindwing Landmarks Using Transfer Learning and High-Resolution Networks. *Biology (Basel)*. 2023 Jul 14;12(7):1006. doi: 10.3390/biology12071006.

ANEXOS

Anexo 1. Autorização CEUA



SECRETARIA DE ESTADO DA SAÚDE
SUPERINTENDÊNCIA DE CONTROLE DE ENDEMIAS
"SUCEN"



Comissão de Ética no Uso de Animais

CERTIFICADO

Certificamos que a proposta intitulada "Inteligência Artificial para Classificação de Espécies de Culicídeos Baseada em Morfometria de Asas.", protocolada sob o CEUA-SUCEN No. 0007, sob a responsabilidade de **Karin Kirchgatter Hildebrand** que envolve a utilização de material biológico de animais invertebrados capturados na rotina de vigilância da Superintendência de Controle de Endemias do Estado de São Paulo, trata-se do uso de asas de mosquitos par fins de morfometria, não havendo manipulação de animais vivos, invertebrados ou vertebrados. A presente proposta está de acordo com os preceitos da Lei 11.794 de 8 de outubro de 2008, com o Decreto 6.899 de 15 de julho de 2009, bem como com as normas editadas pelo Conselho Nacional de Controle da Experimentação Animal (CONCEA), e foi avaliado *ad referendum* em 28/04/2021 e **aprovada** por revisor *ad hoc* pela Comissão de Ética no Uso de Animais da Sucen.

Finalidade da Proposta: [Pesquisa](#)

Vigência da Proposta: de [05/2021](#) a [05/2023](#)

São Paulo, 28 de Abril de 2021

Adriano Pinter dos Santos

Med. Veterinário
Pesquisador Científico VI
Presidente da Comissão de Ética no Uso de Animais
SUCEN

Anexo 2. Ficha do Aluno Janus.

Janus - Sistema Administrativo da Pós-Graduação



Universidade de São Paulo
Faculdade de Medicina
FICHA DO ALUNO

5179 - 12825490/1 - Vinicio Rodrigues de Lima

Email: viniciorodrigues@usp.br
Data de Nascimento: 28/09/1992
Cédula de Identidade: RG - 35464301-0 - SP
Local de Nascimento: Estado de São Paulo
Nacionalidade: Brasileira
Graduação: Bacharel em Ciências Biológicas - Universidade Cruzeiro do Sul - São Paulo - Brasil - 2018

Curso: Mestrado
Programa: Medicina Tropical
Área: Doenças Tropicais e Saúde Internacional
Data de Matrícula: 07/05/2021
Início da Contagem de Prazo: 07/05/2021
Data Limite para o Depósito: 07/11/2023
Orientador: Prof(a). Dr(a). Karin Kirchgatter - 07/05/2021 até o presente. Email: karink@usp.br
Co-orientador: Prof(a). Dr(a). Mauro Cesar Cafundó de Moraes - 14/10/2021 até o presente. Email: mauro_morais@usp.br
Proficiência em Línguas: Inglês, 07/05/2021
Data de Aprovação no Exame de Qualificação: Aprovado em 31/01/2023
Data do Depósito do Trabalho:
Título do Trabalho:
Data Máxima para Aprovação da Banca:
Data de Aprovação da Banca:
Data Máxima para Defesa:
Data da Defesa:
Resultado da Defesa:
Histórico de Ocorrências: Primeira Matrícula em 07/05/2021

Aluno matriculado no Regimento da Pós-Graduação USP (Resolução nº 7493 em vigor a partir de 29/03/2018).

Última ocorrência: Matrícula de Acompanhamento em 17/07/2023

Impresso em: 03/11/2023 12:41:45



Universidade de São Paulo
Faculdade de Medicina
FICHA DO ALUNO

5179 - 12825490/1 - Vinicio Rodrigues de Lima

Sigla	Nome da Disciplina	Início	Término	Carga Horária	Cred.	Freq.	Conc.	Exc.	Situação
IMT5117-4/1	Tópicos Avançados em Doenças Infecciosas Relevantes para a Saúde Internacional	27/08/2021	09/12/2021	30	2	100	A	N	Concluída
IMT5106-3/1	Marcadores Moleculares em Plasmodium e Culicídeos Vetores: Especiação e Infecção	04/11/2021	24/11/2021	45	3	100	A	N	Concluída
BIO5766-3/2	Introdução à Análise da Forma Através da Morfometria Geométrica (Instituto de Biociências - Universidade de São Paulo)	06/12/2021	12/12/2021	15	1	100	A	N	Concluída
IMT5132-1/1	Zoonoses Emergentes e Reemergentes Transmitidas por Vetores	03/05/2022	23/05/2022	45	3	100	A	N	Concluída
BIE5782-5/1	Uso da Linguagem R para Análise de Dados em Ecologia (Instituto de Biociências - Universidade de São Paulo)	06/06/2022	26/06/2022	60	4	100	B	N	Concluída
MZP5017-3/3	Dipterologia (Museu de Zoologia - Universidade de São Paulo)	07/11/2022	04/12/2022	120	8	100	A	N	Concluída

	Créditos mínimos exigidos		Créditos obtidos
	Para exame de qualificação	Para depósito da dissertação	
Disciplinas:	10	20	21
Estágios:			
Total:	10	20	21

Créditos Atribuídos à Dissertação: 80

Conceito a partir de 02/01/1997:

A - Excelente, com direito a crédito; B - Bom, com direito a crédito; C - Regular, com direito a crédito; R - Reprovado; T - Transferência.

Um(1) crédito equivale a 15 horas de atividade programada.

Última ocorrência: Matrícula de Acompanhamento em 17/07/2023

Impresso em: 03/11/2023 12:41:45

Anexo 3. Primeira página do artigo publicado.

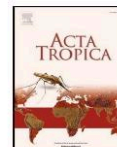
Acta Tropica 249 (2024) 107089



Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica



Integrating artificial intelligence and wing geometric morphometry to automate mosquito classification

Vinício Rodrigues de Lima^a, Mauro César Cafundó de Morais^{b,c}, Karin Kirchgatter^{a,d,*}

^a Programa de Pós-Graduação em Medicina Tropical, Faculdade de Medicina, Instituto de Medicina Tropical, Universidade de São Paulo, São Paulo, SP 05403-000, Brazil

^b Instituto Israelita de Ensino e Pesquisa Albert Einstein (IIEPAE), Sociedade Beneficente Israelita Brasileira Albert Einstein (SBIBAE), São Paulo, SP, Brazil

^c Computational Systems Biology Laboratory (CSBL), Institut Pasteur de São Paulo, São Paulo, SP 05508-020, Brazil

^d Laboratório de Bioquímica e Biologia Molecular, Instituto Pasteur, São Paulo, SP 01027-000, Brazil

ARTICLE INFO

Keywords:

Mosquito-borne diseases
Species identification
Integrative approach

ABSTRACT

Mosquitoes (Diptera: Culicidae) comprise over 3500 global species, primarily in tropical regions, where the females act as disease vectors. Thus, identifying medically significant species is vital. In this context, Wing Geometric Morphometry (WGM) emerges as a precise and accessible method, excelling in species differentiation through mathematical approaches. Computational technologies and Artificial Intelligence (AI) promise to overcome WGM challenges, supporting mosquito identification. AI explores computers' thinking capacity, originating in the 1950s. Machine Learning (ML) arose in the 1980s as a subfield of AI, and deep Learning (DL) characterizes ML's subcategory, featuring hierarchical data processing layers. DL relies on data volume and layer adjustments. Over the past decade, AI demonstrated potential in mosquito identification. Various studies employed optical sensors, and Convolutional Neural Networks (CNNs) for mosquito identification, achieving average accuracy rates between 84 % and 93 %. Furthermore, larval *Aedes* identification reached accuracy rates of 92 % to 94 % using CNNs. DL models such as ResNet50 and VGG16 achieved up to 95 % accuracy in mosquito identification. Applying CNNs to georeference mosquito photos showed promising results. AI algorithms automated landmark detection in various insects' wings with repeatability rates exceeding 90 %. Companies have developed wing landmark detection algorithms, marking significant advancements in the field. In this review, we discuss how AI and WGM are being combined to identify mosquito species, offering benefits in monitoring and controlling mosquito populations.

1. Introduction

Mosquitoes (Diptera: Culicidae) comprise a group with over 3500 described species (Orlandin et al., 2017; Harbach, 2023), distributed throughout the tropical and temperate regions of the planet. They are encountered at all vertical strata within the environments they inhabit, coexisting with human populations in both rural and urban settings. However, their development and greater diversity are particularly prominent in tropical forest regions (Harbach, 2007).

The vertical stratification of mosquitoes in natural environments is directly dependent on the feeding habits of each species and the availability of food sources. Males of some species feed on liquid plant sources such as sap, nectar, and fruits. However, in most species, females

require blood meals from warm-blooded vertebrates. In the absence of warm-blooded hosts, other cold-blooded animals such as snakes, amphibians, and even worms and leeches can serve as a source of nourishment to complete the maturation of their eggs (Harbach, 2007). This hematophagous habit ultimately renders culicids important vectors of arboviruses and other pathogens transmitted through the bite of these insects (Harbach, 2023).

The identification of species of medical importance contributes to a better understanding of the biological characteristics of these disease-vector insects. These insects often face various forms of selective pressure, including parasitic infections, environmental and landscape changes, insecticide usage, among others (Dujardin, 2008). Many mosquito species, although similar in appearance, possess unique biology.

* Corresponding author at: Programa de Pós-Graduação em Medicina Tropical, Faculdade de Medicina, Instituto de Medicina Tropical, Universidade de São Paulo, São Paulo, SP 05403-000, Brazil.

E-mail address: karink@usp.br (K. Kirchgatter).

<https://doi.org/10.1016/j.actatropica.2023.107089>

Received 1 November 2023; Received in revised form 29 November 2023; Accepted 30 November 2023

Available online 1 December 2023

0001-706X/© 2023 Elsevier B.V. All rights reserved.