

Débora Bruna de Morais

**Modelo de *Machine Learning* Aplicado à Identificação do Mosquito *Aedes aegypti***

Goiânia

2023



UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

## TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC no 1240/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei no 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

### 1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG)

Nome(s) completo(s) do(a)(s) autor(a)(es)(as): Débora Bruna Rodrigues de Moraes

Título do trabalho: Modelo de Machine Learning Aplicado à Identificação do Mosquito Aedes Aegypti

### 2. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador) Concorda com a liberação total do documento [ X ] SIM [ ] NÃO<sup>1</sup>

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

#### Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro.

**Obs.: Este termo deve ser assinado no SEI pelo orientador e pelo autor.**



Documento assinado eletronicamente por **Sanderley Ramos Pires, Professor do Magistério Superior**, em 24/08/2023, às 09:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Debora Bruna Rodrigues De Moraes, Discente**, em 24/08/2023, às 10:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3989540** e o código CRC **75FCF37A**.

---

Referência: Processo nº 23070.024218/2023-26

SEI nº 3989540

Débora Bruna de Morais

**Modelo de *Machine Learning* Aplicado à Identificação do Mosquito *Aedes aegypti***

Trabalho de conclusão de curso apresentado na Escola de Engenharia Elétrica, Mecânica e de Computação como requisito para a conclusão do curso de Engenharia de Computação e obtenção do título de Engenheira de Computação.

Universidade Federal de Goiás – UFG  
Escola de Engenharia, Mecânica e Computação (EMC)  
Trabalho de Conclusão de Curso

**Orientador:** Prof. Dr. Sandrerley Ramos Pires

Goiânia

2023

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Morais, Débora Bruna de  
Modelo de Machine Learning Aplicado à Identificação do Mosquito  
*Aedes aegypti* [manuscrito] / Débora Bruna de Moraes. - 2023.  
29 f.

Orientador: Prof. Dr. Sandrerley Ramos Pires.  
Trabalho de Conclusão de Curso (Graduação) - Universidade  
Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de  
Computação (EMC), Engenharia da Computação, Goiânia, 2023.  
Bibliografia.

Inclui fotografias, algoritmos, lista de figuras.

1. *Aedes aegypti*. 2. Mosquito. 3. Classificação de imagens. 4.  
Aprendizado de máquina. 5. Redes neurais convolucionais. I. Pires,  
Sandrerley Ramos, orient. II. Título.

CDU 519.25



UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

## DECLARAÇÃO

### ATA DE AVALIAÇÃO DE PROJETO FINAL

#### CURSO

<input type="checkbox"/> Eng Elétrica	<input type="checkbox"/> Eng Mecânica	<input checked="" type="checkbox"/> Eng Computação PFC 1 <input type="checkbox"/> PFC 2 <input checked="" type="checkbox"/>
---------------------------------------	---------------------------------------	--

#### TÍTULO DO TRABALHO

**Modelo de Machine Learning Aplicado à Identificação do Mosquito Aedes Aegypti**

#### BANCA AVALIADORA

Membro 1	Sandrerley Ramos Pires
Membro 2	Deborah Silva Alves Fernandes
Membro 3	Karina Rocha Gomes da Silva

#### DATA DA DEFESA

21/08/2023

#### DISCENTE

Matrícula	Nome
201602410	Débora Bruna Rodrigues de Morais

#### NOTAS

Matrícula	Membro 1			Membro 2			Membro 3			Média*
	NPT	NTE	NAA	NPT	NTE	NAA	NPT	NTE	NAA	
201602410	10,0	8,0	9,5	10,0	8,0	9,5	10,0	8,0	9,5	8,9

--	--	--	--	--	--	--	--	--	--

**NPT** – Nota plano de trabalho;

**NTE** – Nota do trabalho escrito;

**NAA** – Nota de apresentação e arguição

Para Eng. Elétrica, Mecânica e PFC2 da Eng. Da Computação:  $NF = 0,1 \times NPT + 0,45 \times NTE + 0,45 \times NAA$

Para PFC1 da Eng. Da Computação:  $NF = 0,3 \times NPT + 0,7 \times NAA$

\* A APROVAÇÃO DO(S) ALUNO(S) ESTÁ CONDICIONADA À APRESENTAÇÃO DO TRABALHO FINAL AO ORIENTADOR COM TODAS AS CORREÇÕES SUGERIDAS PELA BANCA.

#### OBSERVAÇÕES:

Preencher com modificações solicitadas, caso existam. Em caso de reprovação, informar a justificativa.

Foi solicitado a aluna que:

1. Numerasse as páginas da monografia, passasse as palavras de língua estrangeira para itálico;
2. Ajustar o documento para que todas as figuras sejam referenciadas e ser comentadas no texto;
3. A conclusão deve-se posicionar com relação aos objetivos propostos. Foram ou não atingidos;
4. Avaliar o texto aumentando o nível de detalhes na parte técnica da solução.



Documento assinado eletronicamente por **Sanderley Ramos Pires, Professor do Magistério Superior**, em 21/08/2023, às 20:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Deborah Silva Alves Fernandes, Professor do Magistério Superior**, em 21/08/2023, às 20:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Karina Rocha Gomes Da Silva, Professor do Magistério Superior**, em 22/08/2023, às 08:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3980143** e o código CRC **17B8A79A**.

## RESUMO

O presente projeto teve como objetivo desenvolver e treinar um modelo de *machine learning* para a classificação de mosquitos *Aedes aegypti* e outras espécies de mosquitos, com o propósito de contribuir para o combate a doenças transmitidas por esses vetores. O processo de desenvolvimento do modelo incluiu a coleta de dados de imagens de mosquitos, a realização de pré-processamentos para o treinamento eficaz do modelo e a implementação de uma rede neural convolucional para a classificação das amostras. O conjunto de dados foi composto por 14.716 imagens de mosquitos de diversas espécies, com um desbalanceamento entre as classes. Foi aplicada a técnica de *Data Augmentation* para aumentar o número de amostras e balancear o conjunto de treinamento. O modelo foi treinado ao longo de 60 épocas, com a utilização de *callbacks* para otimizar o processo de treinamento. Os resultados mostraram uma acurácia de 75,4% na classificação das amostras, indicando um desempenho promissor do modelo. No entanto, foram observados indícios de possível *overfitting* devido à falta de amostras representativas e às limitações de recursos computacionais. Apesar dos desafios enfrentados, o modelo demonstrou a capacidade de distinguir entre as espécies de mosquitos, contribuindo para o avanço no combate ao *Aedes aegypti* e outras espécies de mosquitos. Futuros trabalhos podem explorar a expansão do conjunto de dados e a experimentação com diferentes arquiteturas de redes neurais para melhorar o desempenho do modelo. Em resumo, o projeto representa um passo importante no desenvolvimento de soluções baseadas em *machine learning* para a prevenção de doenças transmitidas por mosquitos vetores.

**Palavras-chave:** *Aedes aegypti*. Mosquito. Classificação de imagens. aprendizado de máquina. redes neurais convolucionais. doenças transmitidas por mosquitos.

## ABSTRACT

The present project aimed to develop and train a machine learning model for the classification of *Aedes aegypti* mosquitoes and other mosquito species, with the purpose of contributing to the control of diseases transmitted by these vectors. The development process of the model included the collection of mosquito image data, the implementation of preprocessing techniques for effective model training, and the implementation of a convolutional neural network for sample classification. The dataset consisted of 14,716 images of mosquitoes from various species, with an imbalance between the classes. Data Augmentation technique was applied to increase the number of samples and balance the training set. The model was trained for 60 epochs, with the use of callbacks to optimize the training process. The results showed an accuracy of 75.4% in classifying the samples, indicating a promising performance of the model. However, signs of possible overfitting were observed due to the lack of representative samples and limitations in computational resources. Despite the challenges faced, the model demonstrated the ability to distinguish between mosquito species, contributing to advancements in the control of *Aedes aegypti* and other mosquito species. Future work may explore expanding the dataset and experimenting with different neural network architectures to improve the model's performance. In summary, the project represents an important step in the development of machine learning-based solutions for the prevention of mosquito-borne diseases.

**Keywords:** Mosquito. Image classification. Machine learning. Convolutional neural networks. Mosquito-borne diseases.

## LISTA DE FIGURAS

Figura 1 - <i>Flavivirus</i> .....	10
Figura 2 - Ciclo de vida do <i>Aedes aegypti</i> .....	11
Figura 3 - Macho e Fêmea do <i>Aedes aegypti</i> .....	12
Figura 4 - Fluxo migratório do <i>Aedes aegypti</i> .....	13
Figura 5 - Correio da Manhã, edição n. 684, 26 de Abril de 1903 .....	16
Figura 6 - Operação de Convolução.....	17
Figura 7 - Operação de <i>Pooling</i> .....	18
Figura 8 - Camada totalmente conectada .....	18
Figura 9 - Pipeline do Projeto .....	19
Figura 10 - Amostra de imagens do <i>dataset</i> .....	20
Figura 11 - Estrutura de diretório do <i>dataset</i> .....	20
Figura 12 - Esquema representativo da organização do <i>dataset</i> .....	21
Figura 13 - Código de importação das imagens .....	23
Figura 14 - Trecho de código que efetua a normalização dos dados.....	23
Figura 15 - Trecho do código que implementa o <i>Autotune</i> .....	24
Figura 16 - Gráficos das funções Sigmoid e ReLU.....	24
Figura 17 - Curva de perda de treinamento e validação.....	27
Figura 18 - Curva de acurácia de treinamento e validação .....	27
Figura 19 - Matriz de Confusão do Modelo .....	29
Figura 20 - Imagens preditas pelo modelo .....	29
Figura 21 - <i>Aedes aegypti</i> e <i>Aedes albopictus</i> , respectivamente.....	30
Figura 22 - Curva ROC do modelo .....	31

## SUMÁRIO

<b>RESUMO</b> .....	3
<b>ABSTRACT</b> .....	4
<b>LISTA DE FIGURAS</b> .....	5
<b>SUMÁRIO</b> .....	6
<b>1 INTRODUÇÃO</b> .....	7
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	10
2.1 ARBOVIROSES.....	10
2.1.1 Aedes aegypti .....	10
2.1.2 Dengue.....	13
2.2 BREVE HISTÓRICO DO AEDES AEGYPTI NO BRASIL .....	13
2.3 MACHINE LEARNING .....	16
<b>3 METODOLOGIA</b> .....	19
3.1 Coleta de Dados .....	19
3.2 Pré-Processamento .....	21
3.3 Treinamento .....	24
3.4 Validação .....	25
<b>4 RESULTADOS</b> .....	27
<b>5 CONCLUSÕES</b> .....	32
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	34

# 1 INTRODUÇÃO

Goiânia é a segunda capital brasileira com o maior registro de casos notificados de dengue. De acordo com os dados contabilizados pela Secretaria de Estado da Saúde do Estado de Goiás, desde o ano de 2010 até a semana 41 do ano de 2022, o município permaneceu no topo do ranking.

Percebe-se pelos boletins epidemiológicos que as métricas utilizadas são do tipo *lagging indicators*, ou seja, são indicadores de desempenho utilizados para medir resultados passados. Nesse caso, os indicadores são quantidade de casos notificados, casos confirmados, taxa de incidência, quantidade de mortes, etc. Essas métricas fazem menção a um cenário epidemiológico em que as pessoas já estão infectadas pela doença.

A dengue ainda é uma doença grave que vem causando várias mortes e tomando muitos leitos de hospitais. Além dos altos custos médicos e hospitalares despendidos para o tratamento das pessoas infectadas, também é necessário levar em consideração os custos emocionais envolvidos em relação à falta do bem-estar do indivíduo, da família e das mortes iminentes.

A dengue é uma doença que pertence à classe das arboviroses, que são doenças causadas por vírus transmitidos principalmente por mosquitos. O *Aedes aegypti* é o mosquito vetor responsável pela transmissão da Dengue, Zika vírus e febre Chikungunya. Seu ciclo de vida é iniciado pela deposição de ovos em locais propícios ao acúmulo de água.

As principais medidas atuais de combate à dengue são feitas por controles químicos e identificação de possíveis focos, realizados pelos Agentes de Saúde, nebulização de inseticidas por bombas costais ou veiculares (fumacês), além de ações educativas que visam a conscientização da sociedade em relação à saúde coletiva e individual.

O município de Goiânia também realiza a coleta de amostras biológicas para a identificação precoce das áreas com circulação viral. Com base no histórico recorrente da dengue na sociedade e no panorama atual de tecnologia avançada em que vivemos, esse projeto tem o objetivo de construir um modelo de *Machine Learning* capaz de identificar o mosquito *Aedes aegypti* através de imagens, com base na coleta realizada pela prefeitura. Dessa forma, a integração dessa tecnologia aos métodos vigentes, proporcionará um aperfeiçoamento das técnicas utilizadas para o monitoramento do vetor. A identificação das amostras biológicas serão realizadas de forma mais eficiente e eficaz, de modo que as regiões com maior densidade do mosquito recebam medidas preventivas de maneira mais urgente e criteriosa. Assim, o combate precoce poderá evitar que esses vetores alcancem a população e, conseqüentemente, garantirá a saúde coletiva e diminuirá as ocupações em leitos hospitalares.

O projeto está organizado em 6 capítulos.

No capítulo 1, é realizada a introdução, apresentado os objetivos, o problema e a solução do projeto. É abordada a alta incidência de dengue em Goiânia e a limitação dos indicadores atuais para uma resposta eficaz. A doença tem consequências médicas, econômicas e emocionais graves, sendo transmitida pelo *Aedes aegypti*. O projeto propõe um modelo de Aprendizado de Máquina para identificar o mosquito por meio de imagens, melhorando o monitoramento e as ações preventivas. Isso poderia reduzir a disseminação do vetor, contribuindo para a saúde pública.

No capítulo 2, são abordadas as arboviroses, doenças causadas por arbovírus transmitidos por mosquitos e carrapatos. São descritas as famílias de arbovírus, com foco nas *Togaviridae* e *Flaviviridae*, que causam doenças como Chikungunya, Dengue, Zika e Febre Amarela. O vetor principal, *Aedes aegypti*, é abordado em relação a suas características, ciclo de vida e hábitos. A Dengue é discutida em termos de sintomas e imunidade. O capítulo também aborda a história do *Aedes aegypti* no Brasil, ligada à febre amarela. Além disso, é introduzido o conceito de *Machine Learning*, com foco em redes neurais convolucionais (CNNs), destacando seu uso na análise de imagens.

No capítulo 3, foi realizado o processo de desenvolvimento do modelo de classificação de mosquitos. Iniciando com a coleta de imagens de diferentes fontes, o capítulo abrange a formação dos conjuntos de treinamento, validação e teste, incluindo a aplicação de técnicas para equilibrar as amostras. Detalhes sobre o pré-processamento são apresentados, destacando a escolha das redes neurais convolucionais (CNNs), configuração do ambiente de desenvolvimento e a implementação de estratégias eficientes de carregamento e normalização. O treinamento do modelo é abordado com foco na arquitetura *Sequential* e nas funções de ativação, além de técnicas para evitar *overfitting*. A seção de validação explora métricas de desempenho como Precisão, Acurácia, Revocação e AUC-ROC, bem como o uso da matriz de confusão.

No capítulo 4, é apresentada a avaliação do modelo de classificação de mosquitos após um treinamento de 60 épocas, com destaque para a utilização de *callbacks* como o *EarlyStopping* e o *ReduceLROnPlateau*. Gráficos de perda e acurácia mostram a evolução do desempenho, indicando possíveis desafios de *overfitting*. O modelo final revelou resultados promissores, embora tenha enfrentado limitações computacionais e de amostras. Métricas como Precisão, Recall, Acurácia e AUC-ROC indicam um bom desempenho geral, sugerindo potencial para abordar o problema do *Aedes aegypti*.

No capítulo 5, é destacada a realização do estudo visando ao desenvolvimento de um modelo de Aprendizado de Máquina para identificar o mosquito *Aedes aegypti* através de imagens, com o objetivo de aprimorar a vigilância de arboviroses e prevenir sua disseminação. O capítulo aborda a análise do mosquito, técnicas de aprendizado de máquina e coleta de dados. Embora desafios como o *overfitting* sejam observados, o modelo alcançou resultados consideráveis, demonstrando potencial para o controle desses vetores. A aplicabilidade em microscópios e aplicativos é destacada, enfatizando seu uso no campo e na coleta colaborativa de dados, unindo tecnologia e conhecimento biológico para criar ambientes mais seguros e saudáveis.

Por fim, no capítulo 6 é listado as referências bibliográficas utilizadas na pesquisa.

## 2 FUNDAMENTAÇÃO TEÓRICA

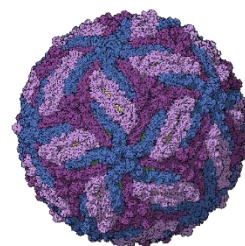
### 2.1 ARBOVIROSES

O termo “arboviroses” designa um grupo de doenças causadas pelos arbovírus, uma abreviação de *AR*thropod *BOR*ne *VIRUS*es, expressão derivada do inglês, que traduzido para o português significa *Vírus Transmitidos por Artrópodes*, especificamente os hematófagos pertencentes às classes *insecta* e *arachnida*, como por exemplo, mosquitos e carrapatos [13].

Os arbovírus são classificados em cinco famílias virais:

- *Bunyaviridae*
- *Togaviridae*
- *Flaviviridae*
- *Reoviridae*
- *Rhabdoviridae*

Figura 1 - *Flavivirus*



Fonte: Manuel Almagro Rivas / Wikimedia Commons

Dentre os arbovírus, as principais causas de patologias em seres humanos e outros vertebrados são os vírus compreendidos pelas *Togaviridae* e *Flaviviridae*. Na família *Togaviridae*, destaca-se o vírus da febre Chikungunya (CHIKV), pertencente ao gênero *Alphavirus*. Já na família *Flaviviridae*, encontram-se os vírus da Dengue (DENV), Zika (ZIKV) e Febre Amarela (YFV), pertencentes ao gênero *Flavivirus*, figura 1.

A transmissão ocorre pela picada de fêmeas dos mosquitos *Aedes aegypti* e *Aedes albopictus* infectadas pelos vírus.

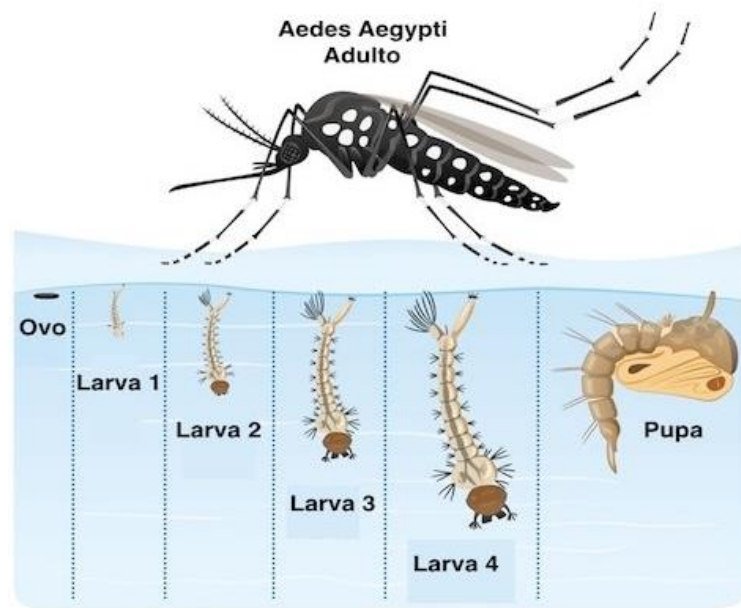
#### 2.1.1 *Aedes aegypti*

O *Aedes aegypti* é um artrópode pertencente à família *Culicidae* do gênero *Aedes*. Trata-se de uma espécie de clima tropical e subtropical, com distribuição global [12].

Vive predominantemente em ambientes urbanos, sendo mais ativo no início da manhã e no final da tarde, antes do pôr do sol [17].

Seu ciclo de vida, figura 2, constitui em uma metamorfose completa que dura de 7 a 10 dias, compreendendo quatro fases distintas: ovo, larva (quatro estágios larvários), pupa e adulto [12].

Figura 2 - Ciclo de vida do *Aedes aegypti*



Fonte: istock photo modificada

**Ovo:** Os ovos são depositados pela fêmea nas paredes dos depósitos de água, posicionando-os milímetros acima da superfície. Com a chegada da chuva, o nível da água aumenta, entrando em contato com os ovos e desencadeando o processo de eclosão, que dura em torno de 30 minutos. O desenvolvimento embrionário no interior dos ovos ocorre aproximadamente de 2 a 3 dias, dependendo da temperatura do ambiente. Esses ovos permanecem viáveis por aproximadamente 1 ano, após o completo desenvolvimento embrionário.

Os criadouros podem ser naturais, como bromélias e bambus ocas, e artificiais, como por exemplo, garrafas plásticas, tampinhas de garrafas, pneus, caixa d'água, copos descartáveis, vasos de plantas, etc.

**Larva:** Essa etapa compreende quatro estágios larvais e engloba o período de nutrição e crescimento do embrião. A duração dessa fase varia conforme a temperatura, a disponibilidade de alimento e a densidade de larvas no criadouro. Sob condições favoráveis, pode-se concluir em até 5 dias, no entanto, em circunstâncias menos propícias, especialmente no quarto estágio larval, essa fase pode estender-se por várias semanas.

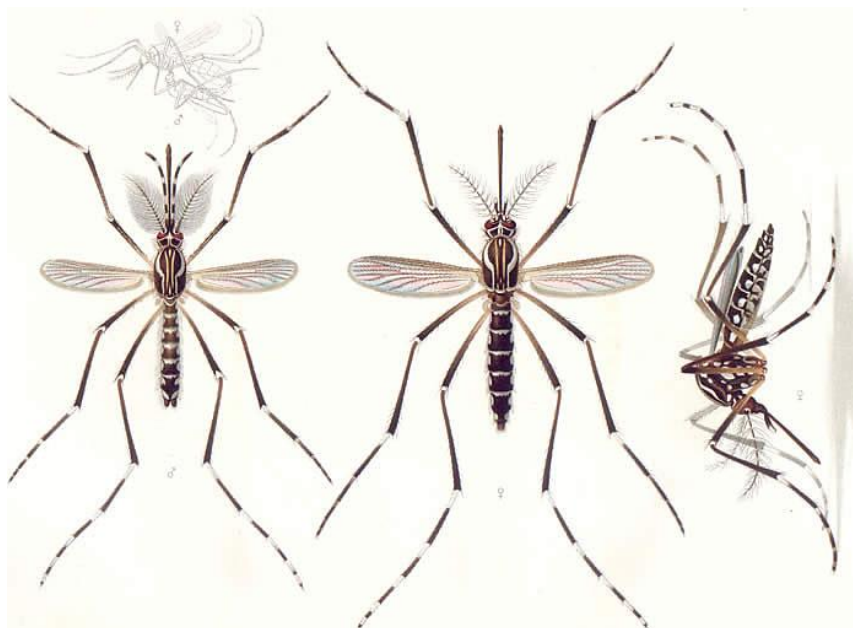
As larvas são sensíveis a movimentos bruscos na água e à luminosidade, respondendo com movimentos rápidos e buscando refúgio no fundo do reservatório de água em que ele se encontra.

**Pupa:** Fase metamórfica que marca a transição da etapa larval para a forma alada. Durante esse período, não ocorre alimentação, sendo a respiração conduzida através das trompas. A pupa fica disposta na superfície, de forma a facilitar a ecdise do inseto adulto. Essa fase dura aproximadamente cerca de dois a três dias.

**Adulto:** Consiste na forma alada e reprodutora do mosquito, representando importante fase de sua dispersão e perpetuação da espécie. O tempo de vida dura em média de 30 a 45 dias, e a fêmea copula apenas uma vez durante o seu curto período de vida.

O mosquito mede cerca de 1 centímetro, tem uma cor escura, possui manchas brancas no abdômen e pernas, e um formato de uma lira no dorso do torác. A distinção entre machos e fêmeas é notadamente caracterizada pela existência de antenas plumosas e palpos mais alongados nos machos, além de uma ligeira redução de tamanho comparado às fêmeas, figura 3.

Figura 3 - Macho e Fêmea do *Aedes aegypti*



Fonte: E. A. Goeldi (1905) / Wikimedia Commons

Se alimentam da seiva das plantas, entretanto, a fêmea necessita de sangue, tendo predileção pelo humano, para que ocorra a maturação dos seus ovos.

O período de maior densidade populacional da espécie ocorre durante o verão, por ser uma época de maior pluviosidade, o que resulta em um aumento significativo de locais propícios para a fêmea depositar seus ovos [14]. Além disso, as temperaturas mais elevadas aceleram o ciclo de desenvolvimento do mosquito, abreviando as etapas de ovo, larva e chegando até a fase adulta.

O *Aedes aegypti* é o principal vetor transmissor da Dengue, Zika e Chikungunya, possuindo o *Aedes albopictus* como vetor secundário.

### 2.1.2 Dengue

A dengue é uma doença infecciosa febril aguda, podendo manifestar-se em formas leves ou severas e possui quatro sorotipos diferentes (**DENV-1, DENV-2, DENV-3 e DENV-4**) [13]. A infecção por um determinado sorotipo proporciona imunidade permanente a esse sorotipo específico, e adicionalmente, induz a uma imunidade cruzada parcial e temporária aos demais tipos.

O ciclo viral da doença tem início quando o vetor transmissor, o *Aedes aegypti* fêmea, pica uma pessoa infectada [16]. O vírus passa por um processo de multiplicação no intestino do vetor e infecta diversos tecidos, até alcançar as glândulas salivares do mosquito. Após o repasto de sangue infectado, e um período de incubação extrínseca que pode variar de 8 a 12 dias, o ciclo de transmissão se inicia quando o mosquito infectado se alimenta do sangue de outros indivíduos humanos. O ciclo se encerra apenas com a morte do mosquito infectado [18].

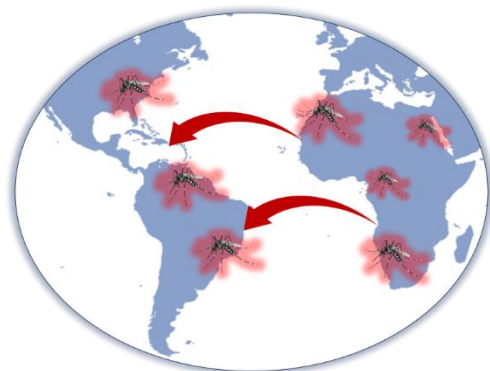
Os sintomas da doença podem variar de sintomáticos a assintomáticos. Geralmente, a primeira manifestação apresenta-se como uma febre alta que varia entre 39°C e 40°C, podendo perdurar por um período de 2 a 7 dias [14]. Esse quadro é acompanhado por cefaleia intensa, dores musculares e articulares, sensação de prostração, fraqueza, desconforto ocular, erupções cutâneas e prurido. Na fase mais grave da enfermidade, caracterizada como febre hemorrágica, os sintomas abrangem dor abdominal aguda e constante, náuseas persistentes, entre outras manifestações, podendo culminar no óbito do indivíduo.

## 2.2 BREVE HISTÓRICO DO AEDES AEGYPTI NO BRASIL

O *Aedes aegypti* é originário do continente africano e se alastrou pelas regiões tropicais e subtropicais do planeta durante as expansões marítimas. Foi introduzido nas Américas durante o período colonial pelos navios negreiros [7].

No Brasil, os primeiros surtos de febre amarela são datados do final do século XVII nas regiões de Pernambuco e Bahia, ficando inerte por cerca de 150

Figura 4 - Fluxo migratório do *Aedes aegypti*



anos e ressurgindo nas últimas décadas do século XIX, na região do Rio de Janeiro. O país se tornou bastante propício à proliferação do mosquito, pois as cidades sofreram rápidas expansões, aumentando a densidade populacional e os constantes fluxos migratórios e imigratórios, além das condições climáticas favoráveis para a sua reprodução.

A doença da febre amarela já havia sido diagnosticada pela primeira vez em 1849 na Bahia, quando eclodiu a sua segunda epidemia, e teria se propagado do norte ao sul do país, alcançando a então capital Rio de Janeiro. Entre os finais do século XIX e início do século XX, a população da região foi gravemente acometida pela doença.

Nesse período, a capital já havia grandes concentrações urbanas, principalmente no centro da cidade. As condições sanitárias não conseguiram acompanhar o exponencial crescimento da capital e, conseqüentemente, a população começou a ser assolada pelos grandes e constantes surtos de doenças infecciosas, como a cólera, varíola, peste bubônica, malária, febre tifoide e a própria febre amarela.

Durante o mandato do prefeito Francisco Pereira Passos, nomeado pelo então presidente Rodrigues Alves (1902-1906), várias reformas de modernização urbana e saneamento básico foram implantadas, a fim de se eximir a reputação do Rio de Janeiro como uma cidade insalubre e “túmulo dos estrangeiros”, colocando-a aos moldes da *Béle Époque* [1].

A política do “Bota-Baixo” [15], como ficou conhecido às ações de modernização urbana iniciada pelo prefeito, acarretou na demolição de diversas habitações coletivas, como cortiços, casas de cômodos, estalagens e velhos casarões. Milhares de famílias pobres foram desabrigadas e despachadas para longe do centro da cidade, como morros e periferias. Em consonância, as reformas sanitárias foram lideradas pelo médico e sanitarista Oswaldo Cruz (1903-1909), que foi designado para a Diretoria Geral de Saúde Pública, a fim de combater as doenças epidêmicas desse período.

Em relação à febre amarela, ainda havia bastante carência de conhecimento sobre as formas de disseminação e métodos de profilaxia para o seu combate. A sociedade científica era dividida entre aqueles que acreditavam que a febre amarela era uma doença contagiosa, causada predominantemente por microrganismos, e entre outros, que acreditavam ser uma doença infecciosa [1].

Em 1881 surge o primeiro relato internacional sobre a forma de disseminação da febre amarela, representado pelo médico cubano Juan Carlos Finlay na Conferência Sanitária Internacional, realizada em Washington. Em sua pesquisa, intitulada como “*O mosquito hipoteticamente considerado como agente transmissor da febre-amarela*”, Finlay declarou que o

mosquito da espécie *Aedes aegypti*, conhecido primeiro como *Culex fasciatus* e, posteriormente, como *Stegomyia fasciata*, era o hospedeiro intermediário do patógeno da doença, sendo o principal responsável pela sua transmissão [4].

Suas pesquisas não foram aclamadas logo de início pela comunidade científica, mas Finlay continuou persistindo na sua teoria, e até 1900 veio aplicando aos seus experimentos o tanto exigido rigor científico, com testes sistematicamente planejados. Durante o processo, o médico obteve suporte de uma comissão de saúde do Exército norte-americano liderada por Walter Reed, que contribuiu com pesquisas sobre a etiologia e a profilaxia da febre amarela.

Finalmente, um ano mais tarde, após a confirmação de sua teoria, seus resultados foram apresentados ao Congresso Médico Pan-Americano realizado em Havana [4]. No mesmo ano, medidas de combate ao mosquito foram implantadas na região, e a doença da febre amarela, que até então estava difundida e dizimando a população há 200 anos, foi dada como erradicada.

A teoria havanesa desembarcou em solo brasileiro durante o seu momento mais crítico, conquistando dois dos médicos mais renomados da história do país. Em São Paulo, atuava o médico sanitarista Emílio Ribas à frente da Diretoria do Serviço Sanitário do estado, que teve contato direto com os experimentos implantados em Havana. Ribas foi o pioneiro na luta contra a febre amarela no Brasil, executando com determinação os trabalhos realizados por Finlay e pela comissão norte-americana, acarretando na erradicação da doença no estado em 1903.

Nesse mesmo ano, Oswaldo Cruz trabalhava na maior reforma sanitária do país, assumindo o compromisso de erradicar a doença no Rio de Janeiro em um prazo de três anos [6]. Para isso, criou o Serviço de Profilaxia Específica da Febre Amarela, concebido de acordo com os padrões tipicamente militares. A cidade foi dividida em dez distritos sanitários, cada qual chefiado por um delegado de saúde, responsável por monitorar os casos notificados, aplicar multas e intimar os donos a reformar ou demolir os imóveis considerados insalubres [9]. As brigadas sanitárias, ou popularmente chamadas de ‘mata-mosquitos’, eram constituídas ainda por médicos, inspetores sanitários, acadêmicos de medicina, auxiliares de limpeza, entre outros.

As ações implantadas durante a reforma sanitária para o combate da febre amarela foram bastante invasivas, violentas e autoritárias. As brigadas sanitárias adentravam os domicílios sem autorização prévia dos inquilinos, internavam compulsoriamente os enfermos com traços da moléstia, interditavam e demoliam habitações consideradas insalubres [2].

Os principais métodos de profilaxia consistiam em isolar o doente, com o objetivo de interromper o ciclo de vida da doença e evitar que novos mosquitos vetores fossem contaminados. O isolamento era realizado transportando os doentes para hospitais ou isolando-os em suas

próprias residências [5]. No último caso, a cama do paciente era protegida por uma tela, a casa inteira era vedada com imensos panos de algodão. Em seguida, era espalhado um pó de piretro no local.

Devia-se também proteger pessoas dos mosquitos já infectados. Assim, as ruas eram percorridas a fim de identificar focos de mosquitos, depositando uma combinação de piretro e querosene em águas paradas. As casas eram expurgadas com a queima de enxofre, com o objetivo de eliminar os mosquitos na sua fase adulta, e nas áreas públicas, o gás Clayton era injetado nas redes de esgoto [5].

Por fim, Oswaldo iniciou campanhas de conscientização para a população conhecidas como “Conselhos ao Povo”, por meio de publicações na imprensa e na distribuição de folhetos avulsos pela cidade, como visto na figura 5 [3]. O objetivo era orientar as pessoas sobre a etiologia da doença, os métodos de contaminação e a profilaxia, com o propósito de cercear do imaginário coletivo a crença de que a doença fosse contagiosa, ou seja, transmitida pelo contato direto com objetos contaminados e secreções dos enfermos [19]. Assim, a reforma sanitária se tornaria mais acolhedora e assertiva por parte da população e outros médicos intransigentes.

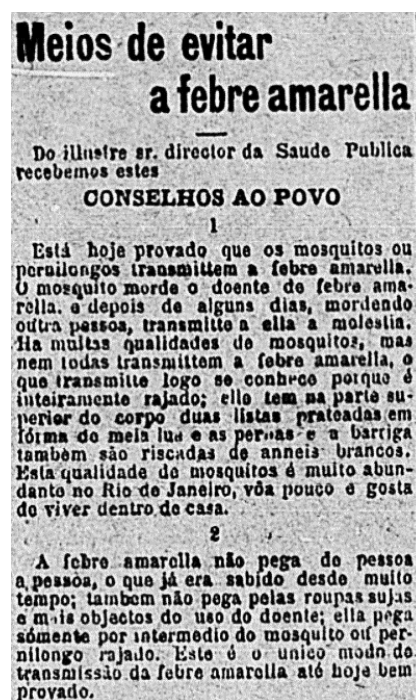
As ações de Oswaldo Cruz começavam a surtir efeitos práticos com a diminuição dos índices de mortalidade causada pela doença. E no ano de 1907, o grande saneador anunciou o fim da epidemia da febre amarela na então capital Rio de Janeiro, o que lhe consagrou reconhecimento internacional, sendo agraciado com uma medalha de ouro no 14º Congresso Internacional de Higiene e Demografia de Berlim, na Alemanha [2].

## 2.3 MACHINE LEARNING

*Machine learning* (Aprendizado de Máquina) é uma área da inteligência artificial que se concentra no desenvolvimento de sistemas computacionais que são capazes de aprender padrões e tomar decisões com base em dados de forma autônoma, usando para isso as redes neurais artificiais e o *Deep Learning* [8].

*Deep Learning* compreende uma vertente do aprendizado de máquina que se baseia em redes neurais artificiais compostas por múltiplas camadas. Quando adicionadas camadas de

Figura 5 - Correio da Manhã, edição n. 684, 26 de Abril de 1903



Fonte: Jornal Correio da Manhã

convolução, essas redes evoluem para o que se conhece como redes neurais convolucionais, também denominadas de *Convolutional Neural Networks (ConvNets ou CNNs)*. As CNNs são particularmente eficazes em tarefas de visão computacional, como reconhecimento de objetos, classificação de imagens e detecção de padrões, devido à sua capacidade de aprender e extrair características significativas das imagens.

Normalmente, uma CNN é constituída por três camadas principais:

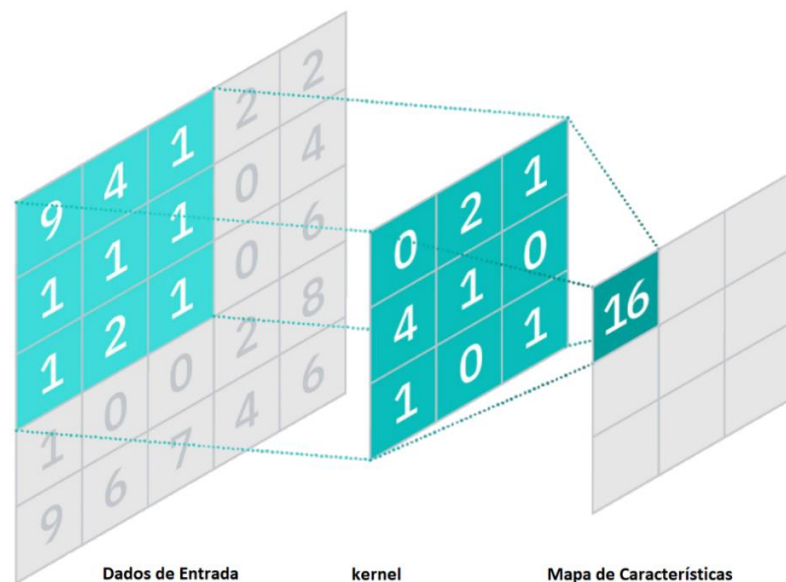
**Camada de Convolução:** Responsável por extrair características significativas das imagens, aplicando operações de convolução aos dados de entrada.

A convolução consiste em operações de multiplicação e soma, envolvendo uma sobreposição de pequenas regiões comumente denominadas de filtros ou kernels, da imagem de entrada com o objetivo de calcular produtos entre os valores dos pixels e os valores correspondentes no filtro. Essa operação ajuda a destacar características como bordas, texturas e outros detalhes em diferentes partes da imagem.

Por exemplo, com base na figura 6, o valor 16 do mapa de características é obtido a partir da seguinte operação de convolução:

$$16 = (9 \times 0) + (4 \times 2) + (1 \times 1) + (4 \times 1) + (1 \times 1) + (0 \times 1) + (1 \times 1) + (0 \times 2) + (1 \times 1)$$

Figura 6 - Operação de Convolução

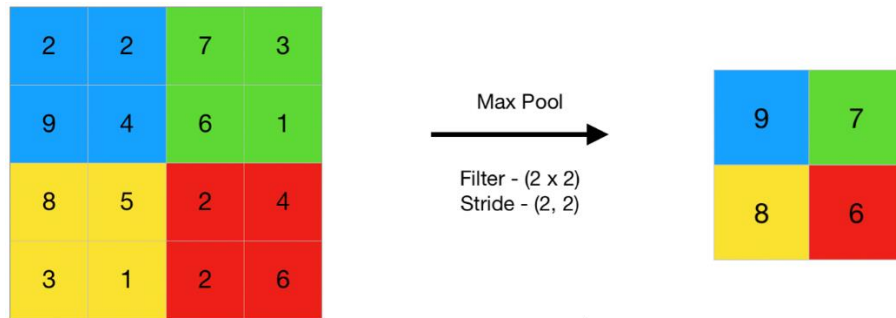


Fonte: IBM

**Camada de Pooling:** Tem o objetivo de reduzir as dimensões dos dados recebidos da camada de convolução, aplicando operações de agrupamento em regiões previamente definidas,

resultado em um único valor característico. O método mais comumente utilizado é o max pooling, no qual a regra estabelece a captura do valor mais elevado do mapa de características, como pode ser visto na figura 7.

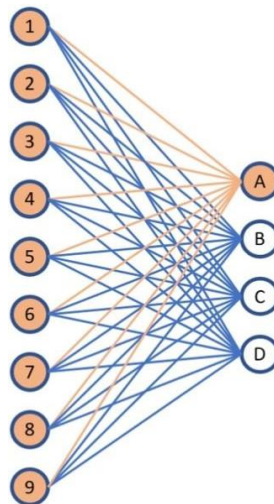
Figura 7 - Operação de *Pooling*



Fonte: Google Images

**Camada Densa:** conhecida também como camada totalmente conectada, é responsável por realizar a classificação do dado, como ilustrado na figura 8.

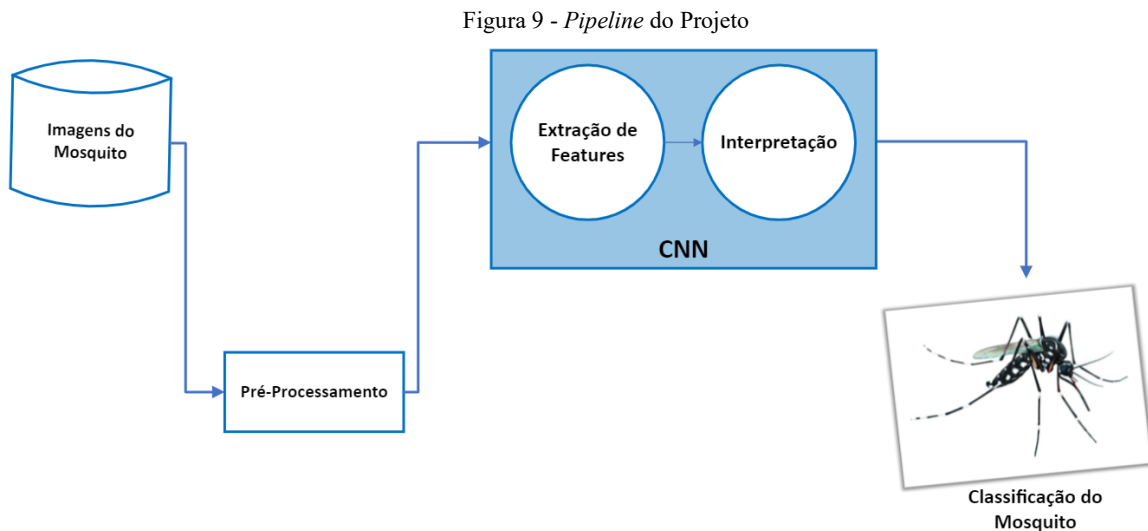
Figura 8 - Camada totalmente conectada



Fonte: Diego Unzueta

### 3 METODOLOGIA

O processo de desenvolvimento do modelo começa com a aquisição das imagens dos mosquitos. Em seguida, são executados pré-processamentos essenciais para assegurar um treinamento eficaz do modelo. A etapa subsequente envolve a implementação do próprio modelo, cujo resultado será a classificação das amostras em duas categorias: *Aedes* e Mosquitos. O pipeline é esquematizado na figura 9.



#### 3.1 Coleta de Dados

A etapa primordial para o desenvolvimento de um modelo de *machine learning* consiste na coleta de dados.

As imagens dos mosquitos foram extraídas de *datasets* públicos localizados na internet<sup>1</sup>, do google imagens e de sites especializados em classificação de insetos.

O *dataset* do projeto constitui em um total de 14.716 imagens de mosquitos, englobando diversas espécies, entre elas o *Aedes aegypti*, *Aedes albopictus*, *Culex quinquefasciatus*, *Anopheles*, entre outros. A figura 10 ilustra algumas amostras do *dataset*.

<sup>1</sup> Kaggle, Google imagens, INaturalist, Gettyimages, CDC, iStockPhoto

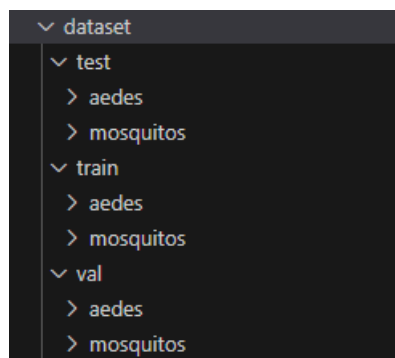
Figura 10 - Amostra de imagens do *dataset*



Fonte: Google Imagens

As imagens foram divididas em três *datasets* localizados em três diretórios distintos, conforme mostrado no esquema da figura 11:

Figura 11 - Estrutura de diretório do *dataset*



Os dados de treino contabilizam um total de 12.316 imagens, sendo 6.158 imagens da espécie *Aedes aegypti* e 6.158 imagens para o restante das espécies.

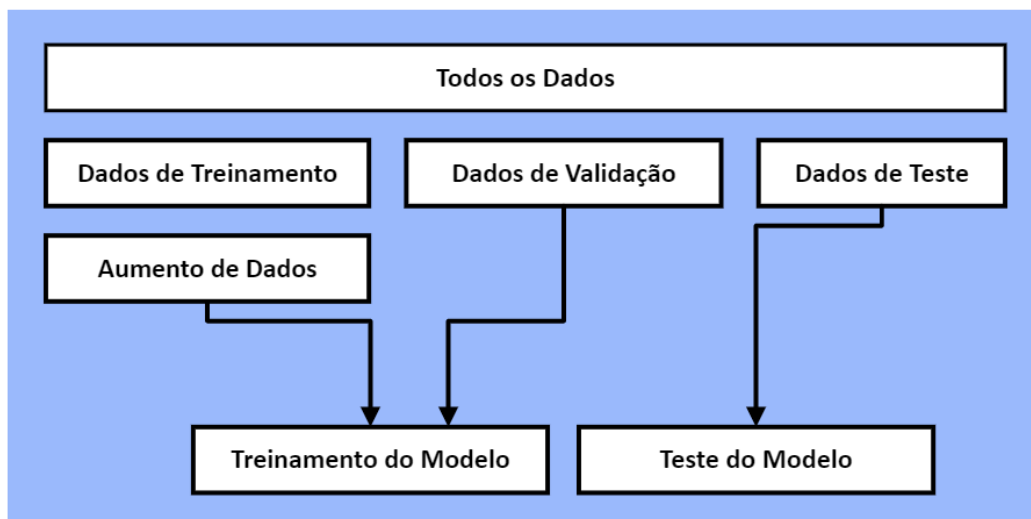
Os dados de validação estão distribuídos num total de 600 imagens para cada classe, assim como os dados de teste.

Para os dados de treinamento, ocorreu um desbalanceamento pela carência de imagens de mosquitos *Aedes aegypti*. Isso é um problema, pois pode acarretar no enviesamento dos resultados para a classe dominante [11]. Ou seja, o modelo pode receber mais informações de uma classe do que de outra e, assim, dificultar a generalização para novos dados. Esse problema foi corrigido utilizando a técnica de *Data Augmentation*, que consiste basicamente em gerar novas amostras a partir das existentes através de transformações geométricas e de cor.

Dentre as possíveis transformações disponíveis para a aplicação dessa técnica, a de rotação foi utilizada. Assim, para o aumento dos dados, as imagens foram rotacionadas em 90°, 180° e 270° e salvas juntamente com as outras amostras originais, alcançando o limite para o balanceamento.

A organização e divisão dos dados é ilustrada na figura 12.

Figura 12 - Esquema representativo da organização do *dataset*



### 3.2 Pré-Processamento

Após a construção do conjunto de dados, deu-se início à implementação do modelo de *machine learning*. A opção pela arquitetura recaiu sobre as redes neurais convolucionais, escolha pautada em sua ampla aplicabilidade e comprovada eficácia em tarefas de visão computacional. Os recursos e infraestrutura disponíveis também foram fundamentais para a tomada de decisão.

O treinamento do modelo ocorreu em um notebook equipado com 8 GB de RAM, um processador AMD Ryzen 7 4800H, e uma GPU Nvidia GeForce GTX 1650 Ti, contendo 4 GB de memória dedicada. Optou-se pelo Python como a linguagem de programação, devido à proficiência no domínio e sua ampla adoção na análise de dados, beneficiada por uma vasta gama de bibliotecas que simplificaram o desenvolvimento.

A capacidade de processamento da GPU foi requisitada para o treinamento do modelo. Dessa forma, foi necessário configurar o ambiente de desenvolvimento de maneira precisa, incluindo a instalação do Python versão 3.9, TensorFlow versão abaixo de 2.11, cudatoolkit versão 11.2, e cudnn versão 8.1.0. A configuração apropriada do ambiente de desenvolvimento é crucial para garantir a correta funcionalidade e desenvolvimento do modelo.

Os métodos de carregamento de *datasets* comumente utilizados para tarefas de *machine learning* não funcionaram para o projeto. Por ser um conjunto de dados próprio, ou seja, dados que não são nativos de bibliotecas, e consideravelmente grande, eles tiveram que ser carregados em lotes, utilizando uma função específica que o tensorflow fornece.

```
tensorflow.keras.utils.image_dataset_from_directory()
```

Essa função carrega as imagens do diretório e cria um objeto de conjunto de dados TensorFlow que pode ser facilmente utilizado para treinar e avaliar modelos de aprendizado profundo. Além de lidar automaticamente com a leitura, pré-processamento e carregamento de imagens de forma eficiente. Isso economiza tempo e esforço durante a preparação dos dados para treinamento. Para o projeto, conforme mostrado no código 1 da figura, os parâmetros configurados para essa função foram:

**data\_path\_train:** caminho para o diretório onde se encontra os dados.

**labels:** Método para inferir as etiquetas das classes.

**label\_mode:** Tipo de rótulo ('int' para rótulos inteiros, 'categorical' para rótulos categóricos, etc.)

**color\_mode:** Modo de cor ('grayscale', 'rgb', etc.)

**batch\_size:** Tamanho do lote para carregar as imagens.

**image\_size:** Tamanho das imagens após o redimensionamento.

**shuffle:** Misturar aleatoriamente as imagens no conjunto de dados.

**seed:** Semente aleatória para embaralhamento.

Para um **batch\_size=32**, a função carrega as primeiras 32 imagens do diretório como um lote, processa essas imagens e as torna disponíveis para treinamento ou validação. Quando o modelo termina de processar esse lote, a função carrega o próximo lote de 32 imagens e assim por diante.

O carregamento em lote oferece várias vantagens:

1. **Eficiência de memória:** Em vez de carregar todas as imagens de uma vez, o carregamento em lote permite trabalhar com quantidades gerenciáveis de dados por vez, evitando o esgotamento da memória.
2. **Velocidade de treinamento:** permite que a GPU execute cálculos em paralelo enquanto o próximo lote é carregado, o que acelera o treinamento.
3. **Processamento em paralelo:** permite que o processamento das imagens ocorra em paralelo, melhorando a eficiência do treinamento.

4. **Adaptabilidade:** permite ajustar o tamanho do lote para caber na memória disponível, se a memória da GPU for limitada.

#### Código 1:

Figura 13 - Código de importação das imagens

```
data_val = tf.keras.utils.image_dataset_from_directory(  
    data_path_val,  
    labels='inferred',  
    label_mode='int',  
    color_mode='rgb',  
    batch_size=batch_size,  
    image_size=(img_height, img_width),  
    shuffle=True,  
    seed=42  
)
```

Considerando a natureza das imagens em formato RGB, optou-se por seguir a prática convencional de normalização. Nesse processo, os valores dos pixels foram divididos por 255, resultando na conversão dos mesmos em números de ponto flutuante, com intervalo entre 0 e 1. O código 2 da figura 14, mostra o processo para os três *datasets*: treino, validação e teste, respectivamente.

#### Código 2:

Figura 14 - Trecho de código que efetua a normalização dos dados

```
data_train = data_train.map(lambda image, label: (tf.cast(image, tf.float32) / 255.0, label))  
data_val = data_val.map(lambda image, label: (tf.cast(image, tf.float32) / 255.0, label))  
data_test = data_test.map(lambda image, label: (tf.cast(image, tf.float32) / 255.0, label))
```

Com a preocupação de otimizar recursos computacionais, foi adotado o parâmetro AUTOTUNE do tensorflow, comumente utilizado para otimizar o desempenho do carregamento dos dados durante o treinamento de modelos de aprendizado de máquina. O TensorFlow tenta automaticamente ajustar o número de operações de carregamento de dados em paralelo com base na capacidade do sistema e na quantidade de recursos disponíveis, como a memória da GPU e da CPU. Isso ajuda a otimizar o uso desses recursos e a acelerar o treinamento do modelo. A técnica foi aplicada para os três *datasets*: treinamento, validação e teste, respectivamente, conforme mostrado no código 3 da figura 15.

### Código 3:

Figura 15 - Trecho do código que implementa o *Autotune*

```
AUTOTUNE = tf.data.AUTOTUNE
data_train = data_train.shuffle(1000).cache().prefetch(buffer_size=AUTOTUNE)
data_val = data_val.cache().prefetch(buffer_size=AUTOTUNE)
data_test = data_test.cache().prefetch(buffer_size=AUTOTUNE)
```

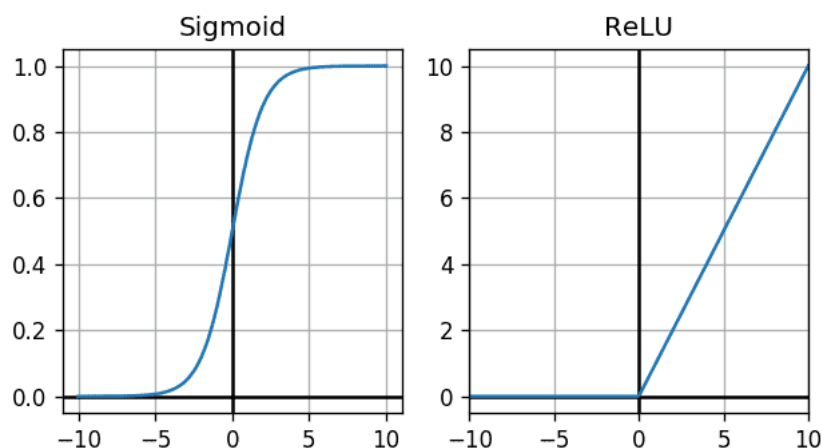
### 3.3 Treinamento

Optou-se pela utilização do modelo *Sequential*, visando a criação de uma estrutura simples e de arquitetura direta. O modelo é composto por duas camadas densas, cada uma contendo 16 filtros de dimensão 3x3. Nessa configuração, não há aplicação de tratamento de borda, e o deslocamento do filtro ocorre de maneira incremental, um pixel de cada vez.

Inclui duas camadas de *pooling* de dimensão 2x2, com um passo de deslocamento de dois pixels a cada iteração. Uma camada de *Flatten* é adotada para transformar os dados multidimensionais de entrada em um vetor unidimensional, permitindo sua propagação às camadas densas subsequentes. Por último, são incorporadas duas camadas densas, uma com 128 neurônios e a seguinte com um único neurônio. Além disso, a técnica de regularização L2 foi adicionada de forma a evitar o *overfitting* do modelo. Essa técnica consiste em adicionar um termo à função de perda durante o treinamento, que penaliza os pesos maiores na rede neural.

As funções de ativação utilizadas foram ReLu e Sigmoid, mostradas na figura 17.

Figura 16 - Gráficos das funções Sigmoide e ReLU



A função ReLU ativa a saída se a entrada for positiva e desativa (zero) a saída se a entrada for negativa. A função é descrita em (1), onde  $x$  é a entrada e zero é a saída, caso  $x$  seja menor ou igual a zero, e o próprio valor de  $x$ , se  $x$  for maior que zero.

$$f(x) = \max(0, x) \quad (1)$$

A função sigmoid produz uma resposta suave e gradual para diferentes valores de entrada. Na equação (2) temos que, à medida que a entrada,  $x$ , se torna positiva, a saída se aproxima de 1, e à medida que a entrada se torna negativa, a saída se aproxima de zero.

$$f(x) = 1 / (1 + \exp(-x)) \quad (2)$$

É bastante relevante para problemas de classificação binária, nos quais busca-se avaliar a probabilidade de pertencimento a uma das duas classes.

### 3.4 Validação

As métricas utilizadas para monitorar o desempenho do modelo são:

**Precisão (*Precision*):** A proporção de verdadeiros positivos em relação a todos os exemplos classificados como positivos. Mede a capacidade do modelo de evitar falsos positivos.

$$PRECISION = \frac{TP}{TP + FP} \quad (3)$$

**Acurácia (*Accuracy*):** A proporção de predições corretas em relação ao total de predições.

$$ACCURACY = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

**Revocação ou Sensibilidade (*Recall* ou *Sensibility*):** A proporção de verdadeiros positivos em relação a todos os exemplos que realmente são positivos. Mede a capacidade do modelo de identificar todos os exemplos positivos.

$$RECALL = \frac{TP}{TP + FN} \quad (5)$$

**Área sob a Curva ROC (AUC-ROC):** Mede a capacidade do modelo de distinguir entre as classes. É uma representação gráfica da taxa de verdadeiros positivos em relação à taxa de falsos positivos para diferentes limiares de classificação [10]. Quanto mais a curva ROC se desloca para o canto superior esquerdo do gráfico, melhor é o

desempenho do modelo, indicando que ele está alcançando altas taxas de verdadeiros positivos com baixas taxas de falsos positivos.

**Matriz de Confusão:** Indica a quantidade de ocorrências que o programa teve para cada uma das quatro categorias:

**Verdadeiro positivo (VP):** número de exemplos corretamente classificados como positivos pelo modelo.

**Verdadeiro negativo (VN):** O número de exemplos corretamente classificados como negativos pelo modelo.

**Falso positivo (FP):** O número de exemplos incorretamente classificados como positivos pelo modelo, quando na verdade são negativos.

**Falso negativo (FN):** O número de exemplos incorretamente classificados como negativos pelo modelo, quando na verdade são positivos.

## 4 RESULTADOS

O processo de treinamento do modelo foi realizado ao longo de 60 épocas, sendo interrompido na 32ª época de acordo com uma condição de parada estabelecida por meio de um mecanismo de *callback*, que é uma função que será executado quando um evento ocorrer ou quando uma determinada condição for atendida. O *EarlyStopping* foi configurado para monitorar a acurácia de validação, encerrando o treinamento quando essa métrica não demonstrou melhorias por um período consecutivo de 10 épocas.

O *callback ReduceLROnPlateau* desempenhou um papel crucial na otimização do treinamento, permitindo um ajuste dinâmico da taxa de aprendizagem (*learning rate*) do otimizador. Caso a perda de validação não apresentasse melhoras durante um intervalo de 6 épocas seguidas, esse *callback* reduzia automaticamente a taxa de aprendizagem em um fator de 0.1. Para evitar que a taxa de aprendizagem fosse excessivamente diminuída, foi estabelecido um limite mínimo de 0.0001.

Essa combinação de *callbacks* contribuiu para um treinamento mais eficiente do modelo, controlando o número de épocas e ajustando a taxa de aprendizagem de acordo com as mudanças no desempenho do treinamento e validação.

Encerrado o treinamento, é possível visualizar o desempenho do modelo pelas curvas de perda e acurácia, conforme a seguir.

Figura 17 - Curva de perda de treinamento e validação

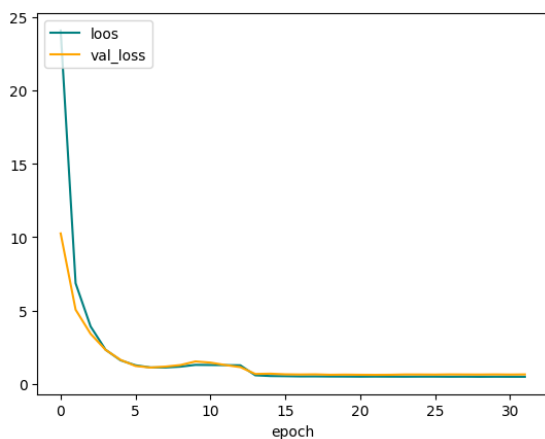
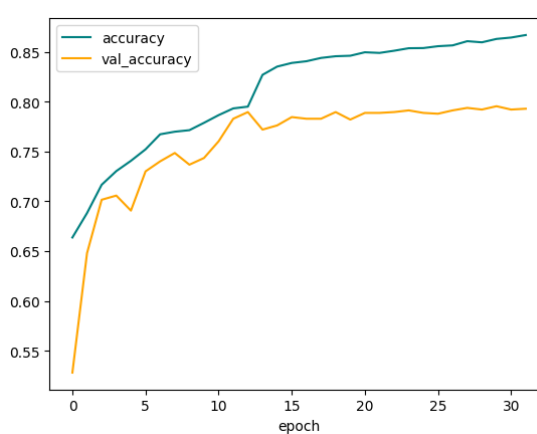


Figura 18 - Curva de acurácia de treinamento e validação



O gráfico de perda (Loss), figura 17, fornece uma representação visual da evolução da função de perda tanto durante o treinamento (em azul) quanto para o conjunto de validação (em laranja). A função de perda é um indicador que quantifica a capacidade do modelo de se ajustar aos dados durante o processo de treinamento.

Ao analisar as épocas, é notável uma redução gradual da perda para ambos os conjuntos, evidenciando o aprimoramento do desempenho do modelo. Contudo, é comum observar flutuações ou oscilações na curva de perda ao longo do treinamento.

Em torno da época 14, a curva atinge um patamar, sugerindo que o modelo pode estar estagnando em seu progresso e não obtendo melhorias significativas. Essa situação pode indicar a possibilidade de *overfitting*, no qual o modelo memoriza os dados de treinamento em vez de generalizar para novos dados.

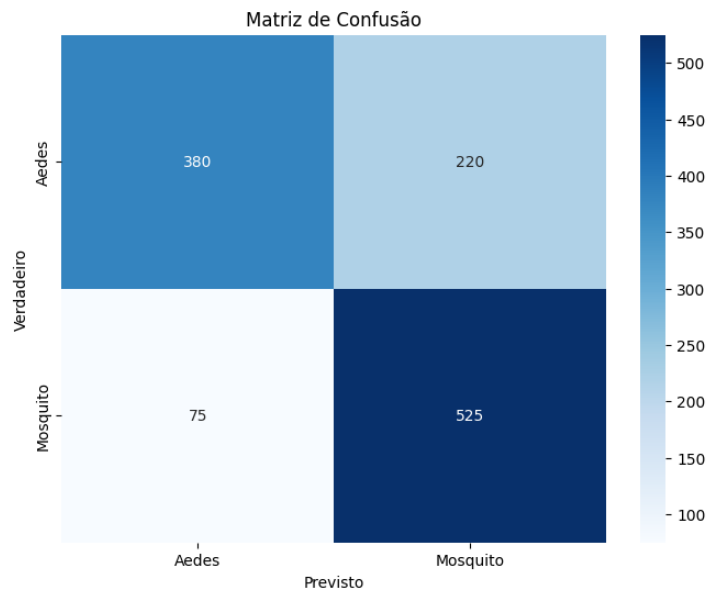
Examinando a curva de acurácia, figura 18, torna-se evidente que a partir da época 14, o processo de aprendizado do modelo se estabiliza. Observa-se um contínuo aumento da acurácia no conjunto de treinamento, porém, ela começa a se estagnar no conjunto de validação. Isso sugere que o modelo pode estar excessivamente adaptado aos dados de treinamento, mas não está conseguindo generalizar de maneira adequada para novos dados.

Foram empreendidas múltiplas tentativas de treinamento, no entanto, todas apresentaram indícios de possível *overfitting*. Uma explicação plausível reside na insuficiência de amostras, na presença de imagens pouco representativas e nas limitações de capacidade computacional, o que dificultou o treinamento de modelos mais complexos. Ao expandir o número de camadas e explorar diferentes hiperparâmetros, o limite de memória da GPU era alcançada, resultando na incapacidade da máquina de executar o programa.

Foram realizados testes em outros ambientes, como o Google Colab, entretanto, os mesmos cenários se manifestaram, uma vez que os recursos disponíveis na conta gratuita são notavelmente restritos para a execução de modelos mais elaborados. Além disso, há um limite de utilização da GPU, resultando na interrupção do acesso a esse recurso quando tal limite é alcançado.

Depois de várias tentativas, o presente modelo foi o mais bem comportado. A figura 19 mostra a sua matriz de confusão:

Figura 19 - Matriz de Confusão do Modelo

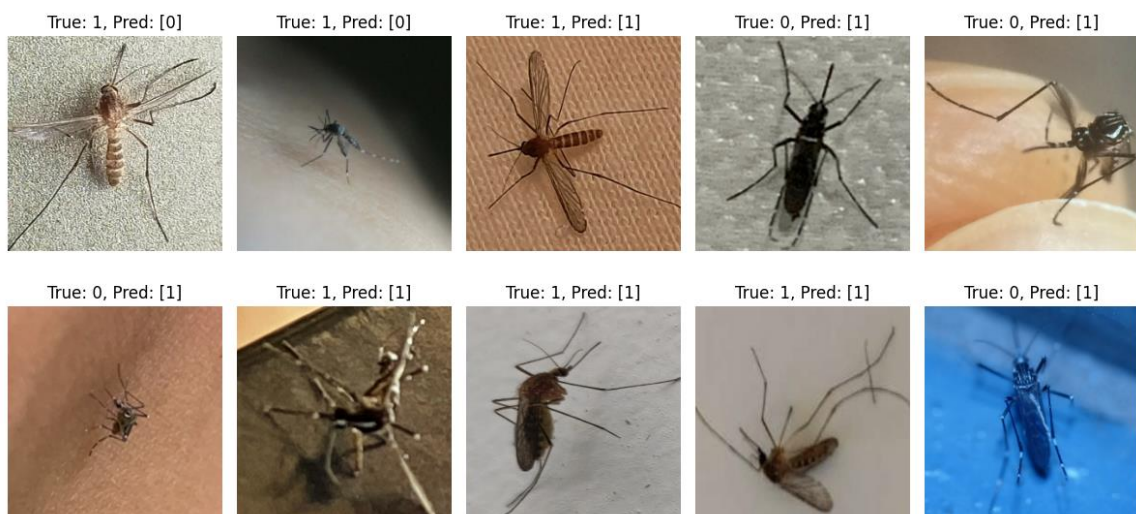


A CNN inferiu que 75 amostras eram *Aedes*, quando na verdade eram mosquitos, assim como inferiu que 220 eram mosquitos, quando na verdade eram *Aedes*.

Na figura 20, é possível visualizar as predições realizadas pelo modelo, onde:

- True = classe verdadeira
  - 0 = Aedes
  - 1 = Mosquito
- Pred = classe predita
  - 0 = Aedes
  - 1 = Mosquito

Figura 20 - Imagens preditas pelo modelo



É razoável conjecturar que essas confusões ocorreram devido a traços similares entre as espécies, especialmente no caso do *Aedes albopictus*, que guarda notável semelhança com o *Aedes aegypti*. Um exemplo ilustrativo pode ser visto na figura 21.

Figura 21 - *Aedes aegypti* e *Aedes albopictus*, respectivamente



Fonte: Google Imagens

Em relação aos outros indicadores, obteve-se os seguintes resultados:

**Precision:** 0.7046979665756226

**Recall:** 0.875

**Accuracy:** 0.7541666626930237

**ROC Curve:** 0.8569180369377136

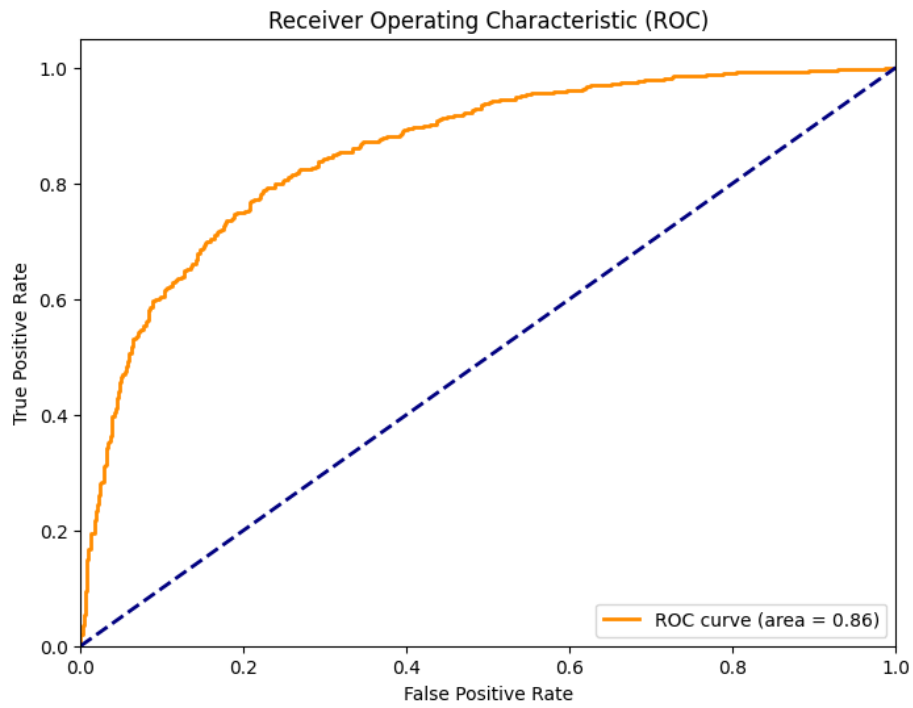
Referente a precisão, cerca de 70,4% das previsões positivas feitas pelo modelo estavam realmente corretas.

Quanto ao *Recall*, cerca de 87,5% dos exemplos positivos foram corretamente identificados pelo modelo.

A taxa de acurácia indica que cerca de 75,4% de todas as previsões estavam corretas.

Ao examinarmos a curva ROC, figura 22, notamos que a área sob a curva atingiu um valor de 0,86. A proximidade desse valor a 1 sugere que o modelo apresenta uma melhor capacidade de distinguir entre as classes.

Figura 22 - Curva ROC do modelo



Apesar das implicações e restrições, é evidente que o modelo apresentou resultados promissores. Há espaço para explorar outras técnicas visando aprimorar seu desempenho, com potencial para torná-lo uma solução viável no enfrentamento do *Aedes aegypti*.

## 5 CONCLUSÕES

O propósito principal deste estudo foi desenvolver um modelo de Aprendizado de Máquina capaz de identificar o mosquito *Aedes aegypti* por meio de imagens, distinguindo-o das diversas espécies potenciais presentes na região de Goiânia. O objetivo era aprimorar a eficiência da vigilância das arboviroses transmitidas pelo vetor, com a finalidade de prevenir antecipadamente a disseminação do patógeno entre a população.

Para alcançar tal objetivo, realizou-se uma análise abrangente da história do mosquito *Aedes aegypti*, compreendendo seu *habitat*, ciclo de vida, mecanismo de transmissão e as doenças associadas. Além disso, foi realizado um estudo das técnicas de Aprendizado de Máquina essenciais para o desenvolvimento do modelo de aprendizado. As redes neurais convolucionais foram selecionadas devido à sua comprovada eficácia em tarefas de visão computacional, conhecimento e recursos computacionais.

Uma vez estabelecida a estrutura de classificação, foi realizado uma coleta de *datasets* públicos pela internet de 14.716 imagens representativas do mosquito *Aedes aegypti* e de outras espécies, como o *Aedes albopictus*, *Anopheles* e *Culex quinquefasciatus*. As imagens foram submetidas a um pré-processamento para corrigir desequilíbrios, como o desbalanceamento, e remover possíveis interferências que poderiam prejudicar o desempenho do modelo. Ademais, a configuração da máquina foi otimizada para aproveitar todo o potencial de processamento da GPU, resultando em um melhor desempenho do modelo.

Os resultados obtidos sugerem uma tendência de superajuste no modelo, resultando em *overfitting*. Apesar disso, o modelo apresentou méritos consideráveis, alcançando uma taxa de acerto de aproximadamente 75,4% sobre o conjunto total de amostras fornecidas.

Apesar das limitações, os objetivos foram alcançados, representando um passo importante no desenvolvimento de uma solução para o controle do *Aedes aegypti* por meio da classificação de imagens de mosquitos. A utilização de técnicas de *machine learning* e visão computacional pode auxiliar no monitoramento e prevenção de doenças transmitidas por mosquitos, contribuindo para a saúde pública e o bem-estar da população.

Com futuras pesquisas e refinamentos no modelo, é possível que essa abordagem evolua significativamente, emergindo como uma ferramenta altamente eficaz no combate ao *Aedes aegypti* e outras espécies de mosquitos transmissores de doenças.

A contínua otimização do modelo permitirá sua integração em diversas plataformas, desde microscópios avançados até aplicativos de fácil acesso. A capacidade de automatizar o processo de classificação das espécies de forma rápida e precisa, superando os métodos

predominantemente manuais empregados atualmente, é um passo crucial para aprimorar os esforços de vigilância e controle de doenças transmitidas por mosquitos.

Ao ser implementado em microscópios, o modelo poderá ser utilizado por pesquisadores e profissionais de saúde em campo, agilizando a identificação de espécies vetores e permitindo respostas mais ágeis em áreas propensas a surtos. Além disso, a sua disponibilidade como um aplicativo acessível a uma variedade de usuários, desde entomologistas até cidadãos comuns, abrirá portas para a coleta colaborativa de dados e vigilância comunitária, fortalecendo assim as estratégias de prevenção e controle.

À medida que essa abordagem se consolida e amadurece, a combinação entre avanços tecnológicos e o conhecimento biológico contribuirá para a redução dos riscos associados às doenças transmitidas por mosquitos, representando um marco significativo na busca por ambientes mais saudáveis e seguros para a sociedade em geral.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BENCHIMOL, Jaime Larry. *Pereira Passos: um Haussmann Tropical. A renovação urbana na cidade do Rio de Janeiro no início do século XX*. Rio de Janeiro: Secretaria Municipal de Cultura, Departamento Geral de Documentação e Informação Cultural, 1992.
- [2] BENCHIMOL, Jaime Larry. *Reforma urbana e Revolta da Vacina na cidade do Rio de Janeiro*. In: Jorge Ferreira e Lucilia de Almeida Neves Delgado (org.) *O Brasil Republicano. O tempo do liberalismo excludente. Da proclamação da República à Revolução de 1930*. 3. ed. Rio de Janeiro: Civilização Brasileira, 2008.
- [3] Bittencourt, Edmundo. *Correio da Manhã*. Ano III. N. 684. Rio de Janeiro. Disponível em: [http://memoria.bn.br/pdf/089842/per089842\\_1903\\_00684.pdf](http://memoria.bn.br/pdf/089842/per089842_1903_00684.pdf). Acesso em: 26 abr 1903.
- [4] COSTA, Zouraide; ELKHOURY, Ana; FLANNERY, Brendan; ROMANO, Alessandro. *Evolução histórica da vigilância epidemiológica e do controle da febre amarela no Brasil*, 2011.
- [5] Franco, Odair. *História da Febre-Amarela no Brasil*. Ministério da Saúde. Departamento Nacional de Endemias Rurais. Rio de Janeiro, 1969
- [6] FIOCRUZ. *Combate à Febre Amarela*. Disponível em: <https://oswaldocruz.fiocruz.br/index.php/biografia/trajetoria-cientifica/na-diretoria-geral-de-saude-publica/combate-a-febre-amarela> . Acesso em: 14 mar 2023.
- [7] FIOCRUZ. *Uma breve história da febre amarela*. [S. l.], 2008. Disponível em: <https://agencia.fiocruz.br/uma-breve-historia-da-febre-amarela>. Acesso em: 9 set. 2021.
- [8] FLACH, Peter A. ROC analysis. In: *ENCYCLOPEDIA OF MACHINE LEARNING AND GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning*. [S. l.]: MIT press, 2016.
- [9] Goldim, José Roberto. *Experimentos sobre Transmissão da Febre Amarela*. Disponível em: <https://www.ufrgs.br/bioetica/finlay.htm>. Acesso em: 31 dez 1998.
- [10] BROWNLIE, Jason. *ROC Curves and Precision-Recall Curves for Imbalanced Classification*. [S. l.], 2020b. Disponível em: <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/>. Acesso em: 7 jan. 2022.

- [11] Júnior, Geraldo Pires Cordeiro. Reconhecimento De Espécies De Mosquitos Por Meio De Métodos Computacionais Inteligentes Baseados Em Aprendizagem Profunda. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Pernambuco, Programa de Pós-graduação em Ciência da Computação, Recife, 2022.
- [12] Ministério da Saúde, Departamento Nacional de Endemias Rurais, História da Febre Amarela no Brasil, Dr. Odair Franco, Rio de Janeiro, 1969
- [13] Ministério da Saúde. Arboviroses Urbanas Causadas por Vírus transmitidos pelo Aedes: Dengue, Chikungunya e Zika. Guia de Vigilância em Saúde.
- [14] Ministério da Saúde. Fundação Nacional de Saúde. Dengue, Instruções para Pessoal de Combate ao Vetor – Manual de Normas Técnicas. 3. ed., rev. - Brasília : Ministério da Saúde : Fundação Nacional de Saúde, 2001.
- [15] Motta, Marly. O Bota-Abaixo. Disponível em: <https://atlas.fgv.br/verbetes/o-bota-abaixo>. Acesso em: 14 mar 2023.
- [16] Neto, D. S; Rocha, M. D. H. A; Cavalcante, P. A. M; Mariano, W. S. Doenças Transmitidas por Vetores. São Carlos: Pedro & João, 2021.
- [17] O Mosquito que Desafia o Brasil. IPEA, ano 13. Ed 87. Disponível em: [https://www.ipea.gov.br/desafios/index.php?option=com\\_content&id=3260](https://www.ipea.gov.br/desafios/index.php?option=com_content&id=3260). Acesso em: 17 jun 2016.
- [18] Revista de Manguinhos, FIOCRUZ, 37, maio 2017, O legado de Oswaldo Cruz.
- [19] Vieira, Cássio Leite. O que foi a Revolta da Vacina? Disponível em: <https://super.abril.com.br/historia/oswaldo-cruz-e-a-variola-a-revolta-da-vacina>. Acesso em: 31 out 1994.