

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
INSTITUTO DE BIOTECNOLOGIA  
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

**INTELIGÊNCIA ARTIFICIAL NO COMBATE À MALÁRIA: CLASSIFICAÇÃO DE  
IMAGENS MICROSCÓPICAS COM REDES NEURAIS CONVOLUCIONAIS**

**ISABELA DOS SANTOS ROCHA**

**Uberlândia - MG  
Novembro/ 2024**

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
INSTITUTO DE BIOTECNOLOGIA  
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

**INTELIGÊNCIA ARTIFICIAL NO COMBATE À MALÁRIA: CLASSIFICAÇÃO DE  
IMAGENS MICROSCÓPICAS COM REDES NEURAIS CONVOLUCIONAIS**

Monografia apresentada à Coordenação do Curso de  
Biotecnologia, da Universidade Federal de  
Uberlândia, para obtenção do grau de Bacharel em  
Biotecnologia.

**Orientador:** Prof. Dr. Carlos Ueira Vieira

**Coorientador:** Prof. Dr. Murilo Vieira da Silva

**Uberlândia - MG**

**Novembro/ 2024**

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**INSTITUTO DE BIOTECNOLOGIA**  
**CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA**

**INTELIGÊNCIA ARTIFICIAL NO COMBATE À MALÁRIA: CLASSIFICAÇÃO DE  
IMAGENS MICROSCÓPICAS COM REDES NEURAIIS CONVOLUCIONAIS**

Isabela dos Santos Rocha

Prof. Dr. Carlos Ueira Vieira  
Instituto de Biotecnologia (IBTEC)

Prof. Dr. Murilo Vieira da Silva  
Faculdade de Medicina Veterinária (FAMEV)

Instituto de Biotecnologia homologado pela Coordenação do  
Curso de Biotecnologia em \_\_\_\_/\_\_\_\_/\_\_\_\_.

---

Prof. Dr. Nilson Nicolau Júnior

**Uberlândia - MG**  
**Novembro/ 2024**

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**INSTITUTO DE BIOTECNOLOGIA**  
**CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA**

**INTELIGÊNCIA ARTIFICIAL NO COMBATE À MALÁRIA: CLASSIFICAÇÃO DE  
IMAGENS MICROSCÓPICAS COM REDES NEURAIIS CONVOLUCIONAIS**

Isabela dos Santos Rocha

Aprovado pela Banca Examinadora em: \_\_/\_\_/\_\_\_\_. Nota:

---

Prof. Dr. Carlos Ueira Vieira

---

Prof. Dr. Murilo Vieira da Silva

---

Prof. Dr. Richard Costa Polveiro

---

Prof. Dr. Mauricio Cunha Escarpinati

Uberlândia, 14 de novembro de 2024.

18/02/2025, 11:36

SEVUFU - 5870427 - Ata de Defesa - Graduação



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
 Coordenação do Curso de Graduação em Biotecnologia - Uberlândia  
 Av. Pará, 1720, Bloco 2E, Sala 205 - Bairro Umuarama, Uberlândia-MG, CEP 38400-902  
 Telefone: (34)3225-8441 - www.ibtec.ufu.br - cbiotec@ibtec.ufu.br



### ATA DE DEFESA - GRADUAÇÃO

|                                                        |                                                                                                                        |                   |       |                       |       |
|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|-------------------|-------|-----------------------|-------|
| Curso de Graduação em:                                 | Biotecnologia                                                                                                          |                   |       |                       |       |
| Defesa de:                                             | Trabalho de Conclusão de curso II - GBT094                                                                             |                   |       |                       |       |
| Data:                                                  | 14/11/2024                                                                                                             | Hora de início:   | 10:30 | Hora de encerramento: | 11:45 |
| Matrícula do Discente:                                 | 11921BTC003                                                                                                            |                   |       |                       |       |
| Nome do Discente:                                      | Isabela dos Santos Rocha                                                                                               |                   |       |                       |       |
| Título do Trabalho:                                    | Inteligência artificial no combate à malária: classificação de imagens microscópicas com redes neurais convolucionais. |                   |       |                       |       |
| A carga horária curricular foi cumprida integralmente? |                                                                                                                        | ( x ) Sim ( ) Não |       |                       |       |

Reuniu-se no Anfiteatro/Sala 8C 123, Campus Umuarama, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Curso de Graduação em Biotecnologia, assim composta: Professores: Dr. Richard Costa Polveiro - FAMEV/UFU; Dr. Mauricio Cunha Escarpinati - FACOM/UFU; Dr. Carlos Ueira Vieira - IBTEC/UFU, orientador(a) do(a) candidato(a).

Iniciando os trabalhos, o(a) presidente da mesa, Dr(a). Carlos Ueira Vieira, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao discente a palavra, para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do curso.

A seguir o(a) senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

( X ) Aprovada. Nota 95

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Carlos Ueira Vieira, Professor(a) do Magistério Superior**, em 14/11/2024, às 16:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Richard Costa Polveiro, Usuário Externo**, em 18/11/2024, às 14:33, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Professor(a) do Magistério Superior**, em 22/11/2024, às 16:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5870427** e o código CRC **93268148**.

## DEDICATÓRIA

Dedico esse trabalho a todos que acreditaram que eu poderia chegar até aqui e concluir mais esta etapa.

A minha avó Astrogilda Xavier dos Santos que não está mais entre nós, porém seu maior sonho era ver-me formando (todas as suas orações valeram a pena obrigada por tudo e por tanto).

A minha mãe Maria Elaine Dias Dos Santos que esteve ao meu lado, cuidou e abraçou esse sonho de ser biotecnologista junto de mim.

Ao meu irmão e parceiro de vida acadêmica Eduardo dos Santos Rocha que contribuiu com meu aprendizado em informática e sobre evolução pessoal.

Ao meu avô Jose Dias Dos Santos por toda paciência e ajuda durante todo meu caminho e por sempre agir como um pai em vários momentos.

A minha tia Elisabeth Dia Dos Santos e meu primo Henric Dias França que estiveram ali por mim e me acolheram no período mais difícil da faculdade o primeiro.

A minha namorada Aurea Camilli Ramos que percorreu todo o processo de escrita desse trabalho, que me apoia na vida e divide a maternidade da nossa pequena gatinha Vanessa Maria.

A minha cadelinha Izzy que foi minha paz durante vários momentos dessa longa jornada.

Aos meus ilustres amigos que não soltaram a minha mão mesmo em momentos difíceis e de crise. Agradeço por cada risada, cada festa, vinho e poesia, conselho e por tanto amor.

## **AGRADECIMENTOS**

Agradeço ao meu orientador Carlos Ueira Vieira, meu coorientador Murilo Vieira da Silva, aos pós doutorandos do Laboratório de Biotecnologia em Modelos experimentais que me auxiliaram no desenvolvimento dos meus projetos.

Agradeço aos componentes da banca Carlos, Murilo, Richard e Mauricio por aceitarem o convite e proporcionarem meu crescimento acadêmico.

Gostaria de expressar minha sincera gratidão ao Instituto de Biotecnologia (IBTEC) por todo o apoio e incentivo durante a realização deste trabalho. A infraestrutura oferecida e o ambiente colaborativo foram essenciais para o desenvolvimento das minhas pesquisas. Aos professores e colegas, meu muito obrigado por compartilharem conhecimento e experiências que enriqueceram este projeto.

## RESUMO

A malária é um dos principais problemas de saúde pública no mundo, exigindo estratégias eficazes para diagnóstico e controle. A classificação automatizada tem o potencial de acelerar o diagnóstico e aumentar a precisão na detecção de infecções. Este trabalho propõe o uso de inteligência artificial, especificamente com Redes Neurais Convolucionais (CNN), para a classificação de imagens microscópicas de células humanas infectadas pelo parasita *Plasmodium spp.* Foi utilizado um conjunto de dados público contendo 27.527 imagens, divididas em células saudáveis e infectadas. As imagens passaram por um rigoroso pré-processamento, que incluiu redimensionamento, normalização e aumento de dados, otimizando assim o treinamento do modelo. A arquitetura da CNN consistiu em camadas convolucionais, *pooling* e *dropout*, garantindo uma extração de características eficaz.

Os resultados indicaram uma acurácia de 93,21%, superando desempenhos de modelos anteriores. Além do modelo, foi desenvolvida uma aplicação web intuitiva, permitindo que os usuários carreguem e analisem imagens com facilidade. Os desafios enfrentados, como o longo tempo de treinamento (12 horas), indicam a necessidade de infraestrutura computacional mais robusta para futuras implementações. Este estudo destaca o papel da inteligência artificial na melhoria dos métodos de diagnóstico da malária, evidenciando sua aplicabilidade em saúde pública e potencial para transformar o manejo da doença por meio de técnicas automatizadas.

**Palavras-chave:** doença parasitaria, IA , CNN, diagnóstico, processamento de imagens, saúde pública.

## ABSTRACT

Malaria remains one of the major public health challenges worldwide, necessitating effective strategies for diagnosis and control. This work proposes the use of artificial intelligence, specifically Convolutional Neural Networks (CNN), for the classification of microscopic images of human cells infected by the *Plasmodium spp.* parasite. Automated classification has the potential to accelerate diagnosis and enhance accuracy in detecting infections. To achieve this objective, a public dataset containing 27,527 images was utilized, divided into healthy and infected cells. The images underwent rigorous preprocessing, including resizing, normalization, and data augmentation, thereby optimizing model training. The CNN architecture consisted of convolutional, pooling, and dropout layers, ensuring effective feature extraction. The results indicated an accuracy of 93.21%, surpassing the performance of previous models. In addition to the model, an intuitive web application was developed, allowing users to easily upload and analyze images. The challenges faced, such as the long training time (12 hours), highlight the need for more robust computational infrastructure for future implementations. This study emphasizes the role of artificial intelligence in improving malaria diagnostic methods, showcasing its applicability in public health and its potential to transform disease management through automated techniques.

**Keywords:** parasitic diseases, IA, CNN, diagnosis, image processing, public health.

## LISTA DE FIGURAS

|                                                                                                                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figura 1:</b> Ciclo de vida do Plasmodium sp. Fonte: Imagem adaptada de Amabis et al., 2004...                                                                                                                                                               | 16 |
| <b>Figura 2:</b> Bancos de Dados de Sequências de Nucleotídeos (INSDC) Fonte: Imagem adaptada de Pevsner, 2015. ....                                                                                                                                            | 19 |
| <b>Figura 3:</b> Comparação entre a rede neural artificial e a rede neural biológica Fonte: Autoria Própria.....                                                                                                                                                | 21 |
| <b>Figura 4:</b> Simplificação da Rede neural convolucional Fonte: Autoria Própria. ....                                                                                                                                                                        | 22 |
| <b>Figura 5:</b> Imagens saudáveis sem contaminação por Plasmodium sp. Fonte: Imagem adaptada de Malária Cell Images Dataset ( <a href="https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html">https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html</a> )..... | 24 |
| <b>Figura 6:</b> Imagens contaminadas por Plasmodium sp. Fonte: Imagem adaptada de Malária Cell Images Dataset ( <a href="https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html">https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html</a> ). ....              | 24 |
| <b>Figura 7:</b> Gráfico da Acurácia e Loss no treino e validação do modelo. Fonte: Autoria Própria .....                                                                                                                                                       | 30 |
| <b>Figura 8:</b> Acurácia e Loos do modelo. Fonte Autoria Própria .....                                                                                                                                                                                         | 30 |
| <b>Figura 9:</b> Página inicial da Interface Web. Fonte: Autoria Própria .....                                                                                                                                                                                  | 31 |
| <b>Figura 10:</b> Página de Classificação. Fonte: Autoria Própria.....                                                                                                                                                                                          | 32 |
| <b>Figura 11:</b> Abertura do Explorador de Arquivos. Fonte: Autoria Própria .....                                                                                                                                                                              | 32 |
| <b>Figura 12:</b> Predição final das Imagens. Fonte: Autoria Própria.....                                                                                                                                                                                       | 33 |

## LISTA DE ABREVIATURAS

|       |                                                                            |
|-------|----------------------------------------------------------------------------|
| AGI   | Inteligência Artificial Geral                                              |
| ASI   | Super Inteligência Artificial                                              |
| CNN   | Redes Neurais Convolucionais                                               |
| DDBJ  | Banco de DNA do Japão                                                      |
| DNA   | Ácido Desoxirribonucleico                                                  |
| EBI   | Instituto de Bioinformática da Europa                                      |
| HIV   | Vírus da Imunodeficiência Humana                                           |
| IA    | Inteligência Artificial                                                    |
| IHC   | Interação Humano-Computador                                                |
| INSDC | Colaboração Internacional de Banco de Dados de Sequências de Nucleotídeos. |
| NCBI  | Centro de Informação Biotecnológica                                        |
| NGS   | Sequenciamento de Nova Geração                                             |
| NIH   | Instituto Nacional de Saúde                                                |
| RNA   | Rede Neural Artificial                                                     |

## SUMÁRIO

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| 1. INTRODUÇÃO.....                                                                   | 14 |
| 1.1 Histórico da Malária e seus Agentes Etiológicos .....                            | 14 |
| 1.2 Transmissão da Malária: Anofelinos e o Gênero <i>Plasmodium</i> .....            | 15 |
| 1.3 Bioinformática Aplicada ao Diagnostico de Doenças: Ferramentas e Inovações ..... | 17 |
| 1.3 Inteligência Artificial e Machine learning .....                                 | 19 |
| 1.4 Redes Neurais (RNA) e Redes Convolucionais (CNN) .....                           | 20 |
| 2. OBJETIVOS.....                                                                    | 22 |
| 3. MATERIAL E MÉTODOS .....                                                          | 23 |
| 3.1 Base de Dados Utilizada .....                                                    | 23 |
| 3.2 Ferramentas e Bibliotecas Utilizadas .....                                       | 24 |
| 3.2.1 Ambiente de desenvolvimento: .....                                             | 24 |
| 3.2.2 Linguagem de programação: .....                                                | 25 |
| 3.2.3 Bibliotecas: .....                                                             | 25 |
| 3.3 Processamento das imagens.....                                                   | 25 |
| 3.3.1 Redimensionamento:.....                                                        | 25 |
| 3.3.2 Normalização: .....                                                            | 26 |
| 3.3.3 Aumento de Dados (Data Augmentation): .....                                    | 26 |
| 3.4 Construção e Treinamento do Modelo .....                                         | 26 |
| 3.4.1 A arquitetura da CNN.....                                                      | 26 |
| 3.4.2 Parâmetros de treinamento: .....                                               | 26 |
| 3.5 Divisão dos Dados.....                                                           | 27 |
| 3.6 Métricas de Avaliação.....                                                       | 27 |
| 3.7 Criação de uma aplicação web.....                                                | 28 |
| 3.7.1 Normas de Interação Humano-Computador (IHC) .....                              | 28 |
| 3.7.2 Ferramentas utilizadas na página web: .....                                    | 28 |
| 4. RESULTADOS E DISCUSSÃO .....                                                      | 29 |
| 4.2 Eficácia das Técnicas de Pré-processamento.....                                  | 29 |
| 4.3 Desempenho e Arquitetura da CNN .....                                            | 29 |
| 4.4 Análise de Métricas de Desempenho .....                                          | 29 |
| 4.4.1 Treino .....                                                                   | 29 |

|                                                                 |           |
|-----------------------------------------------------------------|-----------|
| 4.4.2 Teste .....                                               | 30        |
| <b>4.5 Modelo Final.....</b>                                    | <b>30</b> |
| <b>4.6 Desafios e Limitações .....</b>                          | <b>30</b> |
| <b>4.7 Desenvolvimento e Usabilidade da Aplicação Web .....</b> | <b>31</b> |
| 4.7.1 Página Inicial .....                                      | 31        |
| 4.7.2 Página de predição .....                                  | 31        |
| 5. CONCLUSÃO.....                                               | 33        |
| 6. REFERÊNCIAS .....                                            | 34        |

## 1. INTRODUÇÃO

### 1.1 Histórico da Malária e seus Agentes Etiológicos

A malária é uma doença parasitária, sistêmica, infecciosa, febril, não contagiosa e endêmica, historicamente conhecida por diversos nomes, como impaludismo, paludismo, febre terçã maligna e benigna, além de febre quartã. No Brasil, a enfermidade também ficou popularmente conhecida por termos como sezão, tremedeira, maleita e, de forma simplificada, febre. Trata-se de uma doença antiga, porém reemergente, com registros que remontam à era pré-Cristã. Hipócrates, renomado médico grego considerado o pai da Medicina, já mencionava a ocorrência sazonal de febres recorrentes associadas à malária (MIOTO, 2012).

A malária recebe esse nome a partir do termo italiano "mal'aria", que significa "ar ruim", representando a associação da doença com áreas pantanosas. É uma doença parasitária endêmica transmitida por vetores, causada por parasitas protozoários do gênero *Plasmodium* em regiões tropicais e subtropicais em todo o mundo. O gênero *Plasmodium* é composto por mais de 200 espécies, que infectam mamíferos, aves e répteis, e os parasitas da malária tendem a ser específicos para seus hospedeiros. *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale* e *Plasmodium knowlesi* são as cinco espécies conhecidas do gênero *Plasmodium* que causam malária em humanos. Das cinco espécies de *Plasmodium* que causam malária em humanos, *P. falciparum* é responsável pela malária grave (FIKADU, 2023).

Ao longo da história, a malária tornou-se um significativo problema de saúde pública. Determinados grupos populacionais apresentam maior suscetibilidade à infecção e risco elevado de desenvolver formas graves da doença, como crianças menores de cinco anos, pacientes portadores do vírus da imunodeficiência humana (HIV) e gestantes. Estima-se que, em 2022, a malária tenha acometido 249 milhões de pessoas em 85 países, resultando em 608.000 óbitos (OMS, 2024). Mais de 90% dos casos fatais ocorrem na África Tropical, 80% na Ásia (GARCIA, 2010).

Os locais de maior risco de contaminação pela malária geralmente estão situados em áreas próximas a florestas tropicais úmidas, uma vez que essas regiões proporcionam condições favoráveis para a transmissão contínua da doença. No Brasil, a área endêmica da malária está

concentrada na Região Amazônica, que abrange os estados do Amazonas, Acre, Amapá, Maranhão, Mato Grosso, Roraima, Rondônia e Pará. Entre os anos de 2010 e 2016, observou-se um declínio significativo nos casos autóctones da enfermidade. No entanto, em 2017, houve um aumento de 52,7% no número de casos reemergentes, seguido por uma nova redução nos anos subsequentes. Apesar dessa diminuição, o número de óbitos ainda se mantém elevado, com 50 fatalidades registradas em 2021 (MINISTÉRIO DA SAÚDE, 2024).

## 1.2 Transmissão da Malária: Anofelinos e o Gênero *Plasmodium*

A malária é transmitida por meio de interações entre humanos, mosquitos do gênero *Anopheles* e o patógeno. Durante o período de postura de ovos, as fêmeas dos mosquitos, ao apresentarem deficiência de proteína e sacarose vegetal essenciais para a maturação dos ovos, passam a se alimentar de sangue humano. Nesse processo, ao realizarem picadas, podem transmitir protozoários infecciosos do gênero *Plasmodium* para a corrente sanguínea humana (SAVI, 2022).

O ciclo de vida do *Plasmodium* (Figura 1), causador da malária, envolve dois ambientes distintos: o corpo humano e o mosquito transmissor, geralmente o *Anopheles*. Tudo começa quando o inseto infectado pica uma pessoa e libera os parasitas, chamados esporozoítos, na corrente sanguínea. Esses esporozoítos rapidamente migram para o fígado, onde se instalam nas células hepáticas e se multiplicam intensamente. Ao se dividirem, geram milhares de novos parasitas, os merozoítos, que retornam ao sangue e começam a infectar os glóbulos vermelhos. É aí que surgem os sintomas da malária, pois os parasitas destroem as células à medida que continuam se replicando. Em certo momento, alguns desses merozoítos se transformam em gametócitos, que são as formas que possibilitam a transmissão para um novo vetor. Quando outro mosquito pica a pessoa infectada, os gametócitos ingeridos entram no intestino do inseto, onde amadurecem e se fundem para formar um zigoto. Esse zigoto se desenvolve e atravessa a parede intestinal, formando um oocisto na parte externa. No interior do oocisto, surgem centenas de novos esporozoítos, que migram para as glândulas salivares do vetor, prontos para serem transmitidos a outra pessoa na próxima picada, reiniciando todo o ciclo (SATO, 2021).

Como consequência disso o indivíduo infectado passa a apresentar sintomas como febre alta, tremores, calafrios, sudorese e dor de cabeça e em casos mais graves convulsões, alterações da consciência, dispneia ou hiperventilação, prostração, hipotensão, hemorragia entre outros. Dos tratamentos existentes os mais utilizados são: A interrupção da esquizogonia sanguínea, que é a principal responsável pela patogenicidade e manifestações clínicas da infecção; a eliminação das formas latentes do parasita no ciclo tecidual (hipnozoítos) presentes

em *P. vivax* e *P. ovale*, prevenindo as recaídas; e a interrupção da transmissão por meio de medicamentos que inibem o desenvolvimento das formas sexuais do parasita (MINISTÉRIO DA SAÚDE, 2024).

Anofelinos são mosquitos referentes a ordem Díptera, infra ordem *Culicomorpha*, família Culicidae, gênero *Anopheles* e popularmente são conhecidos como mosquito-prego. Estes possuem em sua totalidade 400 espécies, porém apenas 40 destas são responsáveis pela transmissão da malária mundialmente sendo eles *Anopheles maculipennis*, *A. culicifacies*, *A. minimus*, *A. gambiae* e *A. funestus*; *A. albimanus* e *A. nuñeztovari*; *A. albitarsis*; *A. marajoara*; *A. janconnae*; *A. deaneorum*; *A. cruzii*; *A. bellator* e *A. homunculus*; *A. darlingi*. As fêmeas deste gênero são hematófagas sendo elas as responsáveis da transmissão do plasmodium para o humano, além disto são aladas, possuem pernas e antenas longas características estas que descrevem também o mosquito macho (SAVI, 2022).

*Plasmodium* pertence à classe *Aconoidasida*, ordem *Hemosporodida*, família *Plasmodiidae*. O parasita tem como hospedeiro definitivo o mosquito do gênero *Anopheles* e como hospedeiro intermediário o ser humano. No caso do *P. ovale* e *P. knowlesi*, algumas espécies de macacos também podem ser hospedeiros intermediários. Ele passa por diferentes estágios de desenvolvimento, sendo eles esporozoíto, hipnozoíto, trofozoíto, merozoíto, esquizonte, gametócitos, oocineto e oocisto (MEIBALAN, 2017).

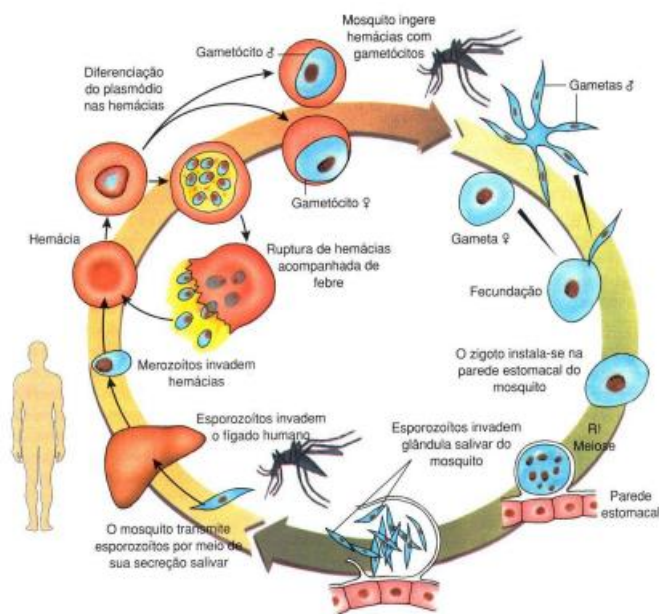


**Figura 1:** Ciclo de vida do *Plasmodium* sp. Fonte: Imagem adaptada de Amabis et al., 2004.

### 1.3 Bioinformática Aplicada ao Diagnóstico de Doenças: Ferramentas e Inovações

A biologia passou por uma transformação ao longo dos anos devido à crescente importância dos computadores e *softwares*, que se tornaram essenciais para seu avanço. Eles permitiram a análise de dados complexos, como sequências de DNA e proteínas, entre muitas outras áreas relevantes (AUTHIER et al., 2019). Esse campo emergente, que combina as ciências computacionais com a biologia, foi denominado bioinformática e é descrito por (PEVSNER, 2015) como a aplicação de bancos de dados computacionais e algoritmos para a análise de genes, genomas e proteínas.

A integração dessas áreas do conhecimento ganhou destaque global em 1990, com o início do Projeto Genoma Humano, idealizado por Craig Venter Bioquímico e Empresário Americano. O objetivo do projeto era identificar e mapear todos os genes da espécie *Homo sapiens*, construindo o genoma completo. Em 2003, sua conclusão marcou um momento histórico para a biologia molecular e a genética, promovendo avanços significativos na compreensão do DNA humano e abrindo novas possibilidades para a pesquisa biomédica. Outro marco importante foi a criação do Atlas Celular, que busca caracterizar os 37,5 trilhões de células humanas (SANCHES, 2021).

O NCBI (*National Center for Biotechnology Information*), o EBI (*European Bioinformatics Institute*) e o DDBJ (*DNA Data Bank of Japan*) (figura 2) são três dos principais centros internacionais dedicados ao armazenamento e compartilhamento de dados biológicos, com foco em sequências genéticas. Essas instituições trabalham de forma colaborativa, integrando informações genômicas e moleculares por meio do International Nucleotide Sequence Database Collaboration (INSDC), garantindo que os dados sejam sincronizados e acessíveis globalmente para a comunidade científica. O NCBI, parte do *National Institutes of Health (NIH)* nos Estados Unidos, hospeda bases de dados como GenBank, PubMed e RefSeq, além de oferecer ferramentas como o BLAST para comparação de sequências. O EBI, localizado no Reino Unido e parte do EMBL, gerencia o banco EMBL-Bank e fornece recursos como Ensembl e UniProt para análise de genomas, transcriptomas e proteomas. Já o DDBJ, no Japão, operado pelo *National Institute of Genetics*, foi o primeiro repositório público de sequências de DNA na Ásia. Juntas, essas instituições recebem submissões de dados desde 1982 e, nas últimas décadas, o volume de sequências de DNA armazenadas tem crescido exponencialmente, especialmente com o surgimento das tecnologias de sequenciamento de nova geração (NGS), permitindo que dados sejam gerados em uma escala sem precedentes. Isso traz novos desafios para a aquisição, análise e distribuição dessas informações, o que torna ferramentas como o RefSeq e a automação de processos, como o uso do comando *tbl2asn* no *GenBank*, essenciais para o controle de

qualidade e padronização das entradas nos repositórios (PEVSNER, 2015).

Além de sua aplicação no gerenciamento de grandes bases de dados, a bioinformática também é utilizada para resolver questões biológicas complexas. Na pesquisa realizada por (PAN et al., 2022) foram utilizadas várias técnicas de bioinformática, incluindo sequenciamento de alto rendimento para gerar perfis moleculares de tumores, análise computacional de redes reguladoras de microRNAs (miRNAs) e modelagem de diafonia entre miRNAs e RNAs endógenos competitivos (ceRNAs). Métodos para integrar dados multiômicos (genômica, transcriptômica e fenômica) são aplicados para identificar interações miRNA-gene e construir redes miRNA-miRNA e ceRNA-ceRNA. Essas abordagens fornecem insights sobre os mecanismos regulatórios no câncer, contribuindo para o avanço da medicina de precisão.

Outra área que se beneficia é a descoberta de fármacos a qual explora a análise computacional de grandes volumes de dados gerados durante o processo de descoberta de medicamentos ajuda a enfrentar desafios, incluindo a criação de *data warehouses* que centraliza e organiza dados de várias fontes, algoritmos para identificar novos alvos de medicamentos e biomarcadores, e programas para avaliar a tratabilidade de alvos e prever oportunidades de reposicionamento de medicamentos já licenciados para novas indicações (WOOLLER et al., 2017).

As ferramentas de bioinformática são amplamente empregadas na identificação de patógenos, utilizando abordagens genômicas no diagnóstico de infecções virais, bacterianas e fúngicas. NGS e a análise filogenética têm sido eficazes na detecção de agentes causadores, como *Clostridium haemolyticum*, e na rotina de estudos microbiológicos. A IA, por sua vez, tem se mostrado essencial no diagnóstico de doenças agudas e crônicas, como apendicite (Inflamação do apêndice vermiforme) e Alzheimer. Algoritmos avançados melhoram a precisão na detecção de células cancerosas e preveem a recorrência do câncer de mama. Sistemas de IA também podem monitorar pacientes em casa, com grandes expectativas para sua expansão no diagnóstico futuro.

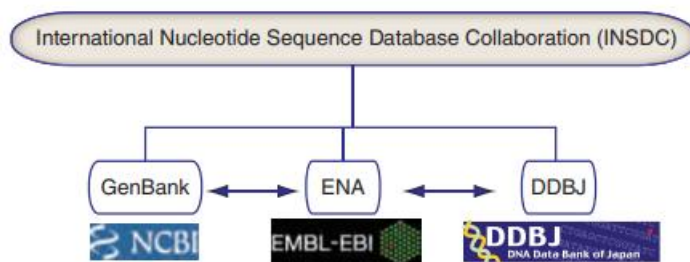


**Figura 2:** Bancos de Dados de Sequências de Nucleotídeos (INSDC) Fonte: Imagem adaptada de Pevsner, 2015.

### 1.3 Inteligência Artificial e Machine learning

Em meados dos anos 1950 o questionamento da capacidade das máquinas pensarem pairava sobre os cientistas da época, Alan Turing que foi um matemático, criptoanalista e cientista da computação trouxe a resposta para tal questão “um computador era capaz de “pensar” se seus resultados fossem tão convincentes que uma pessoa que interagisse com ele não pudesse distinguir suas respostas das de um ser humano real.” (PEREIRA, 2024).

A inteligência artificial (IA) é uma tecnologia que permite a computadores e máquinas simular habilidades de resolução de problemas e processos de tomada de decisão semelhantes aos humanos. Combinada com outras tecnologias, como sensores e robótica, a IA realiza tarefas que normalmente exigiriam intervenção humana, como navegação por GPS, veículos autônomos e ferramentas generativas, como o ChatGPT. A IA inclui aprendizado de máquina e *deep learning*, que desenvolvem algoritmos capazes de aprender com dados e melhorar suas previsões. Recentemente, a IA generativa deu um grande salto no processamento de linguagem natural e agora também trabalha com imagens, vídeos e até estruturas moleculares.

Existem atualmente dois tipos de inteligência artificial a IA fraca, também chamada de IA estreita, é projetada para realizar tarefas específicas e é a forma mais comum de IA usada hoje, como assistentes virtuais (Siri, Alexa) e veículos autônomos. Apesar do nome, ela é bastante eficiente em suas funções. Já a IA forte inclui a inteligência artificial geral (AGI) e a superinteligência artificial (ASI). AGI seria equivalente à inteligência humana, capaz de resolver problemas, aprender e planejar como uma pessoa. A ASI, por sua vez, superaria a capacidade humana. Embora ambas sejam atualmente teóricas, a ASI é mais explorada em obras de ficção científica (IBM, 2024a).

O aprendizado de máquina é um subcampo da inteligência artificial que utiliza dados passados para prever resultados futuros, classificando novos dados conforme são inseridos no sistema. Esse aprendizado ocorre por meio da modificação na estrutura do programa ou da base de dados, o que permite aprimorar continuamente o desempenho do modelo.

As técnicas de aprendizado de máquina são divididas em aprendizado supervisionado e não supervisionado, sendo que o aprendizado supervisionado envolve um processo de treinamento com um conjunto de dados de entrada e um conjunto de teste para avaliar o desempenho. Em contrapartida, no aprendizado não supervisionado, o modelo identifica padrões a partir dos dados fornecidos sem treinamento prévio. Dentre os problemas abordados em aprendizado de máquina, destacam-se a regressão, classificação binária, classificação multiclasse e ranking. Algoritmos como Naive Bayes, Árvores de Decisão e Florestas Aleatórias foram estudados, visto que essas técnicas fornecem um conjunto robusto de ferramentas para a construção de modelos precisos e confiáveis.

#### **1.4 Redes Neurais (RNA) e Redes Convolucionais (CNN)**

Rede Neural artificial (RNA) é compreendida como um algoritmo de *machine learning*, seu surgimento tem como base a inspiração nas redes neurais biológicas. Desta forma cada RNA nos possui que seriam seus “corpos celulares” que se comunicam-se uns com os outros por meio de conexões cujo seu análogo biológico seria os axônios e os dendritos. Assim como, em redes neurais biológicas, as sinapses se tornam mais fortes quando os neurônios disparam de maneira sincronizada seguindo o princípio Hebbiano de que "neurônios que disparam juntos, conectam-se", as ligações entre os nós de uma RNA são ajustadas com base em sua habilidade de gerar a saída esperada (figura 3) (CHOI et al., 2020).

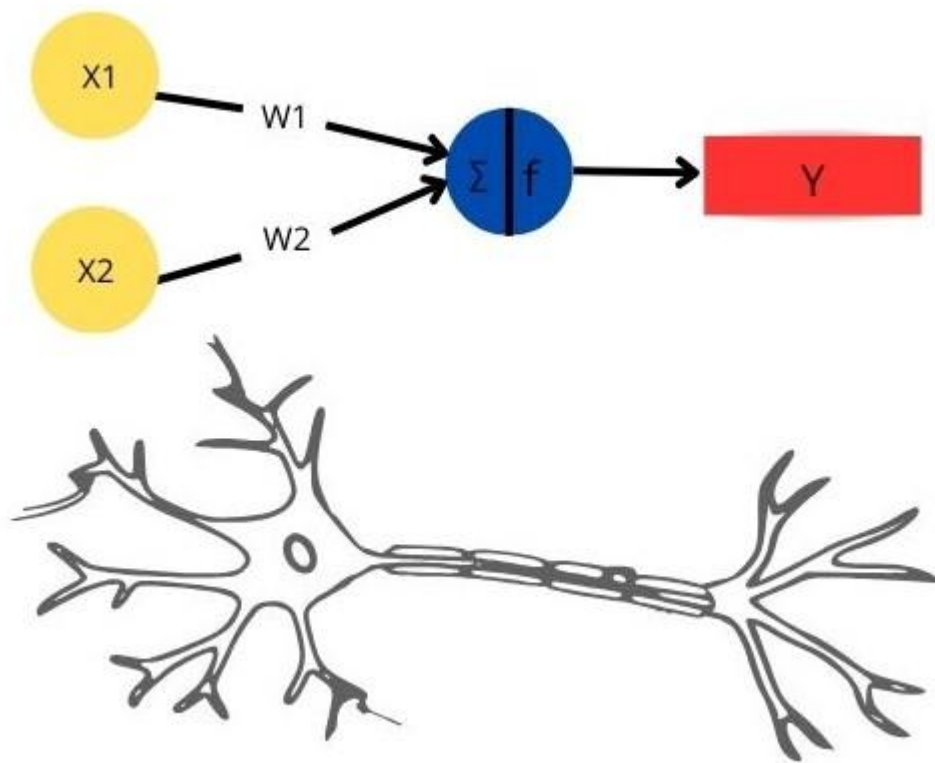


**Figura 3:** Comparação entre a rede neural artificial e a rede neural biológica Fonte: Autoria Própria.

As Redes Neurais Convolucionais (CNNs) (figura 4) são amplamente aplicadas em problemas de visão computacional, como classificação de imagens e detecção de objetos, devido à sua capacidade de capturar padrões complexos em dados tridimensionais. Elas representam uma alternativa mais eficaz aos métodos manuais de extração de características, permitindo a análise automatizada de imagens. Essas redes são formadas por camadas específicas, como as convolucionais, de pooling e totalmente conectadas, que atuam em conjunto para extrair, comprimir e classificar informações visuais. As CNNs têm a habilidade de reconhecer elementos desde detalhes simples, como bordas e cores, até padrões complexos que formam a estrutura de objetos completos. Hoje, as CNNs desempenham um papel crucial em soluções que requerem análise visual e reconhecimento de padrões, sendo um dos principais alicerces da inteligência artificial aplicada à visão computacional (IBM, 2024).

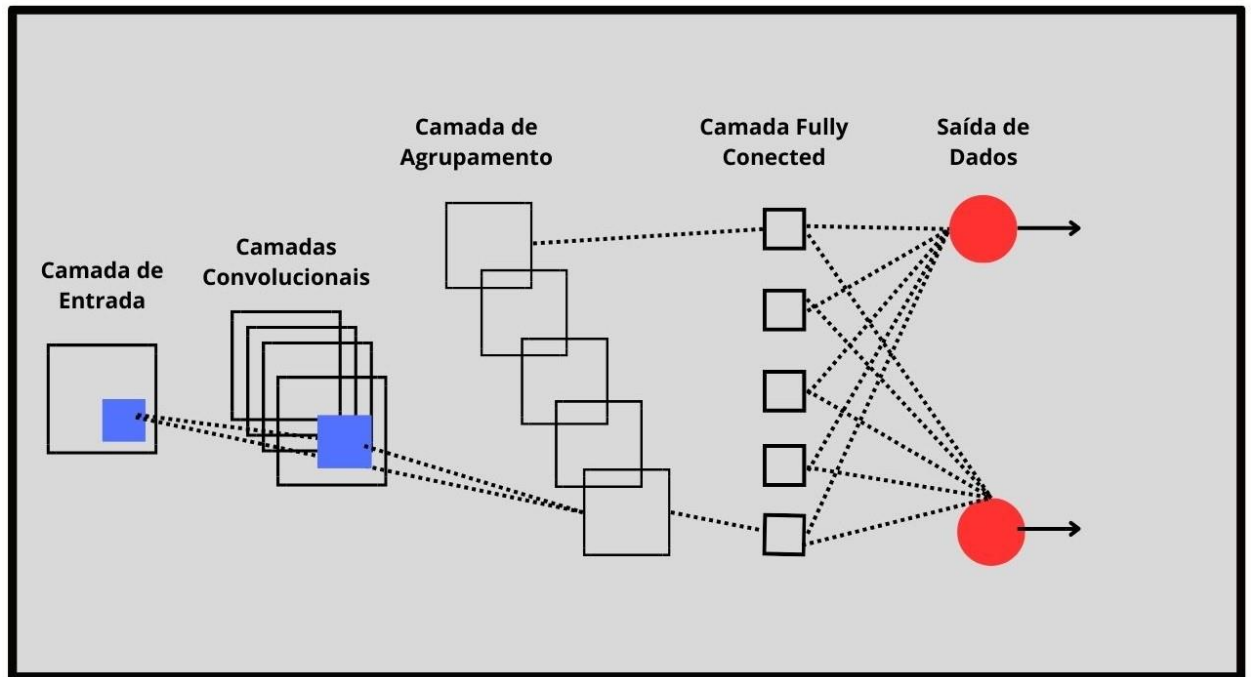


**Figura 4:** Simplificação da Rede neural convolucional Fonte: Autoria Própria.

Diante disso, a malária ainda é um dos principais problemas de saúde pública em regiões tropicais e subtropicais, acometendo milhões de pessoas globalmente. Embora haja tratamentos e diagnósticos significativos, o combate à doença enfrenta desafios, especialmente em áreas com recursos limitados e alta vulnerabilidade. Iniciativas que integram diagnósticos de doenças com biotecnologia, IA, e machine learning têm apresentado grande potencial de detecção precoce, oferecendo perspectivas para o controle e eventual controle dessa enfermidade milenar.

## 2. OBJETIVOS

Desenvolver um modelo de inteligência artificial baseado em CNN para a classificação automática de imagens microscópicas de células infectadas com malária, de modo a auxiliar no diagnóstico e controle da doença, proporcionando maior precisão e agilidade na detecção de casos positivos.

Os objetivos específicos desse projeto foram:

1. Coletar e organizar um conjunto de dados de imagens microscópicas contendo células humanas infectadas e não infectadas por *Plasmodium spp.*
2. Pré-processar as imagens para otimizar o desempenho do modelo, aplicando técnicas de redimensionamento, normalização e segmentação das células.

3. Desenvolver e treinar uma Rede Neural Convolucional (CNN) utilizando bibliotecas de *machine learning* como TensorFlow e Keras para classificar as imagens em células humanas infectadas e não infectadas.
4. Avaliar o desempenho do modelo utilizando métricas como acurácia, precisão, sensibilidade e especificidade, comparando os resultados com métodos tradicionais de diagnóstico.
5. Validar o modelo em um conjunto de dados independente e analisar possíveis erros para refinar o modelo e aumentar a robustez da classificação.
6. Elaborar uma *interface web* automatizada que permita a integração do modelo a sistemas de diagnóstico e pesquisa de campo além de ser de fácil utilização para pessoas leigas.

### 3. MATERIAL E MÉTODOS

#### 3.1 Base de Dados Utilizada

A base de dados utilizada para este trabalho consiste em imagens microscópicas de amostras de sangue humano, classificadas como células saudáveis e células infectadas com o parasita causador da malária (*Plasmodium*). Essas imagens foram obtidas a partir do conjunto de dados público Malaria Cell Images Dataset, fornecido pelo National Institutes of Health (NIH).

O conjunto de dados contém um total de 27.527 imagens, divididas em duas classes sendo elas células saudáveis: 13.789 imagens como as representadas na figura 5 e células infectadas: 13.778 imagens como as representadas na figura 6. As imagens foram capturadas com microscopia de campo claro em um aumento de 100x, e variam em tamanho. Para uniformizar o processamento, todas as imagens foram redimensionadas para 150x150 pixels antes do treinamento do modelo.

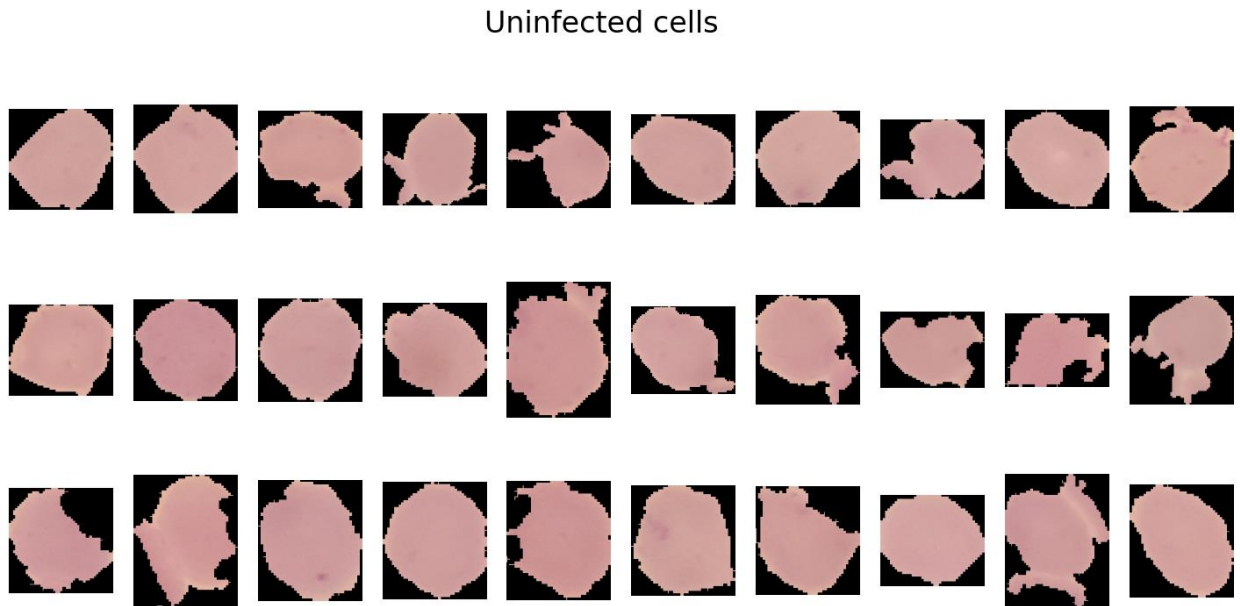


**Figura 5:** Imagens saudáveis sem contaminação por *Plasmodium sp.* Fonte: Imagem adaptada de Malária Cell Images Dataset (<https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html>.)

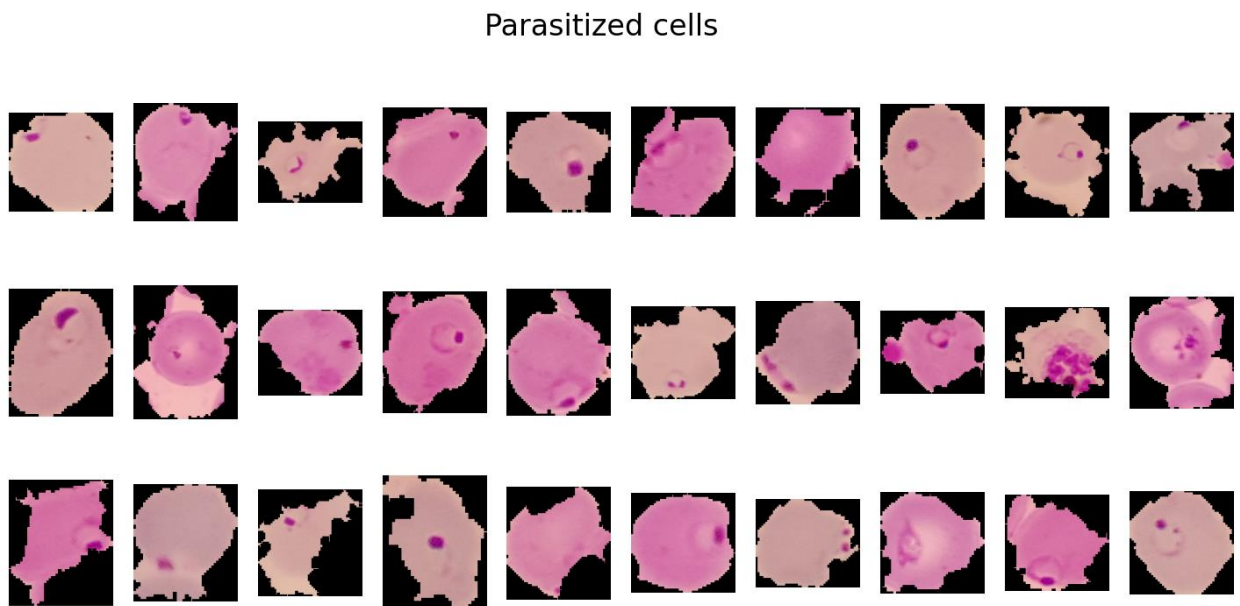


**Figura 6:** Imagens contaminadas por *Plasmodium sp.* Fonte: Imagem adaptada de Malária Cell Images Dataset (<https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html>).

### 3.2 Ferramentas e Bibliotecas Utilizadas

O desenvolvimento do script para a CNN foi realizado utilizando as seguintes ferramentas e bibliotecas:

#### 3.2.1 Ambiente de desenvolvimento:

- **Google Colab:** é uma plataforma de anotação baseada em nuvem fornecida pelo Google que permite gerenciar e compartilhar código Python. É amplamente

utilizado para projetos de mineração de dados e aprendizado de máquina porque fornece poder de computação, como GPU e TPU, sem exigir instalação.

### 3.2.2 Linguagem de programação:

- **Python 3.x.:** É uma linguagem de programação de alto nível e uso geral amplamente usada em áreas como ciência da computação, inteligência artificial e desenvolvimento web. A versão 3.x melhora as versões anteriores e inclui suporte para pacotes e bibliotecas adicionais voltados para aprendizado de máquina, computação e automação.

### 3.2.3 Bibliotecas:

- **TensorFlow** (uma biblioteca de código aberto para aprendizado de máquina e redes neurais);
- **Keras** (uma API de alto nível para facilitar a construção de modelos de aprendizado profundo): Para construção e treinamento de redes neurais convolucionais (CNNs);
- **NumPy** (uma biblioteca que oferece suporte para arrays multidimensionais e várias operações matemáticas eficientes): Para manipulação de arrays e cálculos matemáticos;
- **OpenCV** (uma biblioteca de visão computacional para processamento de imagens) e PIL (Python Imaging Library) (uma biblioteca para manipulação e processamento de imagens em Python): Para pré-processamento de imagens;
- **Matplotlib** (uma biblioteca para criação de gráficos estáticos, animados e interativos) e Seaborn (uma biblioteca de visualização de dados baseada no Matplotlib que oferece uma interface de alto nível para gráficos estatísticos): Para visualização de gráficos e análise de desempenho;
- **Scikit-learn** (biblioteca de aprendizado de máquina para Python que oferece ferramentas para avaliação de modelos): Para avaliação do modelo com métricas de performance.

## 3.3 Processamento das imagens

As seguintes etapas de pré processamento foram aplicadas para preparar os dados de entrada e otimizar o treinamento do modelo:

### 3.3.1 Redimensionamento:

- Todas as imagens foram ajustadas para 150x150 pixels.

### 3.3.2 Normalização:

- Os valores dos pixels foram normalizados para o intervalo  $[0, 1]$ .

### 3.3.3 Aumento de Dados (*Data Augmentation*):

- Técnicas como zoom, espelhamento horizontal, deslocamento e rotações foram aplicadas para expandir o conjunto de treinamento e reduzir o risco de overfitting.

## 3.4 Construção e Treinamento do Modelo

### 3.4.1 A arquitetura da CNN

- Camadas Convolucionais: Três camadas convolucionais com filtros  $3 \times 3$ , ou seja, a imagem é lida de 9 em 9 pixels, e função de ativação ReLU que seria a função responsável para ligar e desligar sinais ajudando a máquina a aprender padrões e evita processamentos desnecessários caso certa parte da imagem não apresente tanta relevância, extração de características.
- Camadas de *Pooling*: tem por função retirar uma amostra de uma imagem utilizando somente suas partes importantes, diminuindo a dificuldade e complexidade do processamento pela máquina.
- Camadas de *MaxPooling*  $2 \times 2$ : caminha sobre a imagem de 4 em 4 pixels e foca nos detalhes mais “fortes” ou importantes, ajudando a diminuir a imagem sem perda das informações principais.
- Camadas *Dropout*: Aplicado *dropout* com taxa de 0,2 para evitar *overfitting*, ou seja, desliga aleatoriamente 20% dos neurônios para evitar que o modelo fique “viciado”.
- Camadas Densas: Uma camada totalmente conectada de 128 neurônios e uma camada de saída com ativação sigmoide para classificação binária (célula saudável ou infectada).

### 3.4.2 Parâmetros de treinamento:

Durante o treinamento do modelo de CNN, foram utilizados os seguintes parâmetros para otimizar o desempenho e garantir um aprendizado eficaz:

- Otimizador Adam: Um otimizador adaptativo eficiente, amplamente utilizado em aprendizado profundo, que ajusta dinamicamente a taxa de aprendizado para cada

parâmetro.

- Função de perda (*Binary Crossentropy*): Métrica de perda usada para problemas de classificação binária, comparando as previsões do modelo com os rótulos verdadeiros e ajustando os pesos com base nos erros.
- *Epochs* (10): Número de vezes que o algoritmo percorre todo o conjunto de dados de treinamento.
- Tamanho do lote (*batch size* de 100): Número de amostras processadas antes de o modelo atualizar seus parâmetros, influenciando a memória e a estabilidade do treinamento.

### 3.5 Divisão dos Dados

A divisão dos dados foi realizada da seguinte forma:

- 70% para treinamento sendo 19.268 imagens.
- 30% para validação sendo 8.258 imagens.

A separação foi realizada de forma estratificada para garantir a representatividade proporcional de ambas as classes.

### 3.6 Métricas de Avaliação

- Acurácia: mede a proporção de previsões corretas (verdadeiras positivas e verdadeiras negativas) em relação ao total de previsões realizadas. Ela indica a porcentagem de classificações que o modelo acertou.
- Precisão: mede a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo.
- *Recall*: mede a proporção de células infectadas corretamente identificadas pelo modelo em relação ao total de células infectadas presentes no banco de dados.
- *F1-Score*: é a média harmônica entre precisão e recall, proporcionando um equilíbrio entre as duas métricas. Ele é útil quando se deseja levar em consideração tanto falsos positivos quanto falsos negativos.
- Matriz de Confusão: A matriz de confusão é uma tabela que permite visualizar o desempenho do modelo, mostrando a distribuição das previsões em relação aos rótulos verdadeiros.

Essas métricas foram utilizadas para avaliar o desempenho do modelo na diferenciação entre células saudáveis e infectadas.

### 3.7 Criação de uma aplicação web

Depois de obter o modelo treinado gerou-se um modelo pré processado (onde todas as informações ficaram armazenadas no mesmo em formato de matriz e quando adiciona uma nova imagem ele a transforma em matriz e identifica por padrão qual é a mais semelhante e a classifica entre *infected* e *uninfected*).

#### 3.7.1 Normas de Interação Humano-Computador (IHC)

Para facilitar a criação da aplicação e garantir uma boa usabilidade, foram seguidas as normas de Interação Humano-Computador (IHC). Essas normas têm como objetivo otimizar a interação entre o usuário e o sistema, proporcionando uma experiência mais intuitiva e eficiente. As seguintes tecnologias foram utilizadas no desenvolvimento da aplicação:

- **Linguagem *Back-end* (Python 3.12.x):** Utilizada para o tratamento das requisições no servidor e manipulação de dados.
- **Framework *Flask*:** Empregado para facilitar a integração entre o *back-end* e a interface web, permitindo a construção de aplicações web de forma simples e eficiente.

#### 3.7.2 Ferramentas utilizadas na página web:

- **HTML *Markup Language* (*Hypertext Markup Language*):** A marcação principal linguagem usada para criar e exibir conteúdo na web. Define a estrutura básica de uma página web e pode incluir textos, imagens, links, tabelas e outros elementos. HTML é responsável por organizar o conteúdo em seções e decidir o que exibir e como classificá-lo no navegador.
- **Case Code Language (CSS):** é uma linguagem usada para definir o estilo e o layout de páginas da web. É usado para controlar a cor, fonte, espaçamento, tamanho dos elementos e posicionamento dos elementos em uma página web. CSS pode ser usado para garantir que um site tenha um design atraente e seja compatível com diferentes dispositivos (responsivos), como smartphones, tablets e computadores desktop.
- **Bootstrap Framework:** é um *framework front-end* amplamente utilizado para simplificar o desenvolvimento de sites e aplicações web responsivos. Ele fornece um conjunto de componentes prontos, como botões, menus de navegação, grades de layout e outros componentes para criar interfaces compatíveis para computadores e dispositivos móveis. Além disso, o *Bootstrap* ajuda a manter a consistência no design e na funcionalidade das páginas da web.
- **Linguagem *Front-End* (JS - *Javascript*):** é uma linguagem de programação

usada no desenvolvimento *front-end* (lado do cliente) para adicionar interação e dinâmica às páginas da web. Com *Javascript* é possível criar efeitos como menus interativos, validação de formulários, atualizar conteúdo sem recarregar a página (AJAX) e manipular elementos HTML/CSS no período. É importante para melhorar a experiência do usuário e o desempenho de muitas novidades na internet.

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Qualidade e Diversidade do Conjunto de Dados

Das 27.527 imagens empregadas no treinamento do modelo, 13.789 correspondem a células humanas saudáveis, enquanto 13.787 representam células humanas infectadas pelo parasita *Plasmodium sp.* A quantidade de imagens em ambas as classes demonstrou ser homogênea para o treinamento do modelo, uma vez que as classes apresentaram quantidades equilibradas de amostras.

### 4.2 Eficácia das Técnicas de Pré-processamento

A alteração das imagens para 150X150 *pixels* preservou a qualidade e a facilidade de processamento, com perda mínima das informações relevantes. A normalização dos valores dos *pixels* melhorou a convergência e a estabilidade durante o treinamento.

### 4.3 Desempenho e Arquitetura da CNN

A arquitetura da CNN apresentou um bom desempenho, com camadas convolucionais de 3x3 que extraíram características visuais essenciais das imagens microscópicas. As camadas de *pooling* foram eficazes na redução da dimensionalidade, ajudando a preservar informações cruciais enquanto diminuía a complexidade computacional. As camadas de *dropout* foram essenciais para mitigar o *overfitting*, o que é fundamental em conjuntos de dados limitados. A camada densa final permitiu a classificação eficaz entre as classes.

### 4.4 Análise de Métricas de Desempenho

#### 4.4.1 Treino

Após o treinamento do algoritmo foram aplicadas métricas para medir seu desempenho sendo elas acurácia e validação loss. No qual a acurácia quantifica a quantidade de acertos obtida

pelo modelo, já o loss identifica quanto o algoritmo erra nas previsões de dados que ele não viu durante o treinamento, ou seja, os dados da validação. Estas métricas foram ilustradas na Figura 7 e 8 constatando que o modelo na etapa de treinamento obteve-se 93.65% de acurácia.

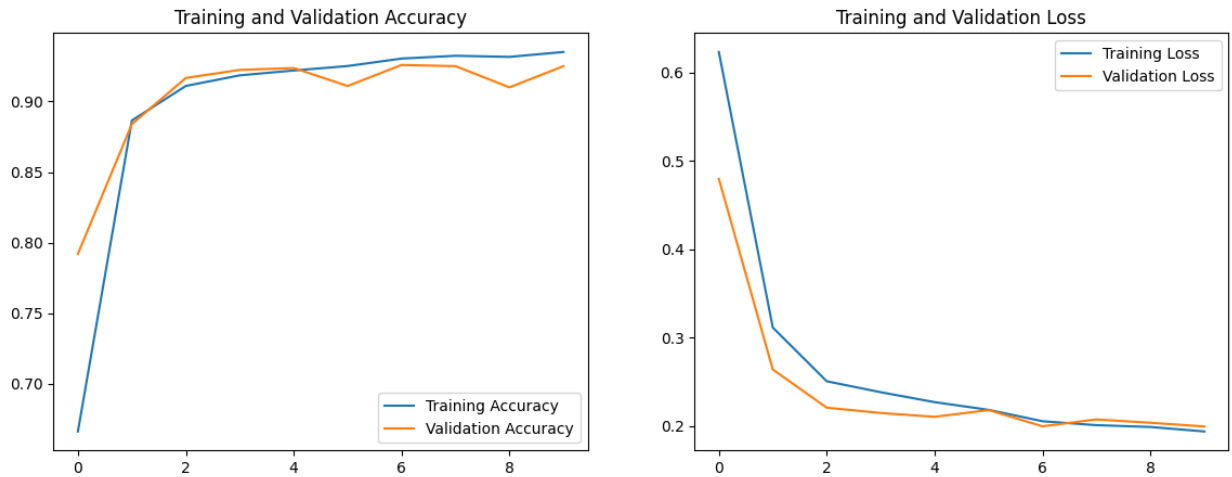


**Figura 7:** Gráfico da Acurácia e Loss no treino e validação do modelo. Fonte: Autoria Própria

#### 4.4.2 Teste

O modelo proposto obteve uma acurácia de 93,21%, demonstrando um desempenho robusto na tarefa de classificação de imagens microscópicas de malária. Esse resultado é superior ao obtido por (RAFIK, 2021) que utilizou uma abordagem similar baseada em CNN para a detecção de parasitas da malária, obtendo uma acurácia de 86%.

```

193/193 ————— 483s 2s/step - accuracy: 0.9347 - loss: 0.1945
Acurácia no conjunto de treino: 93.65%
83/83 ————— 206s 2s/step - accuracy: 0.9330 - loss: 0.1821
Acurácia no conjunto de teste: 93.21%
```

**Figura 8:** Acurácia e Loos do modelo. Fonte: Autoria Própria

#### 4.5 Modelo Final

Logo após verificar as métricas gerou-se um modelo processado que é salvo para evitar a necessidade da repetição do treinamento sempre que for classificar imagens novas. Com esse modelo processado a predição do parasita em novas imagens é facilitada.

#### 4.6 Desafios e Limitações

O modelo ficou sendo treinado por 12 horas devido o tamanho da base de dados e a classificação de imagens demandarem muito tempo visto as etapas da construção de uma

CNN, podendo ser melhorado com a utilização de um cluster (conjunto de computadores) ou um servidor (único computador ou servidor dedicado).

#### 4.7 Desenvolvimento e Usabilidade da Aplicação Web

A criação da aplicação web gerou um site no qual é possível selecionar as imagens desejadas e obter de volta sua predição.

##### 4.7.1 Página Inicial

Na página inicial ilustrada pela Figura 9, o usuário tem seu primeiro contato com a interface onde de maneira intuitiva pode clicar na barra de navegação e selecionar a opção classificação.

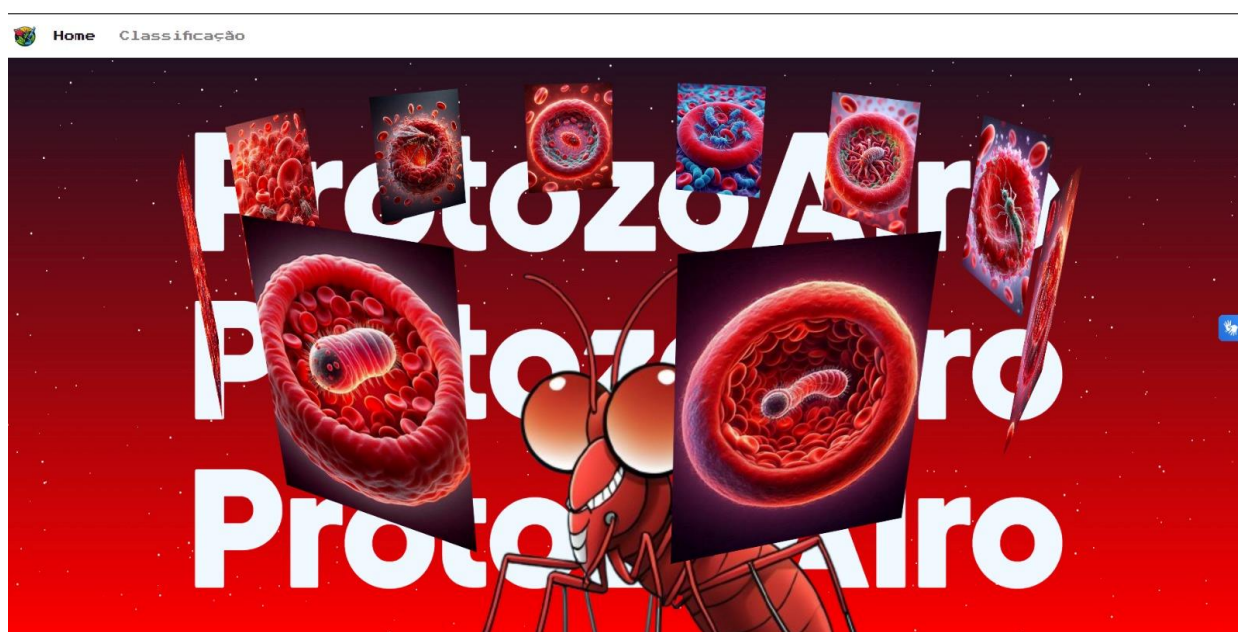


**Figura 9:** Página inicial da Interface Web. Fonte: Autoria Própria

##### 4.7.2 Página de predição

Quando o usuário clicar em classificação a tela representada pela Figura 10 se abrirá para ele de modo que possa selecionar a imagem desejada.

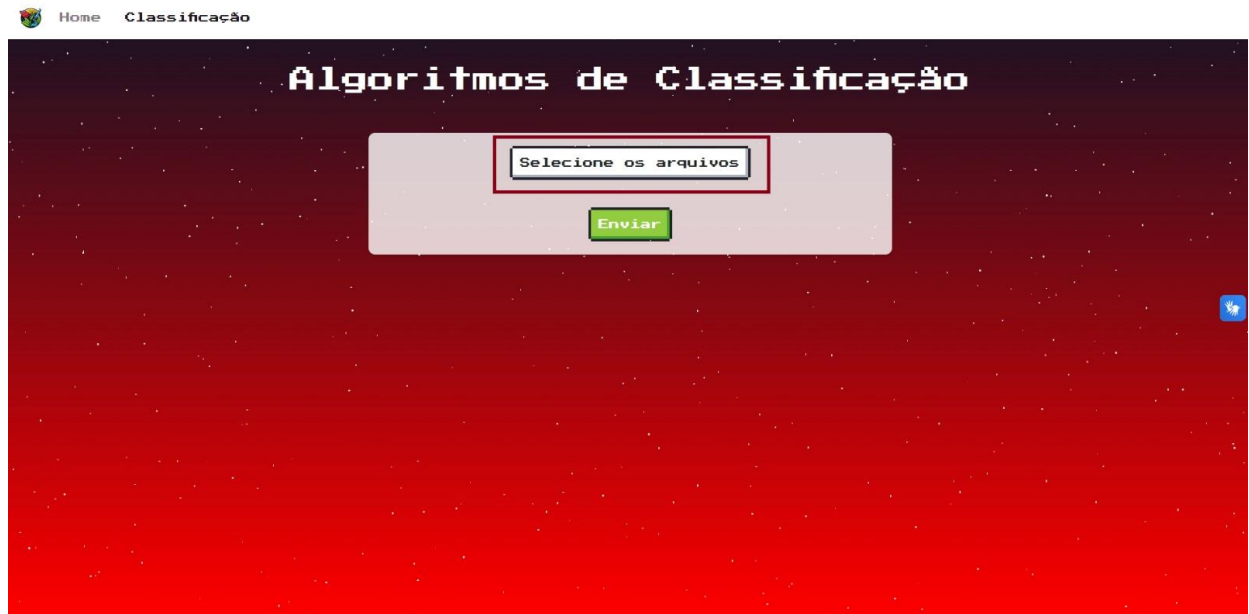


**Figura 10:** Página de Classificação. Fonte: Autoria Própria

Ao clicar no botão “Selecionar Arquivo” mostrado na Figura 11, o explorador de arquivos do usuário será aberto possibilitando a escolha da imagem a ser analisada.



**Figura 11:** Abertura do Explorador de Arquivos. Fonte: Autoria Própria

Após a seleção das imagens o usuário deve clicar em enviar e automaticamente recebera a predição das imagens selecionadas anteriormente como demonstra a Figura 12.

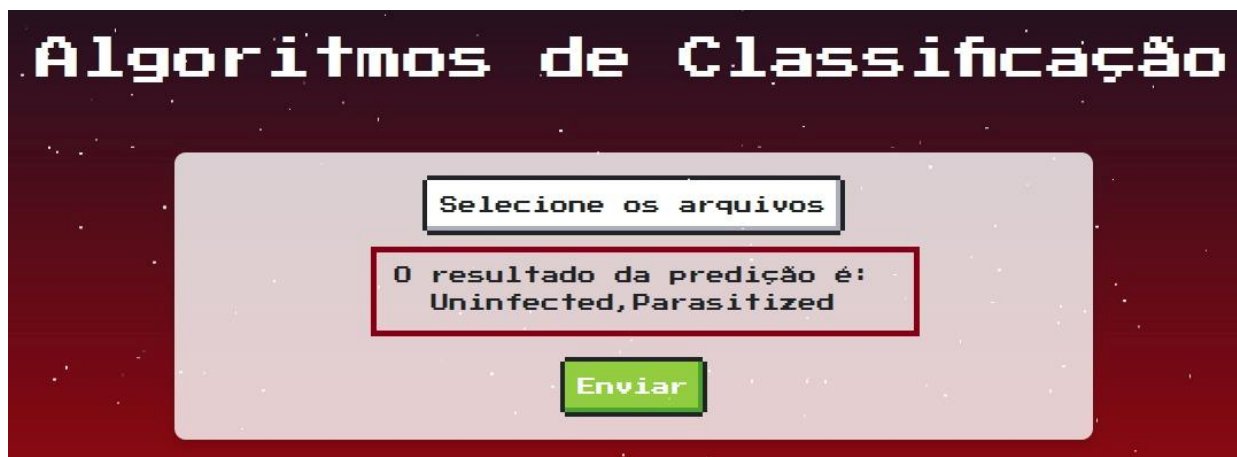


**Figura 12:** Predição final das Imagens. Fonte: Autoria Própria

## 5. CONCLUSÃO

O presente trabalho abordou o desenvolvimento de um modelo de inteligência artificial baseado em Redes Neurais Convolucionais (CNN) para a classificação automática de imagens microscópicas de células infectadas por *Plasmodium* spp., visando facilitar o diagnóstico e controle da malária. Através de um conjunto de dados robusto, contendo 27.527 imagens de células humanas, conseguimos atingir uma acurácia de 93,21% na classificação, demonstrando a eficácia do modelo em comparação com abordagens anteriores.

As etapas de pré-processamento, incluindo redimensionamento, normalização e aumento de dados, foram cruciais para otimizar o desempenho do modelo, permitindo a extração de características visuais essenciais e a mitigação do *overfitting*. A arquitetura da CNN foi projetada de forma a equilibrar complexidade e desempenho, incorporando camadas convolucionais, *pooling* e *dropout*, resultando em um modelo que não apenas apresentou alta acurácia, mas também garantiu estabilidade durante o treinamento.

Além do desenvolvimento do modelo, foi criada uma aplicação web intuitiva, que possibilita a usuários leigos a seleção e análise de imagens, promovendo uma interface de fácil utilização. Essa aplicação representa um passo significativo na integração de tecnologias de

inteligência artificial em ambientes clínicos e de pesquisa, oferecendo uma ferramenta valiosa para o diagnóstico precoce da malária.

Contudo, o projeto enfrentou desafios, como o tempo de treinamento elevado, que poderia ser mitigado pela utilização de recursos computacionais mais robustos, como clusters de máquinas. O trabalho realizado não apenas contribui para a área de saúde pública, mas também abre portas para futuras pesquisas e inovações na aplicação de inteligência artificial em diagnósticos médicos.

Como conclusão, este projeto demonstra que a inteligência artificial, aliada a técnicas de processamento de imagens, pode ser uma ferramenta poderosa no combate à malária, proporcionando um diagnóstico mais rápido e preciso, essencial para o controle eficaz da doença.

## 6. REFERÊNCIAS

CHOI, R. Y. et al. Introduction to machine learning, neural networks, and deep learning. Translational Vision Science and Technology, v. 9, n. 2, 2020.

DAIANA MIOTO, L.; CARLA FACCIN GALHARDI, L.; KARINE AMARANTE, M. Aspectos parasitológicos e imunológicos da malária Parasitological and immunological aspects of malaria Endereço para Correspondência. [s.l: s.n.].

FIKADU, M.; ASHENAFI, E. Malaria: An Overview. Infection and Drug ResistanceDove Medical Press Ltd, , 2023.

GARCIA, L. S. Malaria. Clinics in Laboratory Medicine, v. 30, n. 1, p. 93–129, mar. 2010.

GAUTHIER, J. et al. A brief history of bioinformatics. Briefings in Bioinformatics, v. 20, n. 6, p. 1981–1996, 1 nov. 2019.

IBM. O que é inteligência artificial? .

IBM. O que são redes neurais convolucionais.

MEIBALAN, E.; MARTI, M. Biology of malaria transmission. Cold Spring Harbor Perspectives in Medicine, v. 7, n. 3, 1 mar. 2017.

MINISTÉRIO DA SAÚDE. Malária.

OLIVER PEREIRA. A Máquina de Turing e seu Impacto Fundamental na Computação e Inteligência Artificial.

OMS. OMS.

PAN, T. et al. Bioinformatics Methods for Modeling microRNA Regulatory Networks in Cancer. Em: [s.l: s.n.]. p. 161–186.

PEVSNER, J. BIOINFORMATICS AND FUNCTIONAL GENOMICS third edition. [s.l: s.n.].

RAFIK A. ALIEV, J. K. W. P. M. J. M. B. F. M. S. Malaria Detection Using Convolutional Neural Network. Em: RAFIK A. ALIEV, J. K. W. P. M. J. M. B. F. M. S. (Ed.). 11th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions and Artificial Intelligence - ICSCCW-2021. [s.l: s.n.].

SANCHES, B.; GUSTAVO, B.; CASTRO2, P. Bioinformática: Uma visão geral sobre o campo. [s.l: s.n.].

SATO, S. Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. Journal of Physiological AnthropologyBioMed Central Ltd, , 1 dez. 2021.

SAVI, M. K. An Overview of Malaria Transmission Mechanisms, Control, and Modeling. Medical Sciences, v. 11, n. 1, p. 3, 23 dez. 2022.

WOOLLER, S. K. et al. Bioinformatics in translational drug discovery. Bioscience ReportsPortland Press Ltd, , 31 ago. 2017.



