

**INSTITUTO
FEDERAL**

Goiano

Campus
Rio Verde

MINISTÉRIO DA EDUCAÇÃO
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO - CAMPUS RIO VERDE

**ANÁLISE DE ESPÉCIES DE AMEBAS DE VIDA LIVRE (AVL)
USANDO PROCESSAMENTO DIGITAL DE IMAGENS E
PROJEÇÃO MULTIDIMENSIONAL**

ANA KELLY DE SOUZA FRANCISCO

Rio Verde

2024



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO -
CAMPUS RIO VERDE
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DE ESPÉCIES DE AMEBAS DE VIDA LIVRE (AVL)
USANDO PROCESSAMENTO DIGITAL DE IMAGENS E
PROJEÇÃO MULTIDIMENSIONAL**

ANA KELLY DE SOUZA FRANCISCO

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia Goiano - Campus Rio Verde, como requisito parcial para a obtenção do Grau de Bacharelado em Ciência da Computação.

Orientador: Prof. DOUGLAS CEDRIM OLIVEIRA

Rio Verde

2024

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

FAN532 Francisco, Ana Kelly de Souza
a Análise de espécies de amebas de vida livre (AVL)
usando processamento digital de imagens e projeção
multidimensional / Ana Kelly de Souza Francisco;
orientador Douglas Cedrim Oliveira. -- Rio Verde,
2024.
65 p.

TCC (Graduação em Ciência da Computação) --
Instituto Federal Goiano, Campus Rio Verde, 2024.

1. Amebas de vida livre. 2. Análise morfológica.
3. Descritores de imagem. 4. Projeção
multidimensional. 5. Análise visual de dados. I.
Cedrim Oliveira, Douglas, orient. II. Título.

USANDO PROCESSAMENTO DIGITAL DE IMAGENS E PROJEÇÃO MULTIDIMENSIONAL

Restrições de Acesso ao Documento

Documento confidencial: Não Sim

Justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 06/03/2024

O documento está sujeito a registro de patente? Sim Não

O documento pode vir a ser publicado como livro e/ou artigo? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a docente e/ou autor/a declara que:

- 1 - o documento é seu trabalho original, detém os direitos autorais da produção técnica e não infringe os direitos de qualquer outra pessoa ou entidade;
- 2 - obteve autorização de quaisquer materiais incluídos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- 3 - cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, 6 de março de 2024.

(Assinado Eletronicamente)

Ana Kelly de Souza Francisco (Autora)

(Assinado Eletronicamente)

Douglas Cedrim Oliveira (Orientador)

1058004

(Assinatura do Docente, Autor e/ou Detentor dos Direitos Autorais)

Documento assinado eletronicamente por:

- Ana Kelly de Souza Francisco, 2016102192010420 - Discente, em 06/03/2024 23:14:40.
- Douglas Cedrim Oliveira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 06/03/2024 22:39:10.

Este documento foi emitido pelo SUAP em 06/03/2024. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 580706

Código de Autenticação: 3c7b6fb3bc



INSTITUTO FEDERAL GOIANO

Campus Rio Verde

Rodovia Sul Goiana, Km 01, Zona Rural, 01, Zona Rural, RIO VERDE / GO, CEP 75901-970

(64) 3624-1000



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 2/2024 - GGRAD-RV/DE-RV/CMPRV/IFGOIANO

ATA DE DEFESA DE TRABALHO DE CURSO

Ao sétimo dia do mês de fevereiro de 2024, às 14 horas, reuniu-se a banca examinadora composta pelos docentes: Dr. Douglas Cedrim Oliveira (orientador), Dr. André da Cunha Ribeiro (membro), Me. Fábio Montanha Ramos (membro), para examinar o Trabalho de Curso intitulado "ANÁLISE DE ESPÉCIES DE AMEBAS DE VIDA LIVRE (AVL) USANDO PROCESSAMENTO DIGITAL DE IMAGENS E PROJEÇÃO MULTIDIMENSIONAL" da estudante Ana Kelly de Souza Francisco, Matrícula nº 2016102192010420 do Curso de Bacharelado em Ciência da Computação do IF Goiano – Campus Rio Verde. A palavra foi concedida à estudante para a apresentação oral do TC, houve arguição da candidata pelos membros da banca examinadora. Após tal etapa, a banca examinadora decidiu pela APROVAÇÃO da estudante. Ao final da sessão pública de defesa foi lavrada a presente ata que segue assinada pelos membros da Banca Examinadora.

(Assinado Eletronicamente)

Douglas Cedrim Oliveira
Orientador(a)

(Assinado Eletronicamente)

André da Cunha Ribeiro
Membro

(Assinado Eletronicamente)

Fábio Montanha Ramos
Membro

Observação:

() O(a) estudante não compareceu à defesa do TC.

Documento assinado eletronicamente por:

- Douglas Cedrim Oliveira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 07/02/2024 15:58:18.
- Andre da Cunha Ribeiro, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 08/02/2024 08:05:06.
- Fabio Montanha Ramos, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 13/02/2024 07:06:48.

Este documento foi emitido pelo SUAP em 07/02/2024. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 569738

Código de Autenticação: eb178e166c



INSTITUTO FEDERAL GOIANO

Campus Rio Verde

Rodovia Sul Goiana, Km 01, Zona Rural, 01, Zona Rural, RIO VERDE / GO, CEP 75901-970

(64) 3624-1000

AGRADECIMENTOS

Gostaria de dedicar este espaço para expressar minha profunda gratidão a todas as pessoas que contribuíram para a realização deste trabalho de conclusão de curso.

Em primeiro lugar, gostaria de expressar minha profunda gratidão ao meu orientador, Douglas Cedrim Oliveira, por sua orientação excepcional, paciência infinita e apoio incansável ao longo de todo o processo de pesquisa e redação deste trabalho. Sua atenção meticulosa e disposição para ajudar foram verdadeiramente notáveis e fundamentais para o desenvolvimento deste trabalho e para o meu crescimento acadêmico. As valiosas sugestões fornecidas por ele foram como guias importantes, ajudando a moldar não apenas o conteúdo deste trabalho, mas também a minha trajetória acadêmica como um todo. A sua dedicação e comprometimento em compartilhar seu conhecimento e experiência foram inestimáveis e deixaram uma marca que não será apagada em minha formação.

Aos membros da banca examinadora, André da Cunha Ribeiro e Fábio Montanha Ramos, expresso meus sinceros agradecimentos pela disponibilidade em avaliar este trabalho e pelos comentários e sugestões que contribuíram para sua melhoria. Suas contribuições foram essenciais para o aprimoramento deste estudo.

Estou muito grata às Dras. Natália Karla Bellini e Maria Reyes-Batlle, do Laboratório de Amebas de Vida Livre da Universidade de La Laguna, Espanha, pela assistência na obtenção das imagens das AVLS. Foi crucial para este projeto e contribuiu para o fortalecimento da nossa colaboração científica.

À minha família e ao meu Deus, que sempre estiveram ao meu lado, apoiando-me em todas as etapas da minha jornada acadêmica. Seu amor, encorajamento e compreensão foram fundamentais para minha perseverança e sucesso.

Aos meus amigos e colegas de classe, agradeço pela troca de ideias, pelos momentos de descontração e pelo apoio mútuo durante todo o percurso deste trabalho. Suas palavras de incentivo e apoio foram uma fonte constante de motivação.

Por fim, agradeço a todos os professores, funcionários e demais pessoas que, de alguma forma, contribuíram para a realização deste trabalho, mesmo que não mencionados especificamente. Cada interação, conselho e apoio recebidos agregaram para minha formação e crescimento pessoal e profissional.

A todos, expresso minha mais sincera e profunda gratidão.

RESUMO

FRANCISCO, A. K. de S.. **Análise de espécies de amebas de vida livre (AVL) usando processamento digital de imagens e projeção multidimensional.** 2024. 65 f. Monografia – (Curso de Bacharelado em Ciência da Computação), Instituto Federal de Educação, Ciência e Tecnologia Goiano - Campus Rio Verde. Rio Verde.

Este trabalho tem como objetivo principal abordar a análise morfológica de espécies de Amebas de Vida Livre (AVL) por meio de ferramentas digitais. O processo inclui coleta de imagens, processamento por segmentação, aplicação de descritores morfológicos matemáticos e projeção multidimensional para simplificar a visualização do conjunto de dados. O sistema proposto facilita a inserção de dados organizados em pastas representando classes específicas de amebas de vida livre. Ele permite a análise simultânea de até 12 classes para usuários regulares e até 4 classes para usuários com daltonismo. Cada espécie é representada por um vetor de características (descritores morfológicos) e a análise visual é facilitada por projeção multidimensional dos dados. Além disso, o sistema permite ajustar pesos para cada descritor usando controles deslizantes.

Palavras-chave: Amebas de vida livre. Análise morfológica. Descritores de imagem. Projeção multidimensional. Análise visual de dados.

ABSTRACT

FRANCISCO, A. K. de S.. **Analysis of Free-Living Amoeba (FLA) Species Using Digital Image Processing and Multidimensional Projection.** 2024. 65 f. Trabalho de Conclusão de Curso – Bacharelado em Ciência da Computação, Instituto Federal de Educação, Ciência e Tecnologia Goiano - Campus Rio Verde. Rio Verde, 2024.

This work aims to address the morphological analysis of Free-Living Amoeba (FLA) species through digital tools. The process includes image acquisition, image segmentation, application of morphological operators, computation of image descriptors, and multidimensional projection to simplify data visualization. The proposed system facilitates the insertion of organized data into folders representing specific classes of free-living amoebas. It allows simultaneous analysis of up to 12 classes for regular users and up to 4 classes for users with color vision deficiencies. Each species is represented by a vector of characteristics (descriptors), and visual analysis is facilitated by multidimensional projection of the data. Additionally, the system allows for adjusting the influence weights of each descriptor using sliders.

Keywords: Free-living amoebas. Morphological analysis. Image descriptors. Multidimensional projection. Visual Analytics.

LISTA DE FIGURAS

Figura 1 – Representação dos quatro gêneros das AVLs no estágio de cisto :	1
Figura 2 – Microscópio ótico de campo claro.	4
Figura 3 – Células de amebas do gênero <i>Acanthamoeba</i>	5
Figura 4 – Microscopia digital.	5
Figura 5 – Estágios do ciclo de vida de uma <i>Naegleria fowleri</i> . As setas azuis indicam o tempo de transformação de um estágio no outro (i.e., diferenciação).	6
Figura 6 – Imagem digital.	7
Figura 7 – Processo de aquisição de uma imagem digital	8
Figura 8 – Diferentes conversões de uma imagem colorida para tons de cinza. . . .	9
Figura 9 – Imagem ilustrando o preenchimento de buracos utilizando morfologia matemática.	11
Figura 10 – (a) Imagem Original (b) Depois do processo de Erosão (c) Elemento estruturante em forma de cruz.	12
Figura 11 – (a) Imagem Original (b) Depois do processo de Dilatação (c) Elemento estruturante em forma de cruz.	12
Figura 12 – (a) Imagem Original (b) Depois do processo de Abertura (c) Elemento estruturante em forma de cruz.	13
Figura 13 – (a) Imagem Original (b) Depois do processo de Fechamento (c) Elemento estruturante em forma de cruz.	13
Figura 14 – Exemplo de projeção multidimensional utilizando o <i>dataset</i> de dígitos MNIST (LECUN et al., 1998).	15
Figura 15 – Ilustração da estrutura da projeção multidimensional.	16
Figura 16 – Experimento de silhueta.	18
Figura 17 – Principais passos: Imagem original (a), imagem pré-tratado (b), a região de interesse (c), a imagem binária após a segmentação (d), e imagem final (e).	19
Figura 18 – Interface da proposta do iPCA.	20
Figura 19 – Interface da aplicação desenvolvida, mostrando: a projeção inicial; os descritores calculados para cada imagem; os pesos associados a cada descritor e a imagem em si.	29
Figura 20 – Interação possível feita pelo usuário ao selecionar um vetor de características, visualizando seu ponto correspondente na projeção e imagem associada.	30
Figura 21 – Site ColorBrewer.	31
Figura 22 – Dataset contendo as imagens originais.	32
Figura 23 – Segmentação de uma <i>Acanthamoeba</i> utilizando Ginoris et al. (2007). . . .	33
Figura 24 – Segmentação de uma <i>Acanthamoeba</i> utilizando segmentação manual. . . .	33
Figura 25 – Dataset ilustrando as imagens obtidas após segmentação manual	34
Figura 26 – Visualização da projeção de todo o dataset utilizando todos os descritores ativos.	35
Figura 27 – Ilustração visual da distinção entre os ciclos da espécie <i>Naegleria gruberi</i> . Em a) o melhor resultado, em b) o pior resultado.	36
Figura 28 – Apresentação das imagens das células correspondentes ao pior resultado da espécie <i>Naegleria gruberi</i> , Figura 27b.	36

Figura 29 – Apresentação das imagens das células correspondentes ao melhor resultado da espécie <i>Naegleria gruberi</i> , Figura 27a	37
Figura 30 – Ilustração visual da distinção entre os ciclos da espécie <i>Naegleria australiensis</i> . Em a) o melhor resultado, em b) o pior resultado.	38
Figura 31 – Apresentação das imagens das células correspondentes ao pior resultado da espécie <i>Naegleria australiensis</i> , Figura 30b.	38
Figura 32 – Ilustração visual da distinção entre os ciclos da espécie <i>Vermamoeba vermiformis</i>	39
Figura 33 – Ilustração visual da distinção entre o estágio de cisto de todas as espécies abordadas neste trabalho.	41
Figura 34 – Apresentação das imagens das espécies no estágio de cisto correspondentes à Figura 33b.	41
Figura 35 – Apresentação das imagens das espécies no estágio de cisto correspondentes à Figura 33a.	42
Figura 36 – Representação das células da espécies no estágio Trofozoíto. Em a) o melhor resultado, em b) o pior resultado.	43
Figura 37 – Representação de todo o <i>dataset</i> utilizado neste trabalho. Células no estágio: Cisto e Trofozoíto. Em a) o melhor resultado, em b) o pior resultado.	44
Figura 38 – Representação da espécies da Figura 37a.	44
Figura 39 – Representação da espécies da Figura 37b.	45
Figura 40 – Exibição dos dados analisados neste estudo, ressaltando uma possível progressão desta pesquisa em estudos subsequentes.	48
Figura 41 – Interface	53
Figura 42 – Interface	53
Figura 43 – Interface	54
Figura 44 – Interface	54
Figura 45 – Interface	55
Figura 46 – Interface - Passo 1	57
Figura 47 – Interface - Passo 2	57
Figura 48 – Interface - Passo 3	58
Figura 49 – Interface - Passo 4	59
Figura 50 – Interface - Passo 5	60
Figura 51 – Interface - Passo 6	60
Figura 52 – Interface - Passo 7	61
Figura 53 – Interface - Passo 8	61
Figura 54 – Na figura (A) mostra a imagem original. A figura (B) mostra a figura passada pelo processamento automático da Figura 53.	62
Figura 55 – Interface - Passo 1	62
Figura 56 – Interface - Passo 2	63
Figura 57 – Interface - Passo 3	63
Figura 58 – Interface -Passo 4	64
Figura 59 – Interface - Passo 5	64
Figura 60 – Interface -Passo 6	65
Figura 61 – Na figura (A) mostra a imagem original. A figura (B) mostra a figura passada pelo processamento manual da Figura 60.	65

LISTA DE TABELAS

Tabela 1 – Análise comparativa de descritores entre cistos e trofozoítos da mesma espécie.	40
Tabela 2 – Análise comparativa de descritores entre espécies diferentes considerando as formas cisto e trofozoíto.	46

LISTA DE ABREVIATURAS E SIGLAS

AVL	Amebas de Vida Livre
CCD	<i>Charged-Couple Device</i>
MAP	Meningoencefalite Amebiana Primária
MM	Morfologia Matemática
OLI	<i>Observation-level interaction</i>
PCA	<i>Principal Component Analysis</i>
PI	<i>Parametric Interaction</i>
RF	Requisitos Funcionais
RNF	Requisitos Não-Funcionais
ROI	<i>Region of Interest</i>
TIFF	<i>Tagged Image File Format</i>
WMDS	<i>Weighted Multidimensional Scaling</i>

LISTA DE SÍMBOLOS

λ	Comprimento de onda
\in	Pertence
\subset	Subconjunto
\cap	Interseção

SUMÁRIO

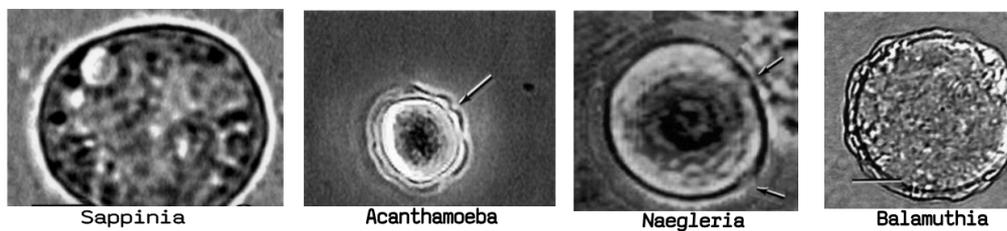
	SUMÁRIO	
1	– INTRODUÇÃO	1
2	– FUNDAMENTAÇÃO TEÓRICA	4
2.1	Microscópio óptico de Campo Claro	4
2.2	Amebas de vida livre	6
2.3	Processamento digital de imagens	7
2.3.1	Representação de imagens digitais	7
2.3.2	Imagem em tons de cinza	8
2.3.3	Filtragem de imagem	9
2.3.4	Segmentação da Imagem	10
2.3.5	Análise Morfológica	10
2.3.6	Operadores Morfológicos	10
2.3.7	Descritores	13
2.4	Projeção multidimensional	14
2.4.1	Técnicas de projeção	15
2.4.2	Métrica de consistência	16
3	– TRABALHOS RELACIONADOS	19
4	– METODOLOGIA	22
4.1	Aquisição das imagens	22
4.2	Requisitos do sistema	23
4.2.1	Requisitos funcionais	23
4.2.2	Requisitos não-funcionais	23
4.3	Pré-Processamento	24
4.3.1	Segmentação	24
4.4	Extração de características das amebas	25
4.4.1	Descritores utilizados	25
4.4.2	Determinando a relevância de cada descritor	27
4.5	Projeção multidimensional	27
4.5.1	PCA	28
4.6	Interface do sistema	28
4.6.1	Qt Designer	28
4.6.2	Estrutura da interface	29
4.7	Outras tecnologias utilizadas	31
5	– RESULTADOS	32
5.1	Geração do <i>dataset</i>	32
5.2	Análise das projeções multidimensionais	33
5.2.1	Projeções por espécie	35
5.2.1.1	<i>Naegleria gruberi</i>	35
5.2.1.2	<i>Naegleria australiensis</i>	37
5.2.1.3	<i>Vermamoeba vermiformis</i>	39

5.2.2	Projeções por estágio celular	40
5.2.2.1	Cisto	40
5.2.2.2	Trofozoíto	42
5.2.2.3	Cisto e trofozoíto	43
6	– CONCLUSÃO	47
6.1	Considerações Finais	47
6.2	Trabalhos Futuros	47
	REFERÊNCIAS	49
	APÊNDICES	52
	APÊNDICE A – INSTRUÇÕES DE USO DO SISTEMA	53

1 INTRODUÇÃO

As amebas de vida livre (AVL) são protozoários amplamente disseminados na natureza, sendo encontrados em uma variedade de ambientes, tais como solo, poeira, ar, água do mar, água potável, piscinas e sistemas de esgoto. Entre as AVLs, destacam-se apenas quatro gêneros: *Sappinia*, *Acanthamoeba*, *Naegleria* e *Balamuthia* (TRABELSI et al., 2012).

Figura 1 – Representação dos quatro gêneros das AVLs no estágio de cisto :



Fonte: (SCHUSTER; VISVESVARA, 2004)

Naegleria fowleri é uma das espécies mais reconhecidas dentro do gênero *Naegleria*, devido à sua capacidade de infectar seres humanos. Esta ameba pode ser encontrada em diversas fontes, incluindo piscinas aquecidas mal tratadas, poças d'água, lagos, córregos, solo e até mesmo partículas de poeira. Popularmente, é conhecida como “ameba comedora de cérebro” (SCHUSTER; VISVESVARA, 2004). A infecção ocorre quando a ameba penetra no corpo do hospedeiro pelo nariz, geralmente quando a pessoa mergulha em água contaminada e a inala. Ela então atinge o sistema nervoso central e se instala no cérebro, resultando na doença conhecida como Meningoencefalite Amebiana Primária (MAP). Apesar de sua importância médica, ainda há muito a ser descoberto sobre essa ameba em termos de pesquisa bioquímica. A *Naegleria gruberi* é frequentemente usada como modelo em estudos relacionados ao gênero *Naegleria*, uma vez que não é patogênica. Isso tem proporcionado contribuições significativas para o estudo de eucariotos (SCORTECCI, 2019).

Mais de 300 casos de Meningoencefalite Amebiana Primária foram registrados internacionalmente, com maior incidência nos Estados Unidos, Austrália e países europeus. Na França, foi relatado o primeiro caso fatal causado por *Naegleria fowleri* em um menino de 9 anos que não apresentava problemas de imunidade. A MAP é uma infecção letal que afeta o sistema nervoso central e, geralmente, ocorre em crianças saudáveis e adultos jovens.

Essa condição é típica do verão, quando as pessoas participam de atividades recreativas em água doce aquecida e poluída (TRABELSI et al., 2012).

Em 2020, um caso de Meningoencefalite Amebiana Primária foi confirmado no estado da Flórida, EUA, de acordo com as autoridades locais. No entanto, não foram divulgadas informações sobre o estado de saúde do paciente ou sobre a maneira pela qual ele contraiu a ameba (GLOBO, 2020). No Brasil, existem diversos tipos de amebas de vida livre (BELLINI et al., 2020), incluindo a *Naegleria fowleri*. No decorrer deste ano, a renomada parasitologista brasileira Denise Leal dos Santos, uma das principais pesquisadoras sobre *Naegleria fowleri* no país, alertou que “Podemos ter casos no Brasil não diagnosticados”, uma vez que a doença pode ser confundida com outras formas de meningite, resultando em casos sub-diagnosticados causados pela *Naegleria fowleri* que não foram devidamente identificados. Ela também enfatizou os riscos associados às mudanças climáticas (GLOBO, 2023).

O uso de ferramentas tecnológicas desempenha um papel de extrema importância no apoio às pesquisas científicas na área biológica, especialmente no estudo de organismos potencialmente prejudiciais à saúde humana. Considerando a alarmante estatística de que cerca de 97% dos infectados acabam perdendo suas vidas (GLOBO, 2020). A finalidade deste projeto é disponibilizar recursos computacionais que possam ter um impacto relevante nas pesquisas associadas a esse parasita no âmbito da biologia.

Com base nessa motivação, o propósito deste estudo é criar um sistema que viabilize a análise morfológica de algumas espécies de amebas de vida livre (AVL) por meio de imagens, utilizando descritores como ferramentas auxiliares na identificação das espécies. Essa abordagem representa um avanço na elaboração de chaves de classificação, pois permite a definição dos melhores descritores morfológicos capazes de distinguir entre duas ou mais espécies distintas.

Neste projeto, foram coletadas imagens de diferentes espécies de AVL para análise por meio do protótipo desenvolvido. Conforme será demonstrado ao longo deste estudo, é possível obter resultados promissores na diferenciação das espécies com um número reduzido de descritores.

O restante desta monografia está estruturada da seguinte forma:

- **Capítulo 2:** apresenta os conceitos e fundamentos que fornecem base teórica para o desenvolvimento deste trabalho;
- **Capítulo 3:** apresenta o processo de aquisição de imagens, bem como o conjunto de imagens coletadas, e as decisões de projeto que definiram a solução proposta no trabalho;
- **Capítulo 4:** são apresentados os resultados mais relevantes obtidos, onde é feita uma discussão quantitativa e qualitativa deles;

- **Capítulo 5:** apresenta as considerações finais, limitações e ideias para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Microscópio óptico de Campo Claro

O microscópio óptico foi inventado para que fosse possível ampliar imagens de objetos não visíveis a olho nu. A resolução dos microscópios atuais é cerca de mil vezes mais que o olho humano. O seu principal objetivo é criar uma imagem real do objeto examinado, quando se observa através da lente ocular se vê uma imagem virtual aumentada da imagem real, ele é o mais usado em diversas áreas, quando se trata de pesquisas e análises. O mesmo é composto por dois jogos de lentes, sendo elas a objetiva e ocular, montadas em extremos opostos de um tubo fechado. As amostras observadas em um microscópio são transparentes, para sua observação é necessário colocar as amostras em uma lâmina que permita que a luz transpasse esta amostra possibilitando assim sua visualização (LEVY et al., 2012).

Na microscopia óptica de campo claro (MOCC), utiliza-se iluminação convencional e os espécimes são preparados com colorações específicas que têm como características a absorção de luz. O tecido é normalmente montado entre uma lâmina de vidro e uma lamela e colocado no microscópio, na platina, para visualização, como pode ser visto na Figura 2.

Figura 2 – Microscópio ótico de campo claro.



Fonte: Extraída de Labor (2016).

Em microscopia óptica, podemos ter diferentes tipos de iluminação utilizadas pelo equipamento: luz refletida, utilizada majoritariamente para amostras opacas; luz transmitida, utilizada majoritariamente para amostras que são (semi-)transparentes, sendo

essa a considerada ao longo desse trabalho. Nesse tipo a luz da fonte é focada pelo condensador para a amostra, as lentes da objetiva recebem a luz transmitida pelo objeto e focam-na para o plano focal das lentes oculares, criando uma imagem ampliada. As imagens desse tipo terão um fundo claro e regiões mais densas da espécie mais escuras, pois irão absorver parte da luz transmitida, como pode ser visto na Figura 3, onde é exibida uma imagem contendo células de amebas do gênero *Acanthamoeba*.

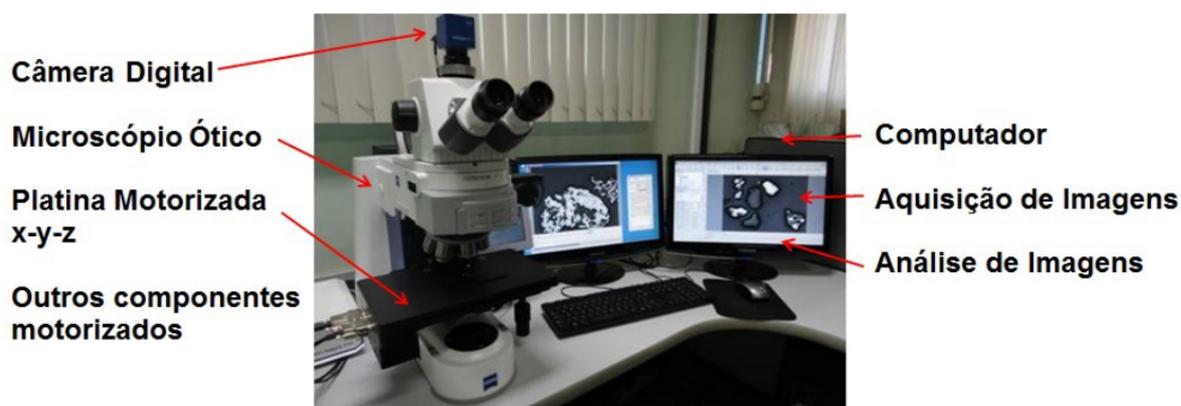
Figura 3 – Células de amebas do gênero *Acanthamoeba*



Fonte: Elaborado pela autora.

Alguns microscópios possuem um diafragma, que dificultam a passagem de luz, tornando a imagem mais ou menos luminosa (LEVY et al., 2012). Quando há uma integração do microscópio com uma câmera digital e um computador, a passagem da luz atinge o sensor da câmera, possibilitando que seja feita a captura em meio digital. Esse processo é chamado de *microscopia digital*, e segue ilustrado na Figura 4.

Figura 4 – Microscopia digital.



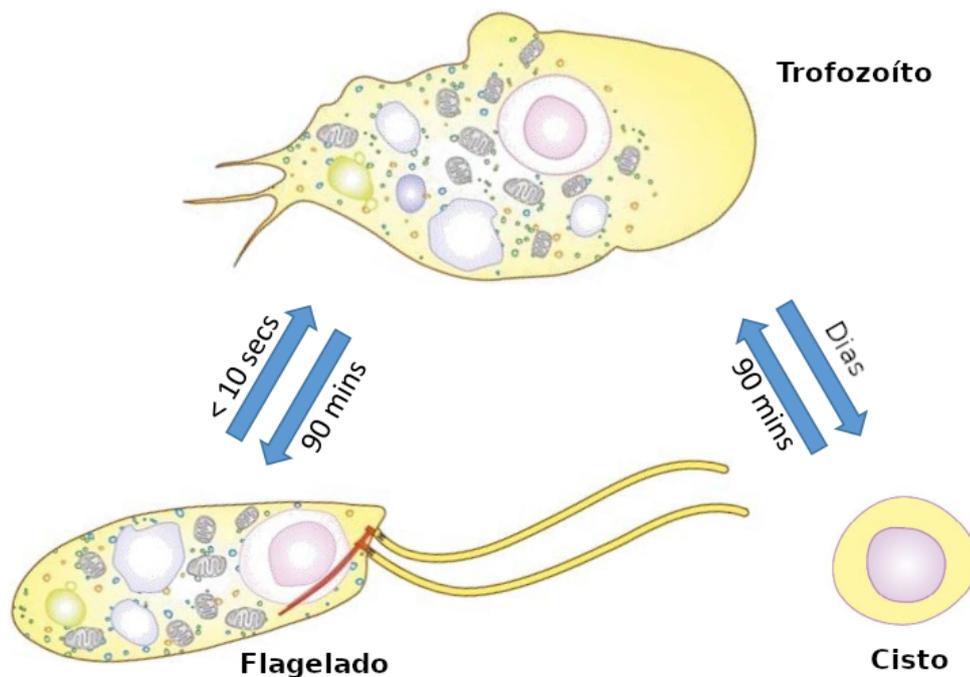
Fonte: Extraída de Mantilla (2013).

2.2 Amebas de vida livre

Amebas de vida livre pertencem ao domínio *Eukaryota* e ao reino *Protozoa*. São amplamente distribuídas na natureza principalmente as pertencentes ao gênero *Acanthamoeba*, que habitam em qualquer ambiente, e as do gênero *Naegleria*, que preferem ambientes aquáticos, natural ou artificialmente aquecidos. Divide-se em três estágios no seu ciclo de vida (TEIXEIRA et al., 2008):

- **Cisto:** Estágio em que a célula está em repouso, capaz de sobreviver desidratada por vários anos em diversos habitats, incluindo condições de temperaturas extremas ou escassez de alimentos. Durante esse período, seu corpo adota uma forma que demanda menos energia, ou seja, uma forma arredondada, conforme ilustrado microscopicamente na parte inferior direita da Figura 5, em um formato circular.
- **Flagelado:** Estágio em que surgem duas pequenas caudas (flagelos) para facilitar o deslocamento em busca de alimento a distâncias maiores, conforme ilustrado na parte inferior esquerda da Figura 5.
- **Trofozoíto:** Estágio no qual sua forma é disforme, sem caudas, mas ainda assim há o movimento em pequenas distâncias no meio, com o objetivo de alimentar-se de bactérias ou alguns outros tipos de células. Também conhecido como estágio *ameboide*, e ilustrado na parte superior da Figura 5.

Figura 5 – Estágios do ciclo de vida de uma *Naegleria fowleri*. As setas azuis indicam o tempo de transformação de um estágio no outro (i.e., diferenciação).



Fonte: Adaptada de Maciver, Piñero e Lorenzo-Morales (2020).

No contexto desse trabalho, por simplicidade e disponibilidade de aquisição, serão

considerados apenas os estágios de cisto e trofozoíto.

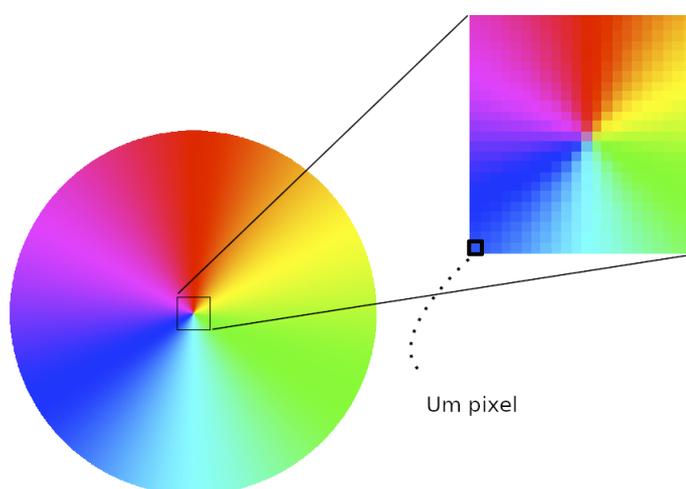
2.3 Processamento digital de imagens

O Processamento Digital de Imagens é quando se utiliza computadores para modificar imagens. Esse método se torna útil em muitas áreas de estudos. Seu objetivo é aprimorar as imagens de acordo com a necessidade em questão, encontrar objetos e reconhecer padrões. Isso ajuda a entender e analisar informações visuais em computadores.

2.3.1 Representação de imagens digitais

Uma imagem digital é representada por duas dimensões de um conjunto de pontos finitos, formando uma matriz $M \times N$, onde cada elemento da matriz é denominado *pixel*. O termo pixel é a junção das palavras em inglês *picture* e *element*, que em definição entende-se como o menor elemento que compõe uma imagem. Pode ser percebido a olho nu ao fazer uma aproximação de uma imagem digital via *zoom*. Com essa ação será visível uma grande quantidade de quadrados em cores possivelmente diferentes, ilustrado na Figura 6.

Figura 6 – Imagem digital.

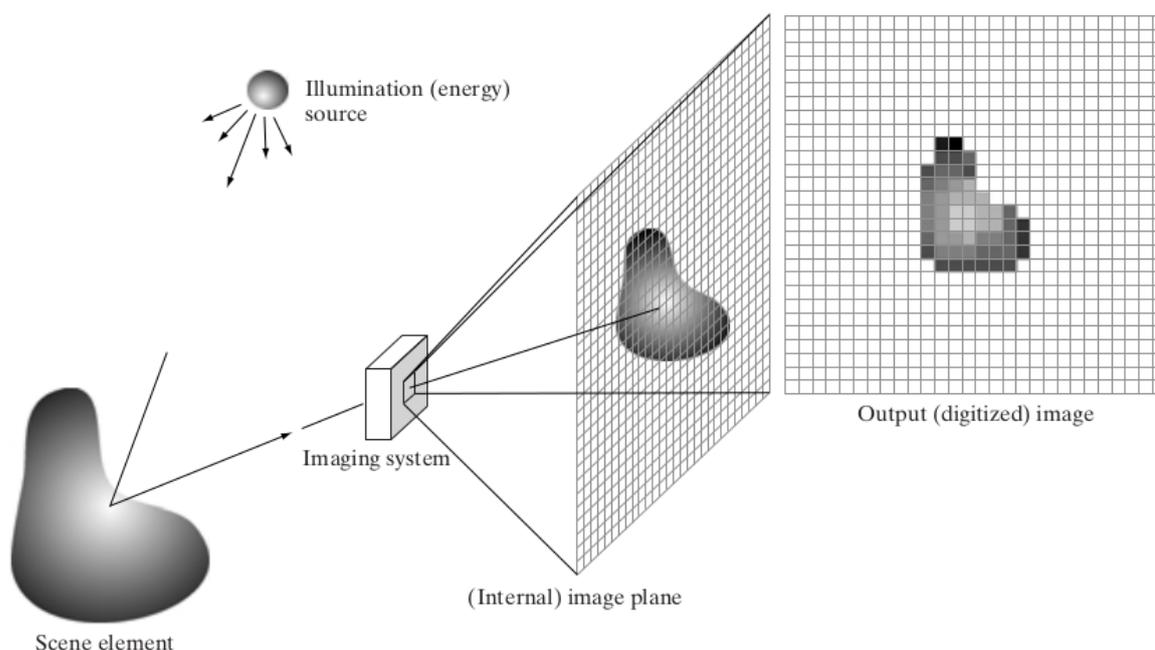


Fonte: Elaborado pela autora.

Sua aquisição dá-se através de um equipamento digitalizador, como por exemplo uma câmera digital. Esta aquisição consiste em converter a cena/imagem que tem uma representação contínua no mundo real para um formato discreto (GONZALEZ; WOODS et al., 2002), como ilustrado na Figura 7.

Cada cor é definida por uma representação discreta da interação da luz com os objetos da cena e com o sensor da câmera na qual a imagem foi capturada. Em geral, para cada pixel da imagem um sensor do tipo CCD de uma câmera digital utiliza filtros para captar a intensidade de luz para três comprimentos de onda diferentes: vermelho, verde

Figura 7 – Processo de aquisição de uma imagem digital



Fonte: Extraído de Gonzalez, Woods et al. (2002).

e azul (*red, green, blue - RGB*, em inglês). Câmeras desse tipo são ditas *câmeras RGB*. Dessa forma, cada pixel da imagem resultante de tal câmera será formado pela combinação da luz em cada uma dessas três componentes.

Esse modelo é inspirado em parte da fisiologia da visão humana, que possui dois tipos de sensores: o primeiro seriam as células bastonetes, que possibilitam a percepção de contraste; e o segundo caracteriza-se pelas células cones, que permitem a percepção das cores através da combinação dos estímulos em três comprimentos de onda dominantes.

Cada componente define o que será chamado um canal de cor: ex. canal R, canal G e canal B, onde a cor final é composta por uma combinação dos três canais. Esse modelo tricromático utiliza assume a formação de cor aditiva, onde a cor final é a soma de cada uma das contribuições individuais na base RGB. Pode ser utilizado em imagens digitais, aparelhos eletrônicos, cinemas e televisões (SCURI, 1999).

Considerando que para representar cada canal computacionalmente sejam utilizados 8 bits, temos 256 valores possíveis por cada canal (2^8 bits). Assim, pode-se representar 16 milhões de cores (256^3). Nessa representação, por exemplo, a cor amarela poderia ser representada por $R = 255$, $G = 255$ e $B = 0$.

2.3.2 Imagem em tons de cinza

Representar imagens em preto-e-branco (e com tons de cinza) basta associar a cada pixel um valor numérico numa escala de tons. Como por exemplo, uma escala com 256

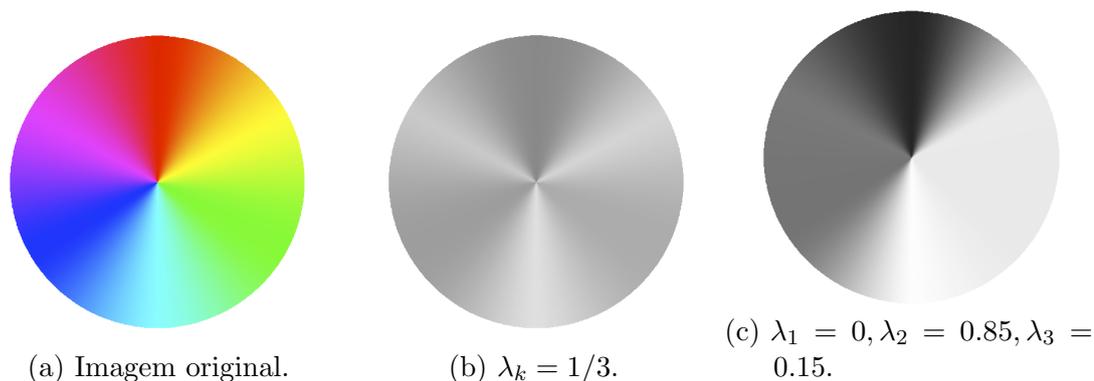
tons de cinzento (do preto absoluto ao branco absoluto) podemos associar a cada pixel um valor inteiro não-negativo de um byte. Convencionamos que o valor 0 corresponde ao preto e 255 corresponde ao branco, os valores intermédios correspondem aos tons de cinzentos (GONZALEZ; WOODS, 2000).

Para a conversão de uma imagem colorida para uma imagem em tons de cinza existem diversas formas de separação de componentes. Um desses métodos é a média dos valores de cada canal para conversão (MENEZES, 2010). Mais especificamente, o valor da cor no pixel (i,j) será dado por

$$C_{ij} = \lambda_1 * R + \lambda_2 * G + \lambda_3 * B. \quad (1)$$

Para diferentes valores de $\lambda_k, k = \{1,2,3\}$, diferentes resultados podem ser obtidos, como pode ser visto na Figura 8. Considerando a mesma quantidade de bits para representação, quando é feita essa conversão em uma imagem RGB passa-se a obter valores que variam entre 256 tons de cinza, onde o valor 0 (zero) representa a cor preta e o valor 255 representa a cor branca.

Figura 8 – Diferentes conversões de uma imagem colorida para tons de cinza.



Fonte: Elaborado pela autora.

2.3.3 Filtragem de imagem

A filtragem de uma imagem é utilizada para corrigir e realçar detalhes. Enquanto a correção é a remoção das características indesejáveis, a melhora/realce é a acentuação das características desejáveis (FILHO; NETO, 1999). As técnicas de filtragem podem ser divididas em algumas categorias, como: filtragem espacial ou no domínio da frequência; Filtragem passa alta, passa baixa, passa banda; filtros lineares ou não lineares; ou ainda, filtros de suavização, de contraste, adaptativos, globais, janelados ou locais (FILHO; NETO, 1999).

2.3.4 Segmentação da Imagem

A segmentação de imagens implica na divisão ou separação da imagem em regiões cujas *pixels* compartilham alguma propriedade (GONZALEZ; WOODS, 2007). A segmentação pode seguir duas estratégias genéricas:

- **Descontinuidade:** A partição da imagem é efetuada com base nas alterações bruscas de intensidade (ex., detecção de contornos);
- **Similaridade:** A partição é efetuada com base na similaridade entre *pixels*, seguindo um determinado critério (ex., binarização, crescimento de regiões, divisão e junção de regiões).

Dentre os métodos para se segmentar uma imagem, diversas estratégias podem ser encontradas: Crescimento de regiões, *superpixels*, limiarização, cortes em grafos, algoritmos Genéticos, redes Neurais, métodos Estocásticos, dentre outros (GONZALEZ; WOODS et al., 2002).

2.3.5 Análise Morfológica

A análise morfológica em processamento de imagens envolve etapas-chave para extrair informações relevantes das estruturas visuais. Começando com o pré-processamento da imagem, seguido pela aplicação de operadores morfológicos para manipular a forma das regiões de interesse. Posteriormente, a segmentação separa essas regiões do restante da imagem. Finalmente, ocorre a análise e interpretação dos resultados para extrair informações úteis. Este processo é interativo, exigindo ajustes para alcançar os resultados desejados.

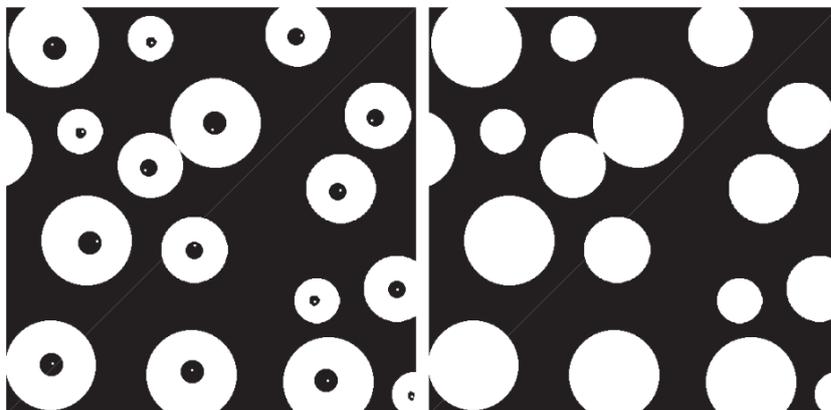
2.3.6 Operadores Morfológicos

O termo Morfologia Matemática (MM) surgiu no ano 1964 pelos pesquisadores G. Matheron e J.Serra. No campo da matemática a MM é usada para extração de componentes de uma imagem, como por exemplo bordas de uma região ou componentes conexas de um objeto (BANON; BARRERA, 1994).

Considere uma imagem binária I , como ilustrado na Figura 9. Os seus *pixels* assumem dois possíveis valores: 0, indicando região de fundo na cor preta; 1, indicando algum objeto (ou partes) de interesse, na cor branca. Dessa forma, há uma analogia direta entre conjuntos de *pixels* que definem o objeto, e o seu conjunto complementar, assim como na Teoria dos Conjuntos. Assim, as operações morfológicas são feitas sobre conjuntos, e as principais são: Erosão, Dilatação, Abertura e Fechamento, sendo as duas últimas combinações das duas primeiras.

Para efetuar as operações, utiliza-se de um elemento estruturante que será denotado como B , também formado por um conjunto binário, pertencente de um conjunto binário de

Figura 9 – Imagem ilustrando o preenchimento de buracos utilizando morfologia matemática.



Fonte: Figura extraída de Gonzalez e Woods (2000).

uma imagem. De forma mais clara um pixel com a cor preta não interage com a imagem, será inativo. Já um pixel com a cor branca interage com a imagem, consequentemente ativo e será elemento importante para a estrutura do objeto na imagem (MATTA, 1998).

A erosão de um conjunto (imagem binária) X pelo elemento estruturante B_x , centrado em cada pixel $x \in X$, é definida por:

$$X \ominus B_x = \{x \in I : B_x \subset X\}, \quad (2)$$

onde I é a imagem resultante.

A erosão fará com que o objeto da imagem fique menor. Além disso, faz desaparecer todos os conjuntos inferiores ao elemento estruturante, e aumenta os “furos” interiores aos conjuntos. O pixel x da imagem que corresponde ao ponto central do elemento estruturante será ativado se o elemento estruturante estiver inteiramente contido na imagem original, caso contrário, será marcado como irrelevante. Obtém-se assim, pixels ativos nas posições onde o elemento estruturante “se encaixa” na imagem.

A Dilatação ocorre de forma oposta a Erosão. Sua formula é dada por :

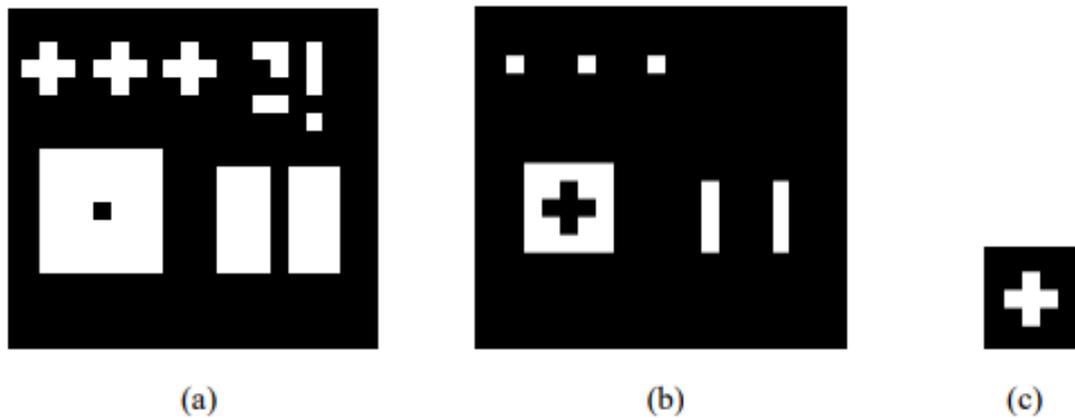
$$X \oplus B_x = \{x \in I : B_x \cap X \neq \emptyset\}, \quad (3)$$

onde I é a imagem resultante.

Da mesma forma como na erosão, o elemento estruturante desliza sobre a imagem. Se houver alguma interseção do elemento estruturante com a imagem, o pixel x da imagem correspondente ao ponto central do elemento estruturante será ativado, caso contrário será marcado como irrelevante.

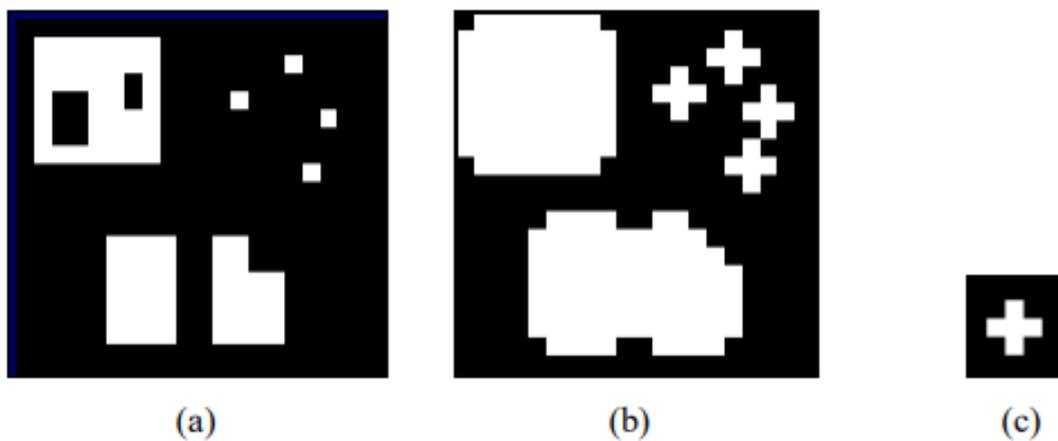
A função do operador morfológico Abertura é permitir a correção, ajuste do objeto

Figura 10 – (a) Imagem Original (b) Depois do processo de Erosão (c) Elemento estruturante em forma de cruz.



Fonte: Extraída de Matta (1998).

Figura 11 – (a) Imagem Original (b) Depois do processo de Dilatação (c) Elemento estruturante em forma de cruz.



Fonte: Extraída de Matta (1998).

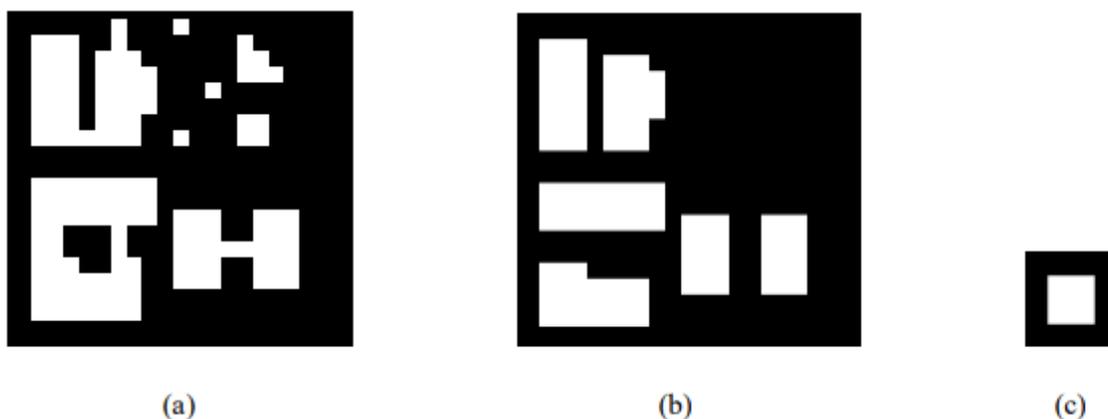
sem modificar estrutura do seu formato. Conseqüentemente os fragmentos que não fazem parte do objeto como pequenos fragmentos são eliminados da imagem com essa operação. Sua formula é dado por:

$$X \circ B_x = (X \ominus B_x) \oplus B_x, \quad (4)$$

Observa se pela formula que para o processo da abertura consiste na aplicação dos operadores erosão seguido por dilatação, ambos pelo elemento estruturante B . Essa operação de abertura faz surgir pequenos espaços entre os objetos vizinhos. Definindo melhor os seus formatos.

Assim como a dilatação é o oposto da erosão. A abertura é o inverso do fechamento.

Figura 12 – (a) Imagem Original (b) Depois do processo de Abertura (c) Elemento estruturante em forma de cruz.



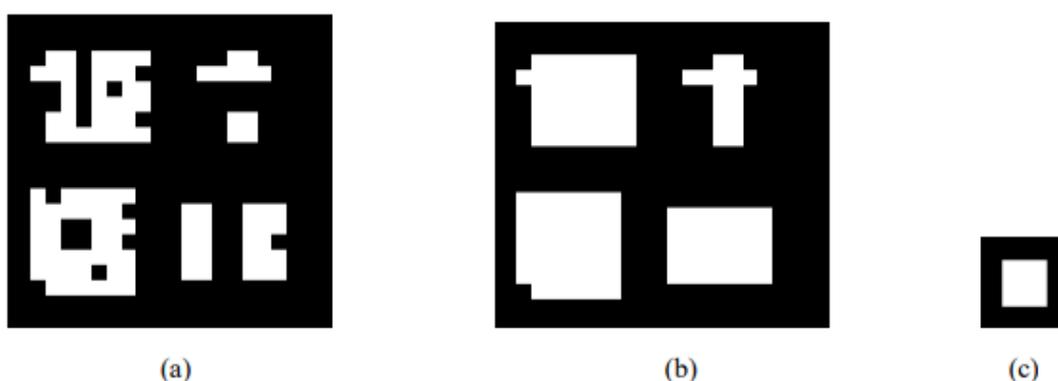
Fonte: Extraída de Matta (1998).

Sua formula é dada por:

$$X \bullet B_x = (X \oplus B_x) \ominus B_x, \quad (5)$$

Já que a abertura cria pequenos “buracos” e sendo que o fechamento é o seu inverso. Logo, a principal função do fechamento é fechar “buracos”. Por exemplo, corrigir pequenas fissuras presentes no objeto da imagem binária.

Figura 13 – (a) Imagem Original (b) Depois do processo de Fechamento (c) Elemento estruturante em forma de cruz.



Fonte: Extraída de Matta (1998).

2.3.7 Descritores

Uma vez que a imagem é segmentada, o próximo passo para a análise das características é buscar as informações individuais de cada objeto, que são obtidas com base no uso de descritores. Com estas medidas podemos comparar a informação dos objetos

com medidas previamente conhecidas para enfim classificar o objeto em uma das várias categorias. Existem vários tipos de descritores capazes de descrever as características de um objeto, como descritores de forma, textura ou cor (SANTOS et al., 2009).

Formas bidimensionais podem ser descritas de duas maneiras diferentes. O primeiro método visa usar o contorno do objeto e suas características (como o tamanho do contorno, curvatura, assinatura, descritores de Fourier), e o segundo visa descrever a região ocupada pelo objeto no plano da imagem. O primeiro método está diretamente ligado à detecção de linhas e bordas e os esquemas descritores resultantes são chamados de representações externas. Estes gozam de certa popularidade, pois produzem uma compacta representação da forma. O segundo método é conectado às técnicas de segmentação de região. Os esquemas de segmentação conduzem diretamente para esquemas de representação do objeto. Descritores de região (como área, momentos e esqueletonização) têm sido usados extensivamente em aplicações de reconhecimento de objetos (SANTOS et al., 2009). Estes podem ser utilizados como representações simplificadas de uma forma bidimensional em uma imagem.

2.4 Projecção multidimensional

Dados multidimensionais são encontrados em diversos contextos. Considere por exemplo uma imagem em tons de cinza, com tamanho 3×3 *pixels*. Podemos aplicar uma transformação \mathbf{T} nessa imagem de forma que cada coluna sua seja concatenada imediatamente ao final da coluna anterior, formando um vetor coluna de tamanho 9, como ilustrado na Equação 6.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \xrightarrow{\mathbf{T}} (a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33})^{\top} = \mathbf{v} \quad (6)$$

Dessa forma, o vetor \mathbf{v} encontra-se em um espaço de dimensão 9, ou 9-dimensional. Na álgebra linear a transformação \mathbf{T} com essas características é conhecida como *vetorização*. Observe que cada entrada a_{ij} irá representar uma coordenada do ponto no espaço 9-dimensional. De forma mais genérica, tem-se que:

Definição 2.4.1 (Dado multidimensional). Seja \mathbf{v} um vetor de m coordenadas $\mathbf{v} = (v_1, \dots, v_m)$. Esse vetor é dito um dado m -dimensional, ou seja, é um ponto (vetor) em um espaço de dimensão m .

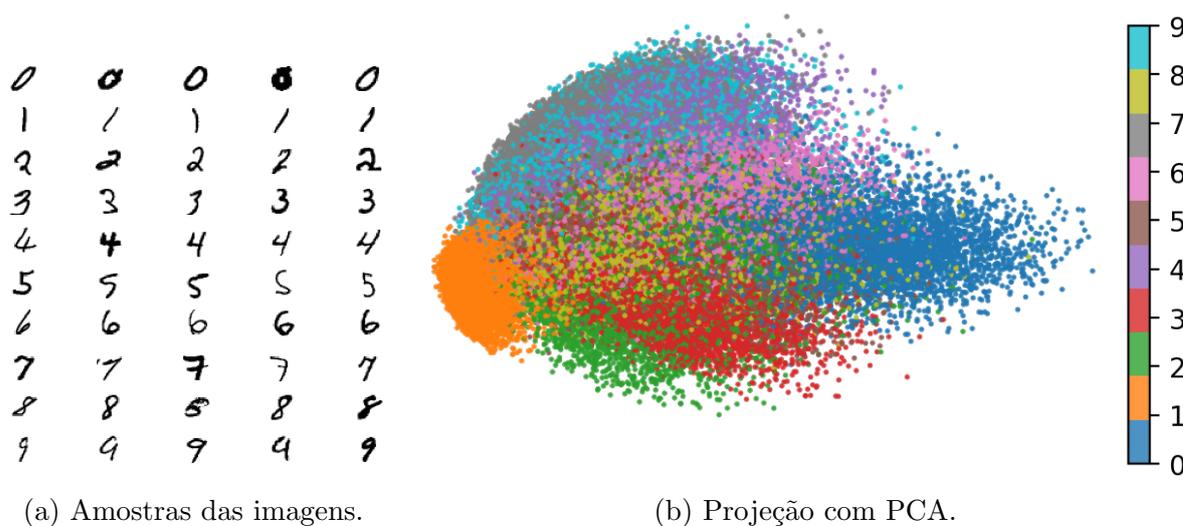
Essa forma de representar a imagem como um dado m -dimensional não é única, nem a mais adequada. Cada representação diferente, ou combinação de diferentes representações, irá definir o chamado *vetor de características* que representa o objeto. O espaço onde ele está imerso é dito *espaço de características*.

2.4.1 Técnicas de projeção

Uma projeção multidimensional é uma técnica usada para transformar um conjunto de dados com múltiplas dimensões em um espaço onde podemos visualizá-los com apenas duas ou três dimensões, ou seja, em um espaço visual mais simples. Essa transformação é realizada com o objetivo de manter as relações de similaridade que existiam entre os pontos no espaço original, como distâncias e vizinhanças entre os objetos. Para visualizar essa projeção, é comum representá-la por meio de gráficos de dispersão (*scatterplots*), nos quais cada ponto representa um dado do espaço de características projetado no espaço visual, como exemplificado na Figura 13b.

Na Figura 14 é ilustrado o *dataset* MNIST (LECUN et al., 1998), que é composto por 70 mil imagens distribuídas entre 10 classes, de 0 a 9, possuindo cada imagem tamanho de 28x28 *pixels* em tons de cinza, de dígitos manuscritos de documentos do Departamento do Censo dos Estados Unidos (*NIST*), e de estudantes do ensino médio estadunidense.

Figura 14 – Exemplo de projeção multidimensional utilizando o *dataset* de dígitos MNIST (LECUN et al., 1998).



(a) Amostras das imagens.

(b) Projeção com PCA.

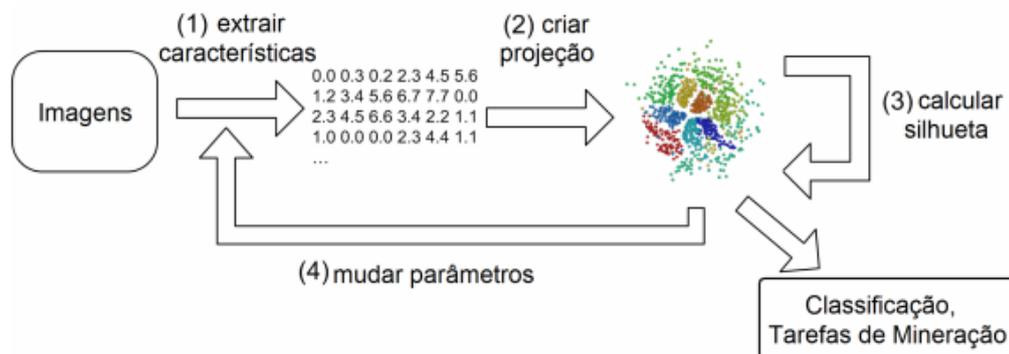
Fonte: Elaborado pela autora.

Na Figura 14a é possível ver alguns exemplos de cada classe, onde nota-se a variação da caligrafia para o mesmo dígito. Já na Figura 14b é possível visualizar a projeção de todo o *dataset*. Com a representação e técnica de projeção utilizadas, os pontos que representam o dígito 1, por exemplo, mantêm-se todos próximos entre si e distantes dos pontos que representam imagens do dígito 0. Cabe observar que dos dígitos da classe 5 e 6 possuem grandes semelhanças entre si, o que reflete nas suas posições na projeção, sendo bastante próximas entre si. O mesmo ocorre entre as classes 4, 7 e 9.

Com os dados projetados no espaço visual, diversos tipos de análise podem ser feitas: desde busca por padrões visuais, ou até mesmo a aplicação de técnicas de mineração

de dados ou aprendizado de máquina. A Figura 15 descreve um *pipeline* para analisar dados utilizando técnicas de projeção multidimensional.

Figura 15 – Ilustração da estrutura da projeção multidimensional.



Fonte: Extraída de Machado et al. (2012).

Para exploração dos dados é permitido que os usuários interajam com a projeção de forma que arraste diretamente os pontos projetados (cada ponto representa um dado do *dataset* selecionado) ou ajuste parâmetros associados à projeção (ex.: os pesos de relevância de cada característica) por meio de controles deslizantes (*sliders*). A primeira forma é dita interação a nível de observação (*Observation-level interaction - OLI*). Já a segunda é dita interação paramétrica (*Parametric Interaction - PI*).

Tanto para a etapa de projeção quanto para a análise dos dados projetados a definição da métrica utilizada impacta diretamente nos resultados obtidos.

A noção de métrica está associada à distância, função de similaridade ou dissimilaridade pontos de dados, e desempenha um papel importante na visualização de dados, aprendizagem de máquina, reconhecimento de padrões e técnicas de mineração de dados. Existem muitos algoritmos de agrupamento que muitas vezes são classificados de acordo com sua relevância para uma determinada consulta com base na pontuação de similaridade. Com isso, o desempenho desses métodos depende da qualidade da métrica: esperamos que ela identifique como é semelhante os pares de casos que são de fato semanticamente próximo (respectivamente diferentes). Existem métricas de uso geral (por exemplo, a distância euclidiana e a similaridade do cosseno de vetores de características), mas muitas vezes eles não conseguem simplificar dos dados de interesse. Melhores resultados são esperados quando a métrica é projetada especificamente para a tarefa (BELLET; HABRARD; SEBBAN, 2013).

2.4.2 Métrica de consistência

O processo de projeção tenta preservar as relações métricas originais tanto quanto possível, isto é: distância entre os pontos, relação de vizinhança. Porém, é um processo

com perdas, ou seja, acaba por descartar informações do espaço original (CEDRIM et al., 2016). Duas coisas podem impactar consideravelmente o resultado obtido por uma projeção: a representação utilizada para o dado multidimensional; a técnica de projeção utilizada. Uma forma de quantificar essa modificação é através de *métricas de consistência*.

A *métrica da silhueta* tem como objetivo classificar quão bem um objeto é pertencente a um grupo. Refere-se a um método de interpretação e validação de consistência dentro de grupos de dados. O resultado do cálculo desta, pode variar entre $[-1,1]$, onde quanto mais o valor for próximo a -1 indica que o agrupamento está incorreto, ou seja, o objeto está mal combinado com os objetos adjacentes. Partindo do mesmo princípio, quanto mais o valor se aproximar de 1 indica que o agrupamento do objeto está bem combinado com seu próprio grupo. A métrica está definida na Equação 7:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (7)$$

com

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (8)$$

e

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j), \quad (9)$$

onde C_k é o agrupamento (*cluster*) que contem os pontos da classe k , e $d(i, j)$ é a distância entre os pontos i e j .

Observe que a silhueta é calculada a partir de dois termos: O primeiro deles ($a(i)$) indica a distância média dos objetos pertencente ao mesmo grupo e i -ésimo ponto; o segundo termo ($b(i)$), calcula a distância média para cada um dos outros grupos, e escolhe a menor entre elas. A intuição associada é que quanto mais compactos os agrupamentos e maior a distância entre os *clusters*, maior o valor da silhueta. Além disso, se um ponto está distante dos demais elementos do seu *cluster* e perto de outros *clusters*, o seu valor de silhueta é reduzido. Esse tipo de métrica possibilita estudar como é a estrutura geral de um agrupamento. Na Figura 16 são exemplificados dois experimentos para ilustrar esse comportamento.

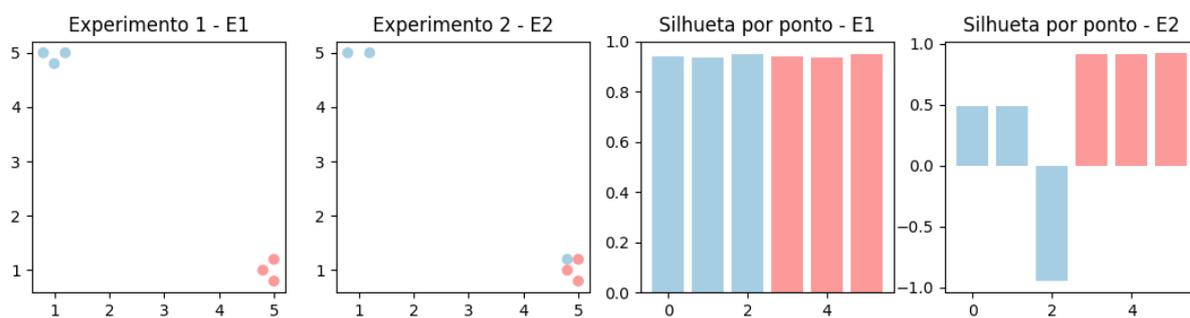
No experimento 1 a análise visual mostra que os dois agrupamentos são bem compactos e estão distantes entre si. Matematicamente isso pode ser observado pois a silhueta para cada um dos pontos obtém valores próximos do máximo, como ilustrado no gráfico de barras.

Já no experimento 2 é notório uma diferença, fazendo uma comparação com o experimento 1. Verificando os dois aspectos novamente, tem-se que no visual é perceptível que um dos pontos em azul está mais próximo do grupo distinto do que o seu próprio grupo. Sendo assim, o seu valor de silhueta deve ser bem próximo do mínimo, o que é

verificado pois obtém silhueta próxima de -1. Note que o valor da silhueta dos outros dois pontos deverá ser menor do que os valores obtidos no experimento 1, pois o *cluster* deixou de ser compacto. No gráfico do agrupamento azul os valores dos dois pontos em azul estão próximos do valor de 0,5.

Por último, o grupo da cor vermelha mostra valor próximo de 1 no gráfico, pois o agrupamento continua compacto e todos os pontos pertencentes ao mesmo grupo continuam distantes à média dos elementos do outro grupo azul.

Figura 16 – Experimento de silhueta.



Fonte: Elaborado pela autora.

3 TRABALHOS RELACIONADOS

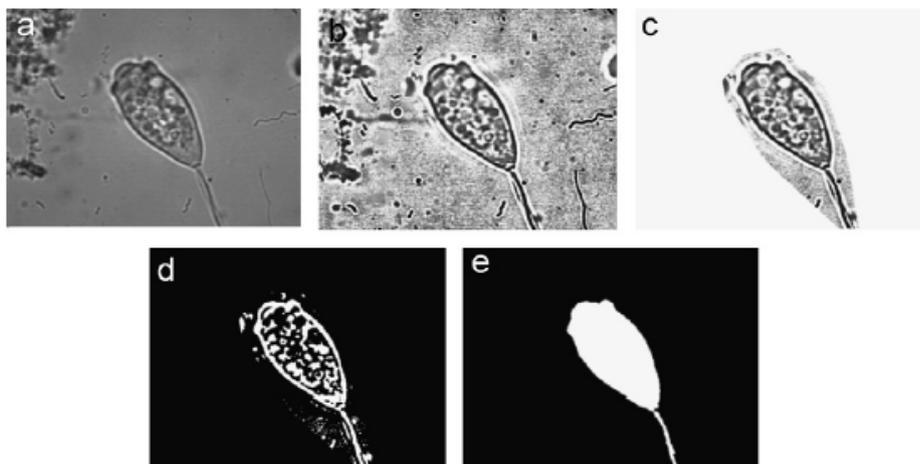
Com o objetivo de entender e analisar estruturas pelas quais foram usadas ao longo desse projeto. Essa seção será destinada aos trabalhos relacionados.

Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system

Neste trabalho, Ginoris et al. (2007) apresentam uma pesquisa experimental sobre o estudo de protozoários que habitam em uma estação de tratamento biológico de águas residuais, popularmente chamado de “esgoto”. Normalmente, a identificação e quantificação de cada protozoário é conseguido por inspeção microscópica e a contagem manual, exigindo tempo e perícia técnica elevada. No entanto, os avanços tecnológicos e a diminuição de custos de cálculo deu a oportunidade para novas técnicas, tais como análise de imagem para ser utilizada na rotina de quantificação de microrganismos numa forma subjetiva automatizada e não ultrapassar alguns dos inconvenientes das técnicas manuais.

Na análise de imagem, a precisão inerente e a precisão de técnicas de microscopia e a velocidade da computação são combinados reduzindo assim o fator de erro humano. Nestes estudos, a análise de imagem foi utilizada para a identificação de protozoários normalmente presente em estações de tratamento de lamas ativas em Nancy (França). O programa de processamento e análise de imagem, consiste em dois módulos principais, ou seja, pré-tratamento e segmentação (GINORIS et al., 2007), com resultados ilustrados na Figura 17.

Figura 17 – Principais passos: Imagem original (a), imagem pré-tratado (b), a região de interesse (c), a imagem binária após a segmentação (d), e imagem final (e).



Fonte: Ginoris et al. (2007)

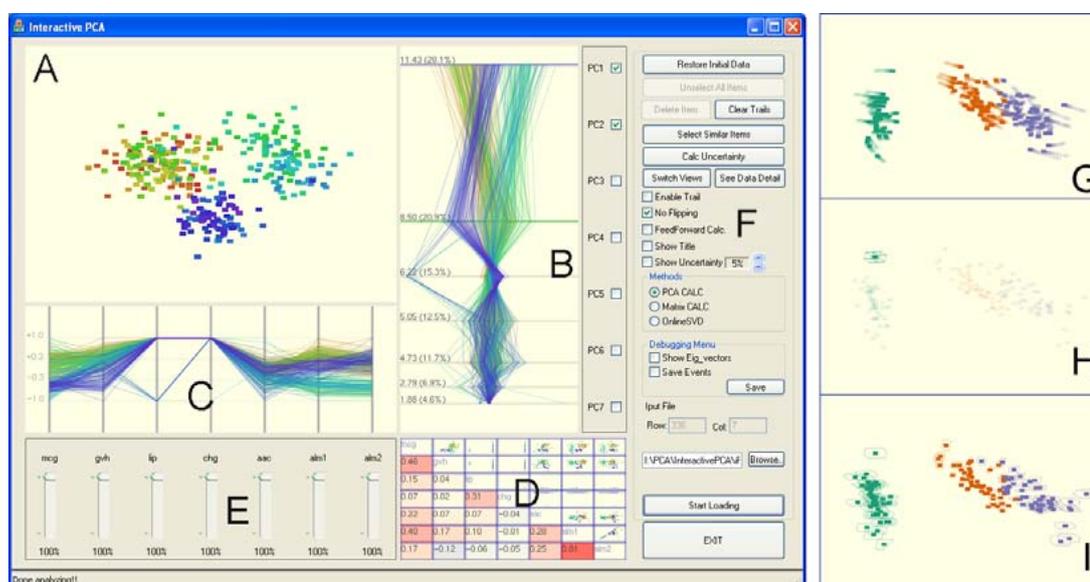
As imagens digitais foram armazenadas em formato TIFF e, posteriormente processadas aplicando a técnica de análise de imagem. A primeira etapa do programa ProtoRec consiste em melhorar a imagem em escala de cinzento inicial (256 níveis). A imagem é inicialmente submetida a equalização de histograma locais, a fim de melhorar o contraste de cada região na imagem, filtragem mediana para executar redução de ruído. Antes da etapa de segmentação, uma região poligonal de interesse (ROI) é definido pelo usuário manualmente em torno do objeto selecionado (recorte).

Aplicado então os filtros de binarização para que a imagem fique em tons de preto e branco, filtro de inversão para diferenciação do objeto e a cor de fundo da imagem. Na fase subsequente, o material de detritos (pequenos artefatos e outros materiais que podem interferir com a análise) é eliminada por uma série de operações morfológicas aplicado às imagens binárias incluindo fechamento morfológica, preenchendo e operações de abertura.

iPCA: An Interactive System for PCA-based Visual Analytics

Com a intenção de ajudar o usuário a entender melhor e interagir com o PCA, Jeong et al. (2009) desenvolveram um sistema que visualiza os resultados da análise de componentes principais utilizando múltiplos pontos de vista coordenados e um rico conjunto de interações do usuário. A ideia principal é apoiar a análise de conjuntos de dados multivariados através de uma extensa interação com a saída do PCA. PCA é uma técnica matemática amplamente utilizado em muitos campos para o fator e análise de tendências, a redução da dimensão. No entanto, é muitas vezes considerado como uma operação de “caixa preta”, cujos resultados são difíceis de interpretar e, por vezes contraintuitiva para o usuário mais leigo.

Figura 18 – Interface da proposta do iPCA.



Fonte: Extraída de Jeong et al. (2009).

As principais partes do sistema estão ilustradas na Figura 18. Na parte (A) é possível ver a projeção inicial dada pelo PCA. Em (E) é possível visualizar *sliders* para ajuste da contribuição de cada dimensão (peso). Já em (G) é possível ver como os pontos da projeção são afetados após a mudança do peso de uma das dimensões.

Para demonstrar a utilidade do sistema, foi realizado um estudo comparativo com um sistema comercial conhecida, do SAS / INSIGHT Interactive Data Exploration. Os participantes do estudo foi proposto a resolução de uma série de tarefas de análise de alto nível com cada interface e avaliado os sistemas na facilidade de aprendizagem e utilidade. Com base na precisão dos participantes, velocidade e *feedback* qualitativo, foi observado que o sistema IPCA sistema ajudou os usuários a compreender melhor as relações entre os dados, o que permite aos participantes de forma mais precisa analisar os dados (JEONG et al., 2009).

A resposta do utilizador sugeriu que a interatividade e transparência do sistema IPCA para com o usuário foram os principais pontos fortes da abordagem, que permitiu que os participantes a analisassem com mais precisão os dados. Os resultados positivos do sistema IPCA se deu por conta da utilização da técnica *Visual Analytics* que pode ser vista como uma abordagem integral que combina visualização, fatores humanos e análise de dados. Essa técnica permite também que os tomadores de decisão combinem sua flexibilidade da técnica e conhecimento em segundo plano com as enormes capacidades de armazenamento e processamento dos computadores para obter uma visão de problemas complexos. Usando interfaces visuais avançadas, os usuários podem interagir diretamente com os recursos de análise de dados do sistema em questão.

Note que essa abordagem possibilita apenas interação a nível de observação (OLI).

4 METODOLOGIA

Neste capítulo, descreveremos os materiais e métodos empregados para atingir os objetivos previamente discutidos. Em síntese:

- Isolar as regiões de interesse (ROI) que contêm células de amebas na imagem;
- Construir um vetor descritores de características, usando descritores de imagem, para mapear essas regiões em um espaço de características;
- Aplicar a projeção multidimensional para analisar o espaço de características relacionado a cada conjunto de descritores.

Python foi a linguagem de programação escolhida para desenvolvimento da ferramenta. Por que tem uma biblioteca padrão grande que possui vários módulos para diversas necessidades que possam surgir durante o código. Além disso, o seu suporte é imenso. Por ser uma linguagem interpretada e de alto nível, roda em qualquer sistema que possua o seu interpretador. É inteligível como o Inglês e sempre evita deixar coisas implícitas. Além de multiplataforma e multi paradigma, com o Python o programador será capaz de desenvolver para qualquer tipo de ambiente, seja desktop, web ou móvel.

4.1 Aquisição das imagens

As etapas aqui descritas foram realizadas em colaboração com Dr. Douglas Cedrim Oliveira, Dra. Natália Karla Bellini e Dra. Maria Reyes-Batlle, laboratório de amebas de vida livre, Universidade de La Laguna, Espanha.

As culturas de amebas de vida livre dos gêneros *Naegleria* e *Vermamoeba* foram mantidas em placas de ágar não nutritivo, temperatura ambiente, conforme protocolos de cultura celular (BELLINI et al., 2020). Por sua vez as culturas de *Acanthamoeba* foram mantidas em meio líquido (ALFIERI et al., 2000). Os trofozoítos (1×10^6 células) foram centrifugados diretamente de cultura líquida, ou re-suspendidos da superfície do ágar com solução salina. Na sequência foram induzidos à diferenciação para estágio de cisto utilizando meios líquidos específicos para cada gênero de ameba (SOHN et al., 2017). Lâminas de vidro foram preparadas com as culturas a fresco. Aproximadamente 1×10^4 cistos foram depositados na lâmina, sobrepostos com lamínula, e diretamente visualizados em microscópio Leica DM1000 LED, magnificação 200x, 400x e 1000x. As imagens foram adquiridas em câmera Leica ICC50W, acoplada ao microscópio, e processadas em software LAS EZ versão 3.4, cujos parâmetros de aquisição da imagem (exemplo: intensidade, exposição e saturação da luz) foram ajustados manualmente e utilizados como padrão para as demais aquisições. Os arquivos foram exportados em formato TIFF e utilizados para análise morfológica computacional.

4.2 Requisitos do sistema

São requisitos diretamente ligados a funcionalidade do software. O mesmo está relacionado à atividade que o sistema realiza por meio do usuário (ÁVILA; SPÍNOLA, 2007). No contexto desse trabalho foram definidos os requisitos funcionais e não funcionais descritos como segue.

4.2.1 Requisitos funcionais

- **RF1** Inserção de Dados (*Dataset*): O sistema deve permitir que o usuário insira dados organizados em pastas, onde cada pasta representa uma classe específica de AVL. Esses dados (imagens já segmentadas anteriormente) serão processados para diferenciar as espécies;
- **RF2** Análise de Múltiplas Classes (Classes): O sistema deve oferecer a capacidade de analisar até 12 classes diferentes simultaneamente para usuários regulares;
- **RF3** Suporte ao Daltonismo (Daltonismo): O sistema deve permitir a análise simultânea de até 4 classes distintas para usuários com deficiência visual, como daltonismo;
- **RF4** Representação de Espécies (Descritores): O sistema deve possibilitar a representação de cada espécie por meio de um vetor de características multidimensional que contém diversos descritores;
- **RF5** Projeção Multidimensional (Projeção): O sistema deve permitir a análise visual das espécies por meio de projeção multidimensional dos dados;
- **RF6** Visualização Gráfica (Visualização): O sistema deve possibilitar a visualização dos dados projetados através de um gráfico de dispersão (*scatterplot*), onde cada ponto representa a imagem de uma espécie;
- **RF7** Ajuste de Pesos (*Sliders*): O sistema deve possibilitar a definição e ajuste de diferentes pesos para cada descritor por meio dos *sliders*, com o objetivo de explorar a influência na projeção multidimensional para diferenciação das espécies.

4.2.2 Requisitos não-funcionais

São requisitos que expressam condições que o software deve atender ou qualidades específicas que o software deve ter. Eles não se relacionam diretamente com as funções do sistema (ÁVILA; SPÍNOLA, 2007).

- **RNF1** (Disponibilidade): É um sistema desenvolvido para ser apenas *desktop*;
- **RNF2** (Interatividade): A técnica de projeção utilizada no sistema deve possuir custo computacional baixo para possibilitar interação em tempo real;
- **RNF3** (Criação das classes): O número máximo de classes possíveis para análise em um dado momento deve ser limitado para garantir interatividade.

4.3 Pré-Processamento

O pré-processamento de uma imagem é visto como a etapa anterior ao seu processamento efetivo, como o próprio nome sugere. Esta etapa é importante pois prepara a imagem para que a mesma possa ter um processo satisfatório. Este trabalho selecionou uma série de procedimentos de pré-processamento e segmentação de imagens com base em técnicas como equalização de histograma, filtragem mediana, binarização, inversão e operadores morfológicos. O pré-processamento é fundamental para preparar a imagem para análise posterior, garantindo que esta esteja em condições adequadas para um processamento eficaz. A equalização de histograma, por exemplo, redistribui as intensidades de forma a obter um histograma mais uniforme, enquanto o filtro de mediana preserva os detalhes finos e bordas da imagem.

No contexto deste trabalho, o ImageJ versão 1.52p foi utilizado em várias etapas do pré-processamento, como filtragem e segmentação de imagens, bem como já possui descritores implementados, o que contribui significativamente para a análise e caracterização das imagens. O emprego dessa ferramenta oferece uma abordagem robusta para as necessidades específicas do projeto em questão.

O ImageJ é um processador de imagens baseado em Java, oferecendo recursos abrangentes para exibição, edição, análise, processamento, armazenamento e impressão de imagens em diversos formatos. É capaz de calcular a área e o valor do pixel selecionado pelo usuário, realizar medições de distâncias e ângulos, gerar gráficos estatísticos e muito mais. O programa oferece suporte a diversas funções de processamento, incluindo operações aritméticas e lógicas entre imagens, ajuste de contraste, análise de Fourier, reconhecimento de margens e aplicação de filtros medianos.

4.3.1 Segmentação

A segmentação de imagens é uma técnica essencial para dividir uma imagem em regiões ou objetos significativos. Isso é fundamental para análises subsequentes, como reconhecimento de padrões e detecção de objetos. Existem várias abordagens. A escolha do método certo depende das características da imagem e do problema em questão.

Segmentação automatizada

Essa etapa de segmentação é feita baseada no artigo da Ginoris et al. (2007), onde são aplicados a equalização de histograma e o filtro da mediana antes da segmentação em si.

A equalização do histograma procura redistribuir as intensidades do histograma original, de modo a obter um histograma uniforme, com praticamente todas as intensidades com o mesmo percentual. Pode ser aplicada em partes da imagem, por exemplo, em

uma janela $n \times m$. A aplicação local serve principalmente para realçar detalhes sutis de pequenas porções da imagem.

Já o filtro de mediana substitui o pixel central da máscara pela mediana de seus vizinhos, visando contornar a deficiência do filtro de média, que é a perda dos detalhes finos e bordas da imagem.

Após essas etapas, são aplicados: a inversão, a binarização utilizando o método de Otsu e por fim são usados para remover imperfeições os operadores morfológicos de fechamento, preenchimento (*hole filling*) e abertura. Esses procedimentos visam alcançar a segmentação automática da imagem, um passo crucial para a identificação precisa dos elementos de interesse.

Segmentação manual

Com o objetivo de gerar um resultado mais aproximado da célula real no sentido morfológico, ou seja, uma melhor segmentação, o sistema oferece uma outra alternativa para que o usuário que funcionará da seguinte forma: Ao obter a imagem original o usuário faz o recorte manual da região de interesse da célula (recorte ao redor do formato do “corpo” da célula), após esse processo esse recorte gera uma nova imagem com o fundo preto onde é preservado o recorte da região de interesse.

4.4 Extração de características das amebas

A fase de extração de características envolve a obtenção de medidas relevantes que podem ser utilizadas na classificação de padrões. Essa etapa também funciona como um meio de compressão de dados, pois reduz a quantidade de informações, eliminando redundâncias e ampliando a distinção entre classes. Durante essa etapa, foram obtidos descritores em cada célula, essenciais para gerar pontos no espaço de projeção.

4.4.1 Descritores utilizados

Ao longo desse trabalho, usamos descritores previamente implementados e validados no ImageJ versão 1.52p. Especificamente:

- **Área:** Pode ser expressa com numero de *pixels* que compreende o objeto ou região de interesse;
- **Perímetro:** O perímetro representa o comprimento da borda de um objeto. Um aproximação para este descritor pode ser obtida pela contagem dos pixels ao longo da borda do objeto na imagem. O comprimento do limite externo da seleção do objeto;
- **Circularidade:** Mede o quão próximo um objeto é de um círculo perfeito. É calculada como a razão entre a área do objeto e o quadrado do seu perímetro. Quanto mais

próximo de 1, mais próximo de um círculo. A circularidade de uma região é dada pela fórmula:

$$C = \frac{P^2}{A} \quad (10)$$

em que P e A são perímetro e a área da região, medidos em unidades de *pixels*. De acordo com o descritor, um círculo possui o menor valor de circularidade, igual a 4π ;

- **Feret:** A maior distância entre dois pontos ao longo do limite de seleção. Também conhecido como comprimento da pinça. A macro Diâmetro do Feret desenha o Diâmetro do Feret da seleção atual na imagem;
- **MinFerret:** A menor distância entre dois pontos ao longo da borda do objeto. Indica a menor largura do objeto;
- **Solidity:** Mede a densidade de um objeto. Uma medida de solidez pode ser obtida pela relação entre a área de um objeto e a área do fecho convexo do objeto.

$$\text{Solidez} = \frac{\text{Área do objeto}}{\text{Área do fecho convexo}} \quad (11)$$

O valor da solidez é igual a 1 para um objeto sólido e menor que 1 para objetos apresentando borda irregular ou contendo buracos.

Um objeto mais sólido terá uma solidez maior, indicando menos "buracos" ou concavidades em sua forma. Uma solidez de 1 indica que o objeto preenche completamente o seu envelope convexo, o que sugere uma forma mais compacta e menos irregular. Valores menores que 1 indicam concavidades ou buracos na forma do objeto.

- **Round (Roundness):** A roundness, ou redondeza, é uma medida da similaridade da forma de um objeto com um círculo perfeito. Pode ser calculada de várias maneiras, mas uma fórmula comum é:

$$\text{Roundness} = \frac{4 \times \text{Área do Objeto}}{\pi \times \text{Maior Diâmetro}^2} \quad (12)$$

Interpretação prática: Quanto mais próximo o valor da roundness estiver de 1, mais a forma do objeto se assemelha a um círculo. Isso é útil para identificar objetos aproximadamente circulares em uma imagem. Se a roundness for menor que 1, isso indica uma forma mais alongada ou irregular em comparação com um círculo perfeito. Similar à circularidade, é uma medida da forma esférica de um objeto. Quanto mais próximo de 1, mais próximo de uma forma esférica;

- **Skew:** O skew, ou inclinação, refere-se à assimetria de um objeto em relação ao seu eixo principal. Em outras palavras, mede o quão deslocado ou inclinado um objeto é. O skew é frequentemente calculado em relação aos momentos de inércia do objeto. Interpretação prática: Se o skew for igual a zero, o objeto é simétrico em relação ao seu eixo principal. Valores diferentes de zero indicam deslocamento ou inclinação. Esse descritor é útil para distinguir entre objetos simétricos e assimétricos.

Para acessar os descritores foram necessárias algumas instalações, listadas a seguir.

Conda: O Conda é um sistema de gerenciamento de pacotes de código aberto e um sistema de gerenciamento de ambiente compatível com Windows, macOS e Linux. Ele instala, executa e atualiza pacotes e suas dependências de forma eficiente. Além disso, permite a criação, salvamento, carregamento e alternância fácil entre ambientes no computador local. Embora tenha sido originalmente desenvolvido para programas em Python, o Conda pode empacotar e distribuir software para qualquer linguagem.

Pyimagej: A biblioteca PyImageJ oferece um conjunto de funções de wrapper para facilitar a integração entre o ImageJ e o Python. Funciona como um ponto de entrada de alto nível para invocar as APIs do servidor ImageJ.

Após a conclusão dessas instalações, procedeu-se à importação da classe `imagej` por meio do módulo `pyjnius`, que possibilita o acesso a classes Java (do ImageJ) como se fossem classes Python. Em seguida, utilizou-se o método `init` para executar os arquivos do ImageJ localmente. Após essa etapa, ocorreu o acesso à classe do ImageJ chamada 'ParticleAnalyzer', a qual oferece comandos relacionados a medições estatísticas em dados de imagem, plotagem de perfil, histograma e plugins relacionados à análise de imagem. Posteriormente, os descritores utilizados pelo ImageJ foram recuperados por meio dessa ação.

4.4.2 Determinando a relevância de cada descritor

Para determinar a relevância de cada descritor para diferenciação de cada espécie. Foi considerado o seguinte cálculo:

$$P(\mathbf{v} \cdot \mathbf{w}) = [v_1 \cdot w_1, v_2 \cdot w_2, \dots, v_n \cdot w_n],$$

onde \mathbf{v} é o vetor de características de cada célula $v = [v_1, v_2 \dots v_n]$, e \mathbf{w} é o peso do slider $w = [w_1, w_2, \dots, w_n]$.

O cálculo será realizado para cada célula individualmente. Em essência, a multiplicação dos valores do vetor de características pelos pesos do slider contribui para uma avaliação ponderada da importância de cada descritor no contexto da diferenciação das espécies.

4.5 Projeção multidimensional

A projeção multidimensional é uma técnica que reduz a complexidade de conjuntos de dados, mapeando-os de espaços de alta dimensionalidade para espaços de menor dimensionalidade. Isso facilita a visualização e compreensão dos dados, ajudando a identificar padrões e relações entre os pontos.

4.5.1 PCA

Neste trabalho foi utilizado para a projeção multidimensional a Análise de Componentes Principais que é uma técnica que visa descobrir um conjunto de novas variáveis que seja substancialmente menor do que o conjunto original, enquanto retém a maior parte das informações contidas nos dados. Essencialmente, o objetivo principal é reduzir a dimensionalidade de um conjunto de dados, minimizando ao máximo a perda de informação (ARAÚJO; COELHO, 2009).

Com o objetivo que os dados do dataset estejam na mesma ordem de grandeza foi utilizado a normalização de dados. Essa técnica Z-Score irá inserir os valores dos dados dentro de um intervalo de 0 e 1, caso tenha resultado negativo -1 e 1. Dado a fórmula do Z-Score para normalização dos dados:

$$z = \frac{x - \mu}{\sigma} \quad (13)$$

onde x valor bruto, μ é a média e σ é o desvio padrão.

4.6 Interface do sistema

Foi desenvolvido um sistema desktop para auxiliar o profissional especialista na área de biologia a analisar espécies distintas de amebas. A ferramenta funcionará da seguinte forma:

- Dada uma imagem contendo as amebas coletadas via microscópio ótico de campo claro. A mesma será dividida em imagens menores separando cada ameba para o processamento;
- O processamento vai ser obtido através de etapas de segmentação de imagem;
- Após a segmentação, será montado um vetor de características multidimensional descrevendo cada ameba de acordo com alguns descritores, como: Área, Perímetro, Circularidade, Feret, MinFeret, Round e Solidity;
- De forma visual será utilizado uma projeção multidimensional para avaliar o espaço de características associado aos descritores utilizados. Cada célula será projetada em formato de ponto na projeção;
- Caso o usuário perceba que uma determinada espécie A está mais próxima na projeção visual do grupo B, ele terá a possibilidade de alterar os pesos do descritor até que a espécie fique ajustada de forma próxima ao grupo correto.

4.6.1 Qt Designer

No desenvolvimento da interface, utilizado a ferramenta Qt Designer, que oferece a capacidade de projetar e criar interfaces gráficas de usuário de maneira eficiente. Essa ferramenta facilita a vida do seu usuário ao permitir a criação de widgets, diálogos ou

janelas principais por meio de uma interface intuitiva de arrastar e soltar, utilizando formulários na tela. De maneira prática, o Qt Designer viabiliza a visualização dos designs criados para garantir que funcionem conforme desejado, assemelhando-se a um protótipo, antes mesmo de começar a escrever qualquer código.

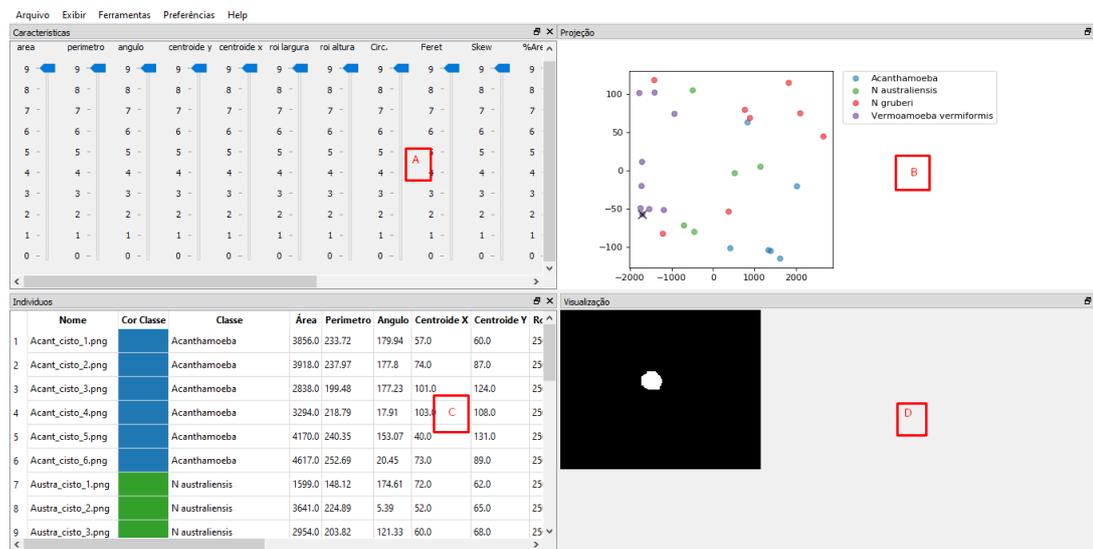
O Qt Designer utiliza arquivos XML (.ui) para armazenar os designs, mas não gera código por si só. O Qt inclui o utilitário uic, que é responsável por gerar o código C++ que cria a interface do usuário. Adicionalmente, é disponibilizada a classe QUiLoader, que possibilita que um aplicativo carregue um arquivo .ui e crie dinamicamente a interface do usuário correspondente.

No contexto do desenvolvimento em Python no Qt Designer, foi optado por utilizar o PyQt5, que representa um abrangente conjunto de ligações Python para o Qt v5. Com mais de 35 módulos de extensão, o PyQt5 permite que o Python seja empregado como uma linguagem alternativa no desenvolvimento de aplicativos em C++ em todas as plataformas suportadas.

É importante mencionar que o PyQt5 não substitui a classe QUiLoader, mas inclui o módulo uic Python. Assim, este módulo possibilita que a classe QUiLoader carregue arquivos .ui para criar uma interface de usuário dinamicamente. Similar ao utilitário uic, o PyQt5 pode gerar o código Python responsável por criar a interface do usuário. O utilitário pyuic5 do PyQt5 funciona como uma interface de linha de comando para o módulo uic.

4.6.2 Estrutura da interface

Figura 19 – Interface da aplicação desenvolvida, mostrando: a projeção inicial; os descritores calculados para cada imagem; os pesos associados a cada descritor e a imagem em si.



Fonte: Elaborado pela autora.

A interface é composta por quatro blocos, onde a junção destes tem o objetivo de realizar de forma visual a distinção das espécies de ameba. Como mostrado na Figura 19 é dividida em quadro blocos, são eles:

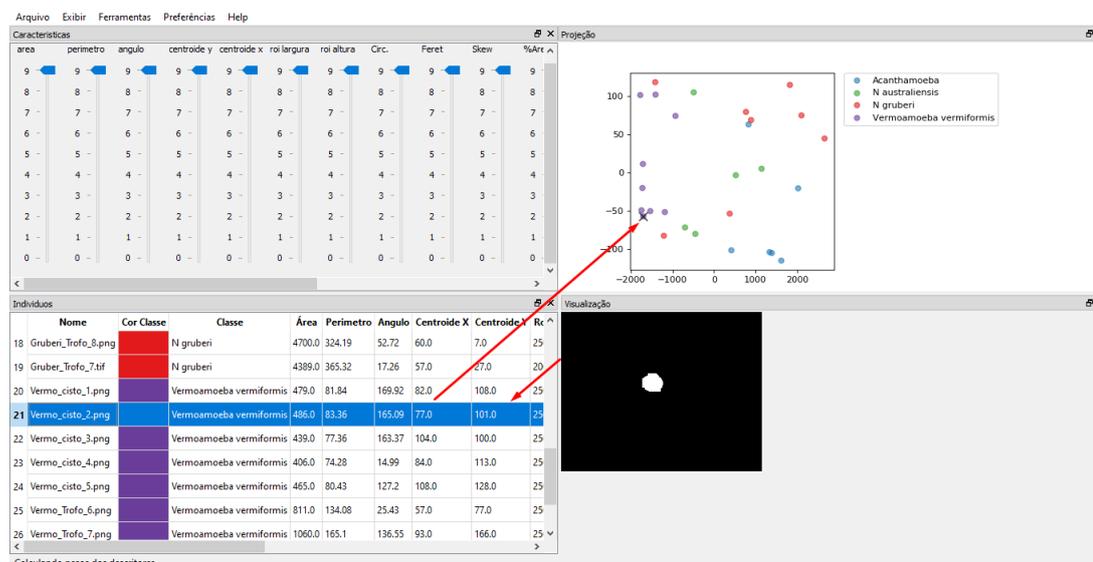
A : Representa a relevância de cada descritor de característica da espécie. Onde o usuário terá a liberdade para alterar os pesos que variam de 0-9, totalizando entre dez opções de pesos, pesos estes que serão determinantes para averiguar se determinada característica é de suma importância para estabelecer uma determinada espécie. Sendo 9 relevância máxima e 0 relevância mínima de cada descritor.

B : Projeção multidimensional das imagens representadas. Onde é feito o cálculo dos vetores de cada descritor e representado no plano. A projeção que é apresentada mostra cada círculo significa uma imagem e cada cor uma classe. O sistema aceita o número máximo de doze (12) classes.

C : A tabela é composta pela primeira coluna o nome de cada imagem para facilitar o entendimento do usuário, segunda coluna a cor que cada imagem pertence. A partir terceira coluna mostra o valor de cada descritor morfológico.

D : Ao clicar em qualquer coluna da tabela descrita no bloco C ou em qualquer círculo da projeção descrita no bloco D. O sistema informa para o usuário a imagem que está sendo representada em outros formatos, como mostrado na Figura 20.

Figura 20 – Interação possível feita pelo usuário ao selecionar um vetor de características, visualizando seu ponto correspondente na projeção e imagem associada.



Fonte: Elaborado pela autora.

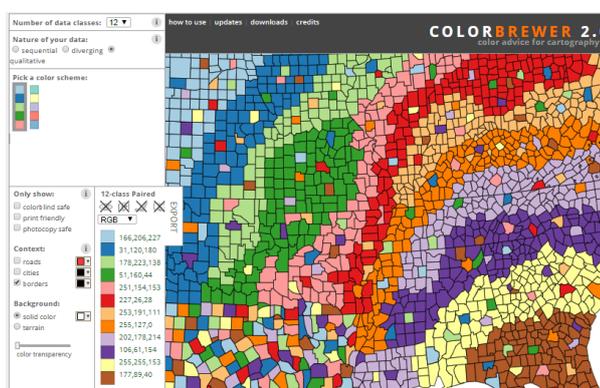
4.7 Outras tecnologias utilizadas

ColorBrewer

No Daltonismo umas das formas mais comuns é não possui a capacidade de diferenciar as cores. Visando essa questão foi utilizada uma ferramenta online, chamada ColorBrewer, disponível em <<http://www.colorbrewer.org>>. Criada por Mark Harrower e Cynthia Brewer no ano de 2003, a ferramenta se baseia no estudo feito por Judy Olson (OLSON; BREWER, 1997) que se refere avaliação de escolha de cor do indivíduo com deficiência de visão de cores vermelho-verde.

Como pode ser visto na Figura 21, o site é estruturado com trinta e cinco esquemas de cores são divididos em três grupos: sequencial (18 esquemas), divergente (9 esquemas) e qualitativo (8 esquemas). Cada grupo tem seu próprio número de classes possíveis para os esquemas de cores, com base na natureza do grupo. Todos começam com três classes e com o grupo divergente ao máximo com 12 cores. O ColorBrewer fornece a escolha de cores em três formatos, são eles: CMYK, RGB, HEXADECIMAL, sendo esse último o adotado aqui.

Figura 21 – Site ColorBrewer.



Fonte: <http://www.colorbrewer.org>

Quando o usuário da aplicação não possui qualquer deficiência de visão, foi utilizado o esquema de 12 cores: #a6cee3, #1f78b4, #b2df8a, #33a02c, #fb9a99, #e31a1c, #fdbf6f, #ff7f00, #cab2d6, #6a3d9a, #ffff99 e #b15928. Já quando o usuário apresenta tal deficiência, foi utilizado o esquema de cores: #a6cee3, #1f78b4, #b2df8a e #33a02c, limitando em quatro o número de classes simultâneas para análise.

5 RESULTADOS

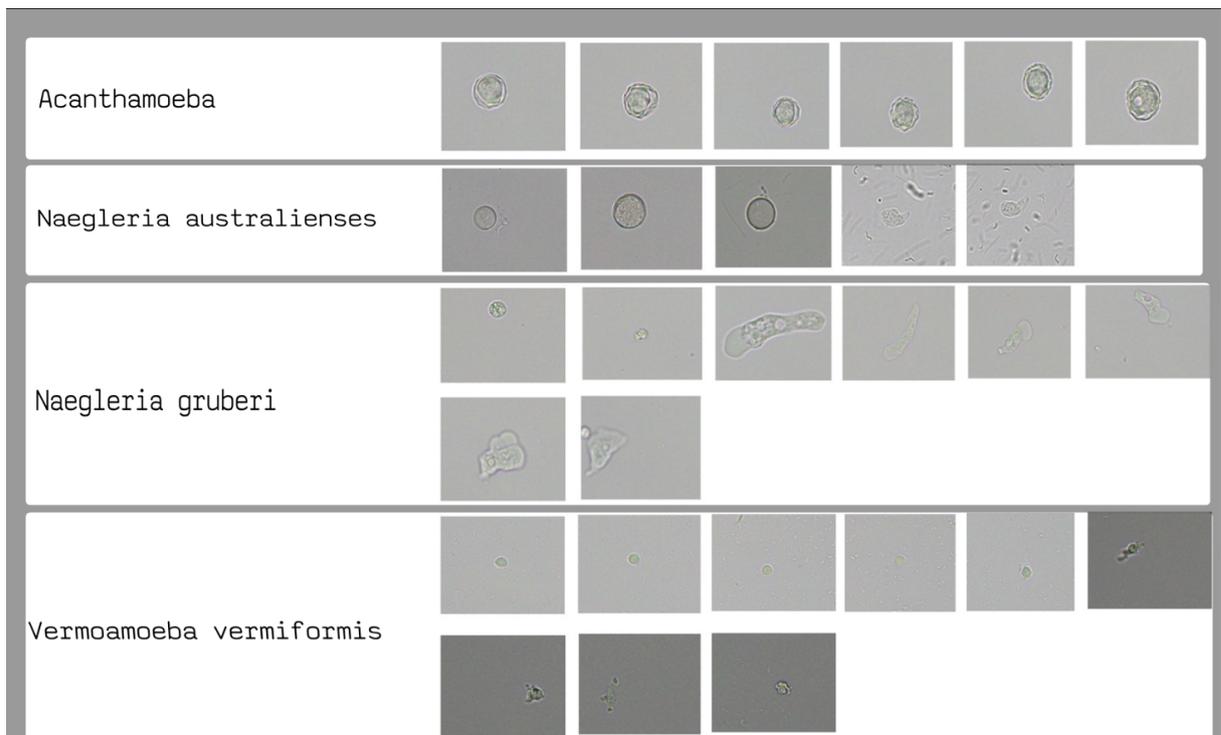
Para promover a transparência e a colaboração, o código-fonte da ferramenta está disponível publicamente no endereço <<https://github.com/Anakellydesouzaa/tcc>>. Essa disponibilidade no GitHub não apenas oferece a oportunidade para outros desenvolvedores contribuírem, mas também permite o acesso ao código-fonte, promovendo uma abordagem colaborativa e enriquecedora para o desenvolvimento do projeto.

5.1 Geração do *dataset*

Nessa seção é ilustrado o conjunto de dados gerado após as etapas de segmentação, para as diferentes espécies analisadas.

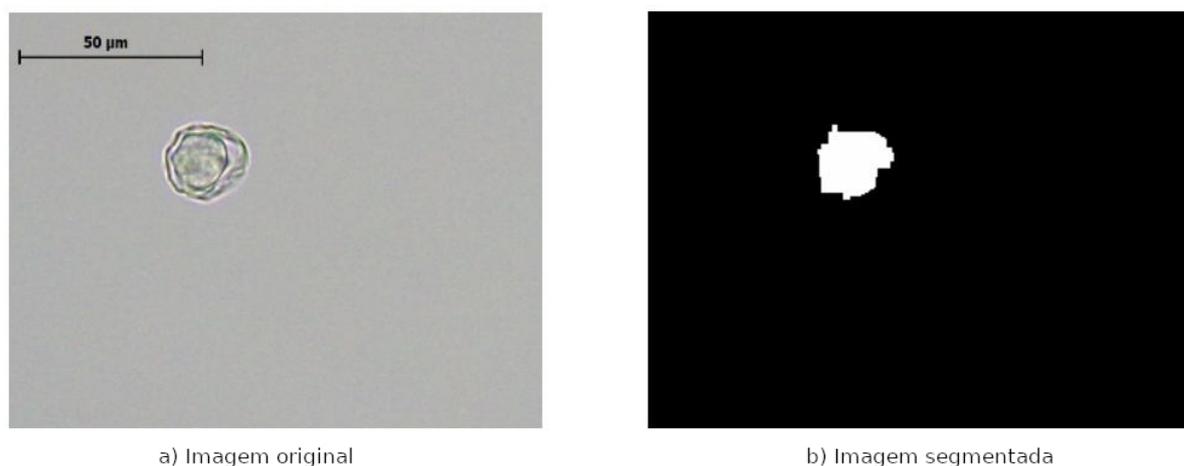
Na Figura 22 são ilustradas todas as espécies consideradas para esse trabalho, bem como as respectivas imagens coletadas.

Figura 22 – Dataset contendo as imagens originais.



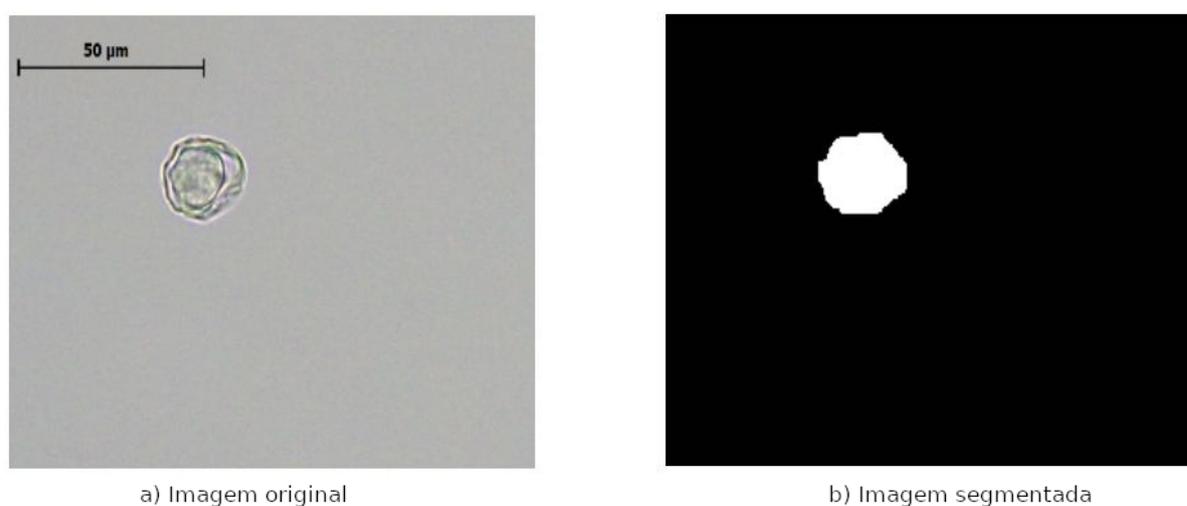
Fonte: Elaborado pela autora.

Inicialmente foi utilizada a metodologia proposta por Ginoris et al. (2007), para efetuar o pré-processamento na imagem. Então foi verificado que esse modelo de processamento acarretou em um resultado não tão bom do ponto de vista do aspecto morfológico da célula, como mostra na Figura 23, um exemplo para a espécie *Acanthamoeba*.

Figura 23 – Segmentação de uma *Acanthamoeba* utilizando Ginoris et al. (2007).

Fonte: Elaborado pela autora.

A partir desses resultados obtidos, foi optado por ser feita a segmentação manualmente, obtendo melhores resultados como mostrado na Figura 24.

Figura 24 – Segmentação de uma *Acanthamoeba* utilizando segmentação manual.

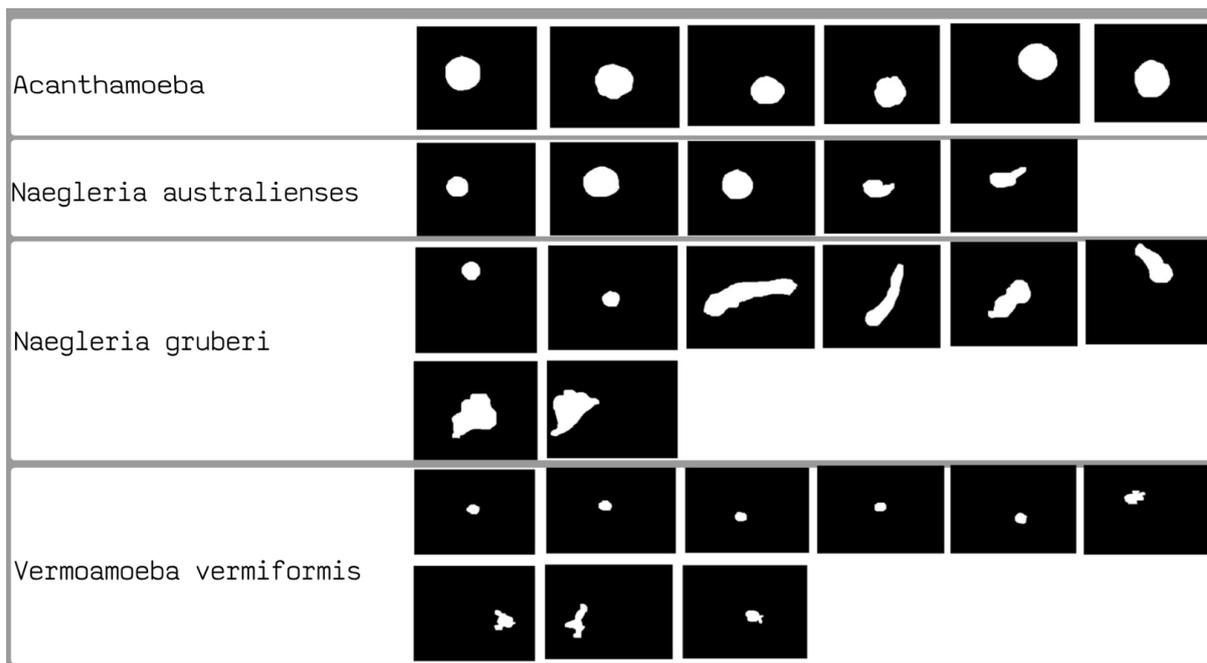
Fonte: Elaborado pela autora.

Em seguida, na Figura 25 são mostradas as imagens das amebas após passarem pelo processo de segmentação manual, onde são obtidas imagens que estarão aptas para análises das espécies de amebas. Essas imagens compõem o *dataset* analisado neste trabalho.

5.2 Análise das projeções multidimensionais

É relevante destacar que ao longo desta sessão, espera-se alcançar melhores resultados, nos quais as células da mesma espécie estejam mais próximas entre si e, conse-

Figura 25 – Dataset ilustrando as imagens obtidas após segmentação manual



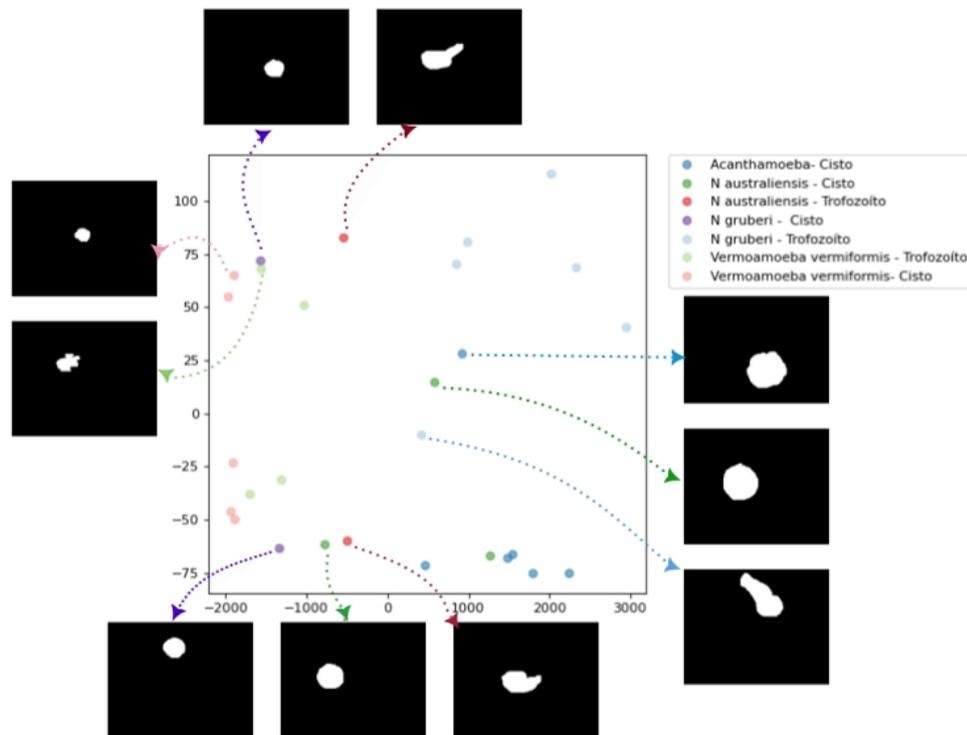
Fonte: Elaborado pela autora.

quentemente, mais distantes das células pertencentes a espécies diferentes. Isso ocorre ao comparar os ciclos (Cisto e Trofozoítos) entre diferentes espécies ou ao realizar comparações entre ciclos distintos da mesma espécie.

O pior resultado, por sua vez, ocorre na situação oposta ao melhor resultado, em que as células de espécies diferentes estão suficientemente próximas, impossibilitando a definição de uma “linha reta” para separar ambas as espécies. E também é importante ressaltar que, ao longo da apresentação na seção de resultados, a referência a esses descritores como “ativos” implica que foi atribuído o peso máximo a cada um deles.

Como um primeiro experimento, foram incluídas todas as espécies listadas na Figura 25, bem como foram considerados todos os descritores ativos para todas as espécies em ambos os ciclos disponíveis. A Figura 26 ilustra a visualização do resultado dessa projeção. Se os descritores fossem adequados o suficiente, deveríamos conseguir separar de forma simples todas as diferentes classes observadas. No entanto, observou-se que com essa abordagem não foi possível alcançar a separação desejada entre as espécies. Isso ajuda a motivar a necessidade de avaliar como o ajuste de descritores pode alterar os resultados. Nos experimentos feitos ao longo do trabalho, essa estratégia não demonstrou resultado ótimo, que melhorasse significativamente esse caso ilustrado, no sentido de conseguir separar bem todas as espécies e ciclos considerados. A partir disso, foram feitos outros tipos de análise descritas em seguida.

Figura 26 – Visualização da projeção de todo o dataset utilizando todos os descritores ativos.



Fonte: Elaborado pela autora.

5.2.1 Projeções por espécie

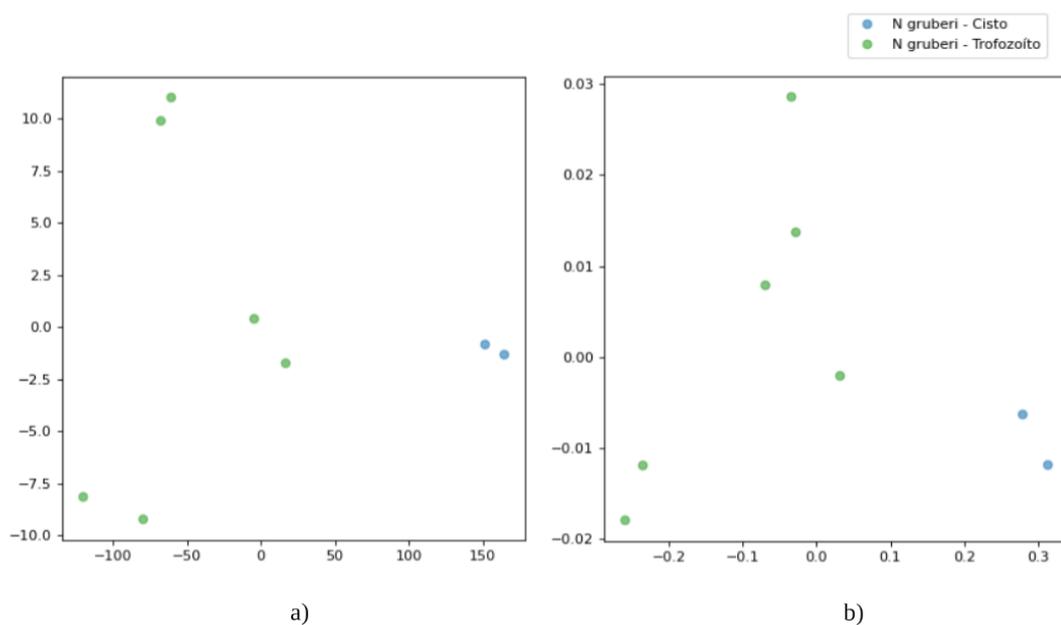
Nesta parte, exploramos as análises que examinam as interações dentro da mesma espécie, porém em fases diferentes, como cisto e trofozoíto. O objetivo é avaliar a importância de cada descritor na distinção de estágios entre as espécies.

5.2.1.1 *Naegleria gruberi*

Na espécie *Naegleria gruberi*, alguns descritores foram notáveis para distinguir claramente os dois ciclos. Quando ativos, Perímetro e Feret mostraram uma Silhueta Média significativa de 0,742, conforme mostrado na Figura 27a. No entanto, descritores como Circularidade e Solidez tiveram um desempenho inferior, revelando uma Silhueta Média de -0,268 na Figura 27b.

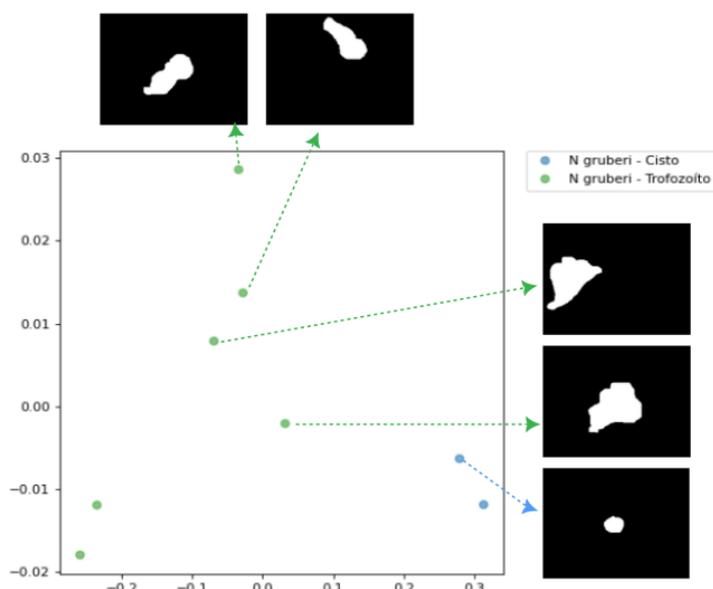
A Figura 28 exibiu as imagens correspondentes ao pior resultado, Figura 27b, ilustrando como cada célula é representada na espécie *Naegleria gruberi*. Observou-se que, mesmo compartilhando o mesmo estágio trofozoíto, as células mostraram uma proximidade visual entre objetos distintos, resultando em um desempenho insatisfatório. Para facilitar

Figura 27 – Ilustração visual da distinção entre os ciclos da espécie *Naegleria gruberi*. Em a) o melhor resultado, em b) o pior resultado.



Fonte: Elaborado pela autora.

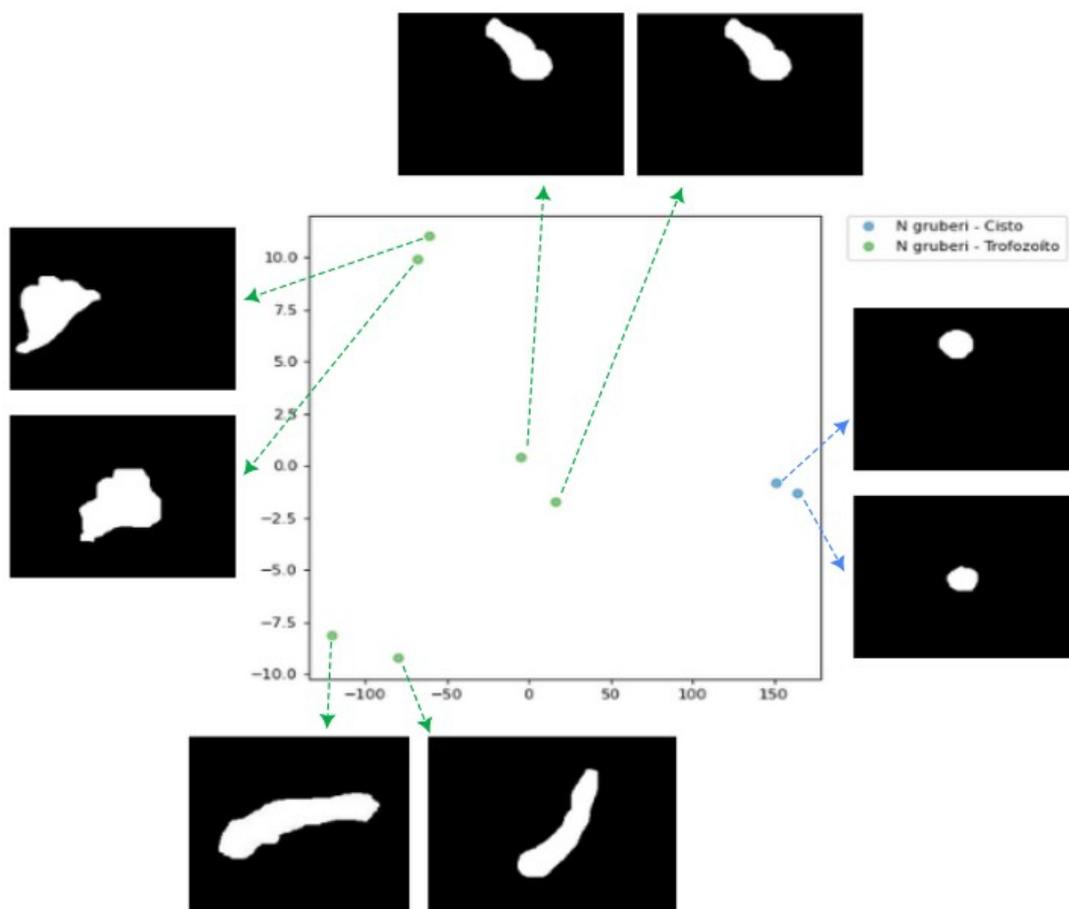
Figura 28 – Apresentação das imagens das células correspondentes ao pior resultado da espécie *Naegleria gruberi*, Figura 27b.



Fonte: Elaborado pela autora.

a compreensão, a Figura 29 apresenta as imagens referente à Figura 27a.

Figura 29 – Apresentação das imagens das células correspondentes ao melhor resultado da espécie *Naegleria gruberi*, Figura 27a



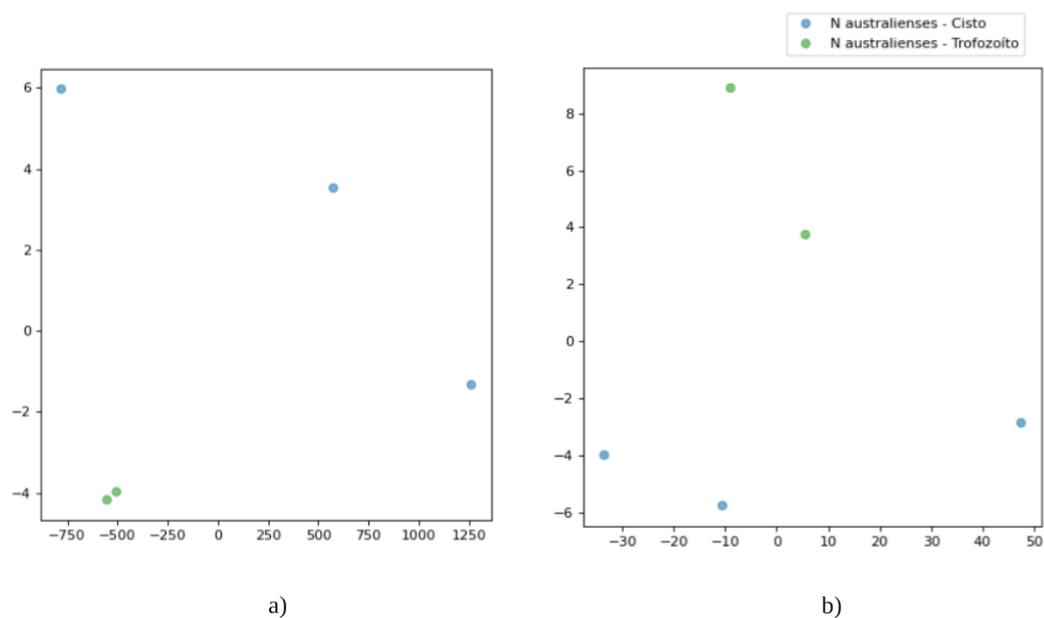
Fonte: Elaborado pela autora.

5.2.1.2 *Naegleria australiensis*

Após análise, foi observado que na espécie *Naegleria australiensis*, os descritores ativos, como Área e MinFeret, foram eficazes na distinção entre os dois ciclos, exibindo uma silhueta média de 0,274, conforme representado na Figura 30a. Em contrapartida, descritores ativos como Perímetro, Feret e Circularidade mostraram menor eficiência, com uma silhueta média de -0,027, como evidenciado na Figura 30b.

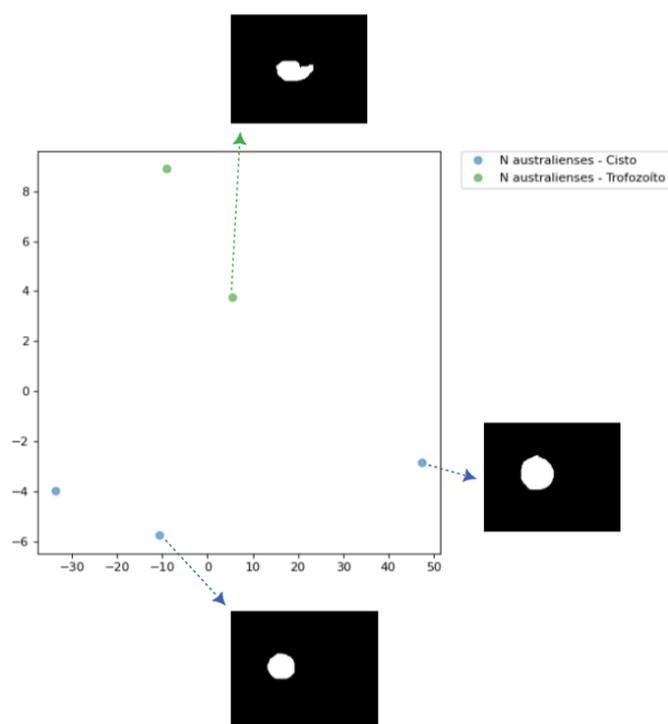
Já na Figura 31, uma representação das imagens das células relacionadas ao pior resultado, Figura 30b, foi realizada com o propósito de exemplificar qual célula representa cada imagem na espécie *Naegleria australiensis*. Notou-se que ao utilizar os descritores menos eficazes, as células em ambos os estágios apresentaram distâncias semelhantes, conforme evidenciado nas imagens das células.

Figura 30 – Ilustração visual da distinção entre os ciclos da espécie *Naegleria australiensis*. Em a) o melhor resultado, em b) o pior resultado.



Fonte: Elaborado pela autora.

Figura 31 – Apresentação das imagens das células correspondentes ao pior resultado da espécie *Naegleria australiensis*, Figura 30b.

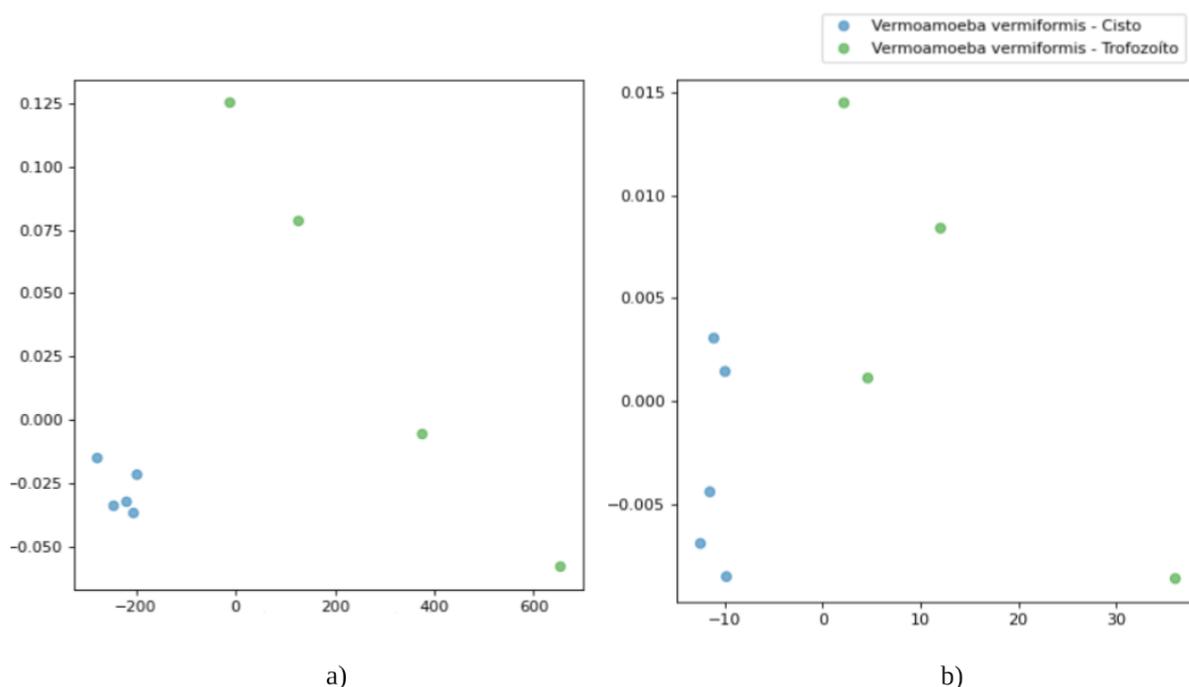


Fonte: Elaborado pela autora.

5.2.1.3 *Vermamoeba vermiformis*

Os descritores que demonstraram maior eficácia na separação da espécie *Vermamoeba vermiformis* entre seus dois ciclos, quando ativos, foram Área e Circularidade, registrando uma silhueta média de 0,781, conforme ilustrado na Figura 32a. Em contraste, os descritores que apresentaram menor eficiência quando ativos foram Feret e Solidity, revelando uma silhueta média de -0,495, conforme evidenciado na Figura 32b.

Figura 32 – Ilustração visual da distinção entre os ciclos da espécie *Vermamoeba vermiformis*.



Fonte: Elaborado pela autora.

Cabe observar que que visualmente para qualquer uma das três espécies, mesmo nos piores casos obtidos, os descritores utilizados possibilitam diferenciar bem as classes, facilitando uma posterior tarefa de classificação.

Metrica de consistência

Na Tabela 1 não foi analisada a espécie *Acanthamoeba*, pois esta não possui o estágio trofozoíto pertencente a esse dataset gerado. Contudo, dentre as três espécies restantes foi notado que o descritor Área (quando ativo) foi retornado como pelo menos um dos descritores com melhor resultado para diferenciação dos ciclos em duas das três espécies analisadas. Já em contrapartida os descritores: Circularidade e Solidity (quando ativos) não foram retornados com uma boa eficiência dentre as de três espécies analisadas. É importante destacar que o descritor Feret pertence ao grupo dos descritores que consegue separar a espécie *Naegleria gruberi*, pois suas células possuem um tamanho considerável

de diferença em relação a cistos e trofozoítas, porém analisando as espécies *Naegleria australiensis* e *Vermamoeba vermiformis* essa situação já não ocorre, pois o tamanho das células são semelhantes como mostrado na Figura 22, sendo assim o descritor Feret se classifica como um descritor não eficiente nessa espécie. Dessa forma, pode se concluir que mesmo que um descritor seja relevante para determinada espécie, este pode não ocorrer com uma outra espécie dependendo de suas características específicas.

Tabela 1 – Análise comparativa de descritores entre cistos e trofozoítos da mesma espécie.

ESPÉCIES	MELHOR RESULTADO		PIOR RESULTADO	
	SILHUETA	DESCRITORES	SILHUETA	DESCRITORES
<i>N. gruberi</i>	0,742	Perímetro, Feret	-0,268	Circularidade, Solidity
<i>N. australiensis</i>	0,274	Área, MinFeret	-0,027	Circularidade, Perímetro, Feret
<i>V. vermiformis</i>	0,781	Área, Circularidade	-0,495	Feret, Solidity

Fonte: Elaborado pela autora.

5.2.2 Projeções por estágio celular

Nesta seção são abordadas as análises que consideram as relações entre as espécies em um mesmo estágio celular, seja ele cisto ou trofozoíto. O propósito é examinar a relevância de cada descritor na diferenciação de estágios entre diferentes espécies.

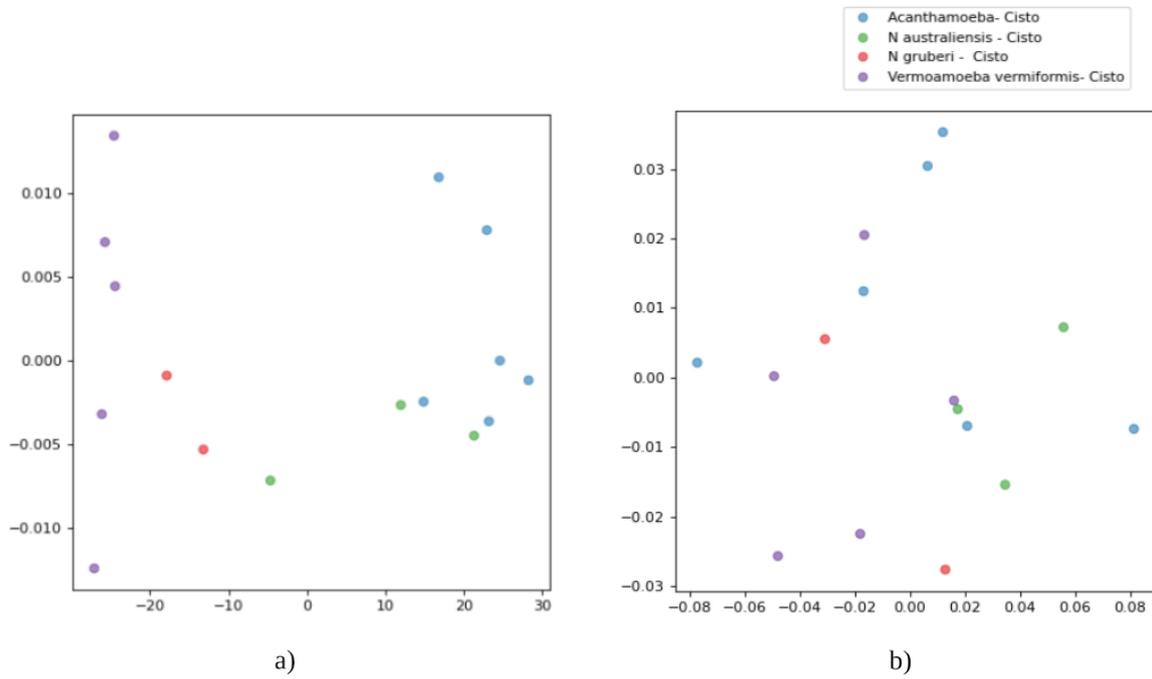
5.2.2.1 Cisto

Na Figura 33, foi realizada uma comparação entre todas as espécies analisadas neste estudo, focando especialmente no estágio de Cisto. Na Figura 33a, os resultados mais satisfatórios foram apresentados utilizando os descritores ativos, Feret e Solidity, que resultaram na melhor silhueta, avaliada em 0,433. Por outro lado, na Figura 33b, foram utilizados descritores que não demonstraram boa eficácia, Circularidade e Round, resultando em uma silhueta de -0,262.

A Figura 34 exibe a imagem das células correspondentes ao pior resultado ilustrado na Figura 33b. Essa representação tem o propósito de exemplificar qual célula representa cada imagem no estágio de cisto. Observa-se que os descritores que não apresentaram boa eficiência quando ativos não conseguiram separar adequadamente as espécies, conforme o esperado.

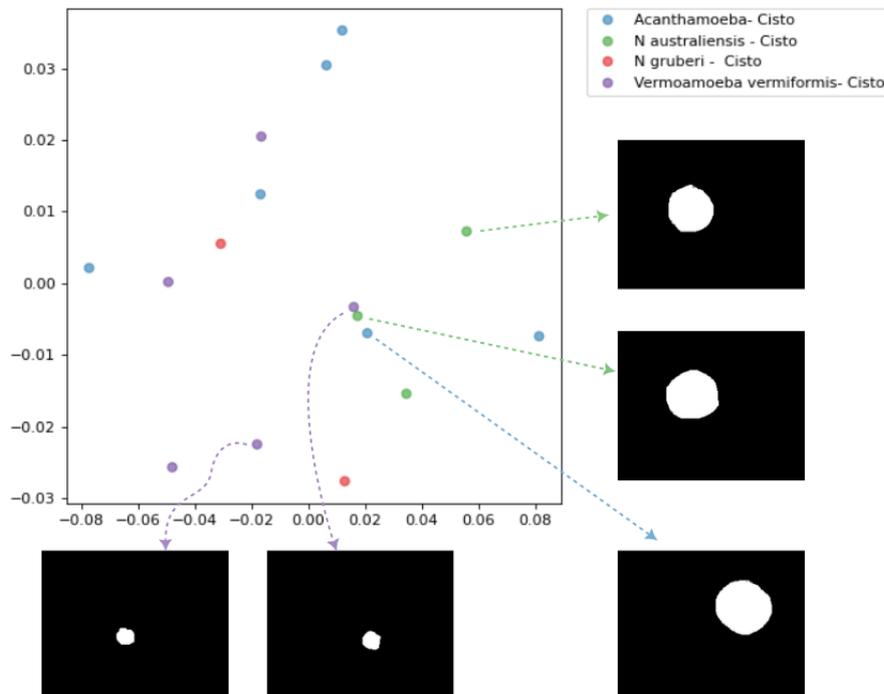
Ainda sobre essas células no estágio de Cisto, na Figura 35, apresenta-se a representação das células correspondentes à Figura 33a. Como evidenciado, as células das espécies *Acanthamoeba* e *Naegleria australiensis* encontram-se mais próximas, o que é considerado um resultado positivo, uma vez que essas espécies são bastante semelhantes entre si. Essa proximidade entre células parecidas está alinhada com a proposta deste estudo de aproximar as células equivalentes.

Figura 33 – Ilustração visual da distinção entre o estágio de cisto de todas as espécies abordadas neste trabalho.



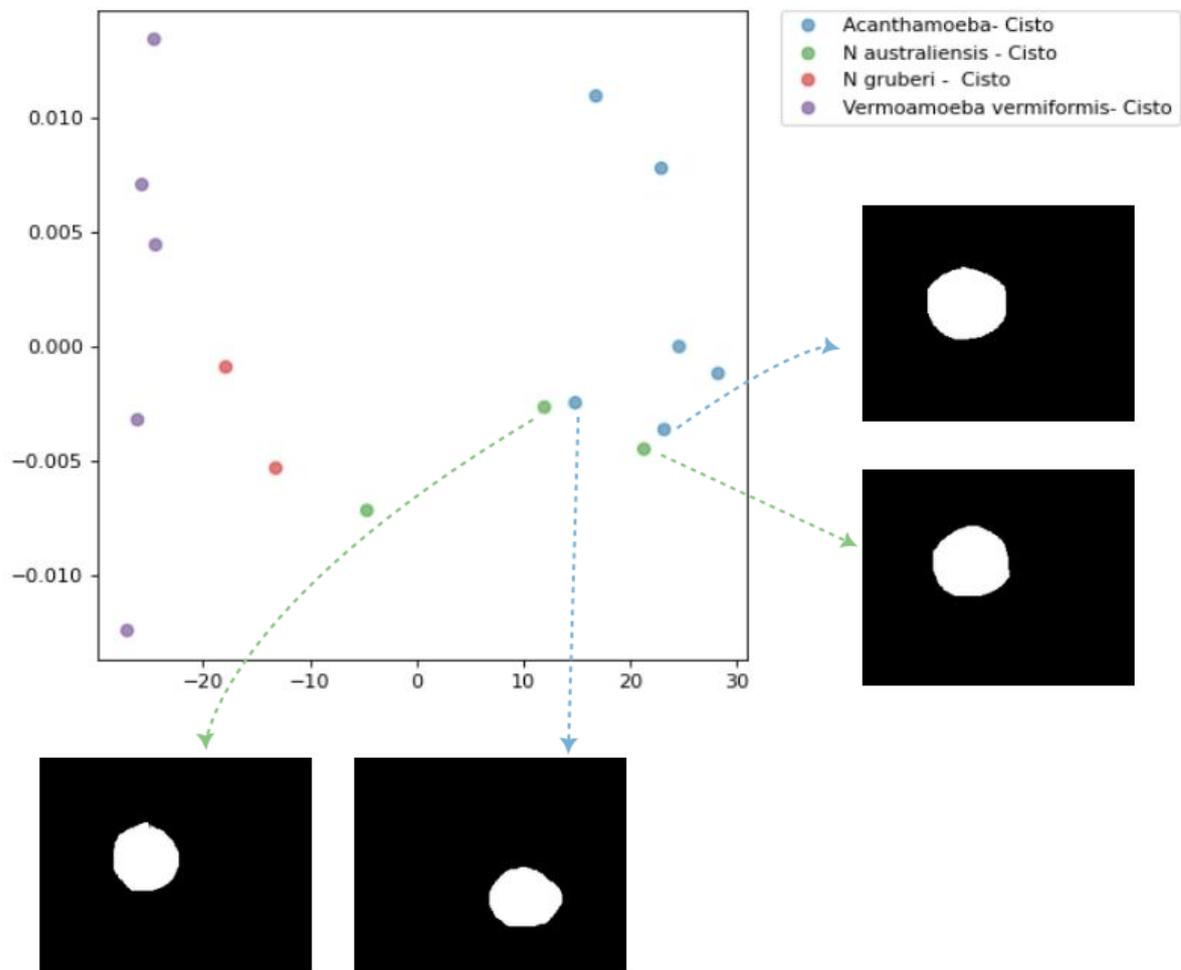
Fonte: Elaborado pela autora.

Figura 34 – Apresentação das imagens das espécies no estágio de cisto correspondentes à Figura 33b.



Fonte: Elaborado pela autora.

Figura 35 – Apresentação das imagens das espécies no estágio de cisto correspondentes à Figura 33a.



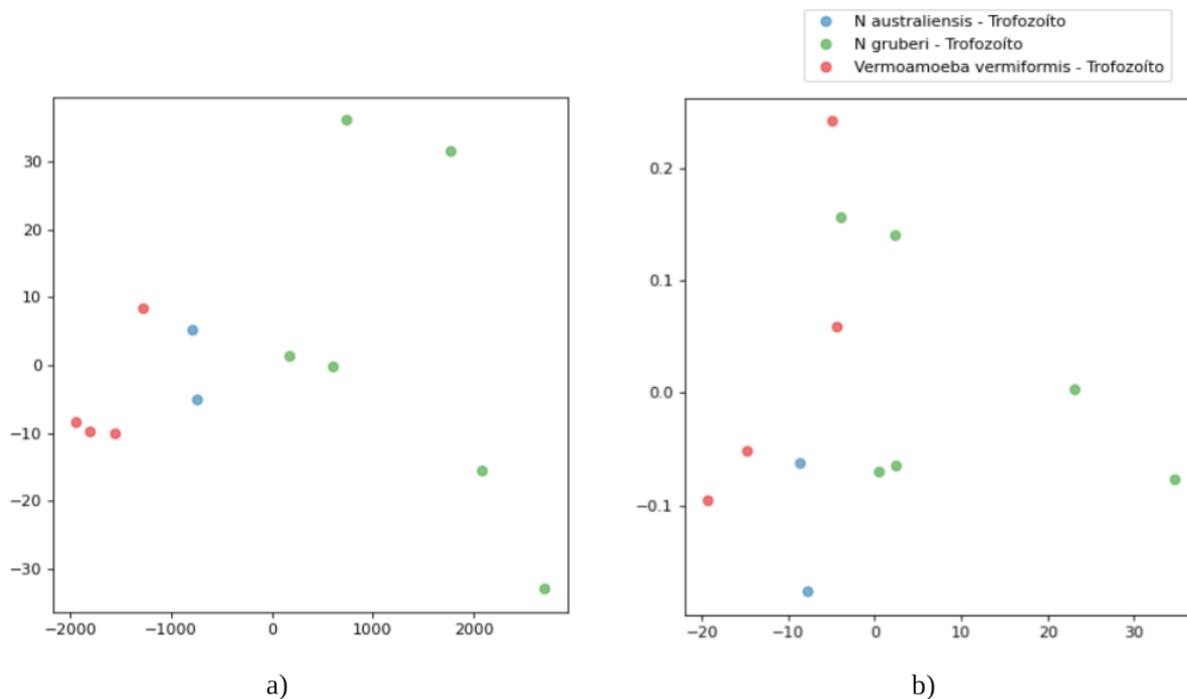
Fonte: Elaborado pela autora.

5.2.2.2 Trofozoíto

Um exemplo ilustrativo desses resultados pode ser observado na Figura 36, ao analisar todas as células das espécies *Naegleria australiensis*, *Naegleria gruberi* e *Vermamoeba vermiformis* no estágio trofozoíto, foi observado um desempenho superior na diferenciação das espécies quando utilizados os descritores ativos: Área, Feret e MinFeret, como evidenciado na Figura 36a, apresentando uma métrica de silhueta mais favorável, alcançando 0,489. Em contrapartida, na Figura 36b, a utilização dos descritores ativos: Circularidade e Round resultou em uma eficácia inferior na separação das espécies, a silhueta é inferior registrando -0,165.

Vale notar que é possível separar as espécies através de linhas verticais no melhor resultado, o que não acontece no pior resultado.

Figura 36 – Representação das células da espécie no estágio Trofozoíto. Em a) o melhor resultado, em b) o pior resultado.



Fonte: Elaborado pela autora.

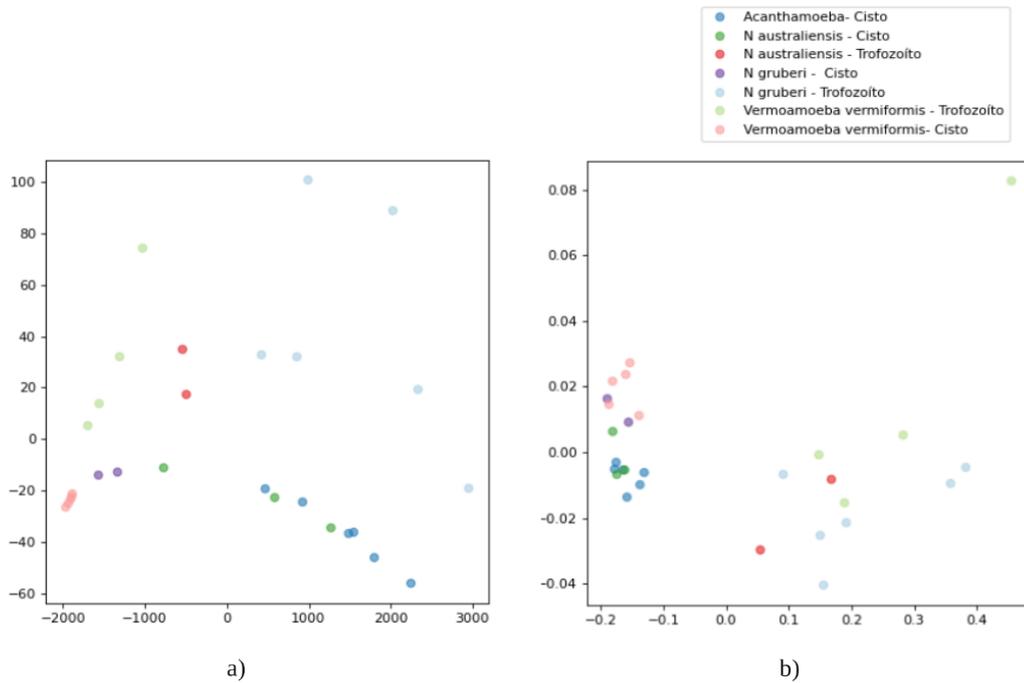
5.2.2.3 Cisto e trofozoíto

Na Figura 37, é demonstrado o resultado entre a comparação de todas as espécies analisadas neste trabalho, levando em consideração os estágios de Trofozoíto e Cisto. Na Figura 37a, são exibidos os resultados utilizando os descritores ativos: Área, Feret e Perímetro, que proporcionaram a melhor silhueta avaliada em 0,074, indicando um resultado mais eficaz. Por outro lado, na Figura 37b, foram utilizados descritores ativos que não apresentaram boa eficácia: Solidity, Circularidade e Skew, resultando em uma silhueta de -0,187.

A Figura 38 por sua vez, exibe o resultado da Figura 37a, com o propósito de exemplificar qual célula representa cada imagem das espécies, as quais, conseqüentemente, estão mais próximas conforme sua similaridade.

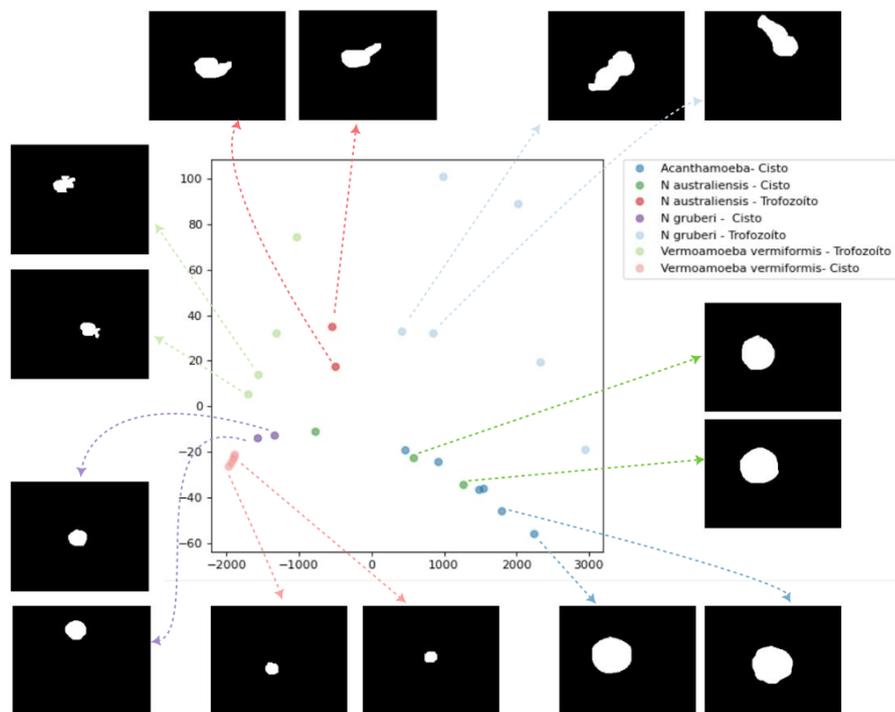
Conforme demonstrado na Figura 39, destacando as espécies da Figura 37b, observa-se que os descritores ativos não foram eficazes. Eles não foram capazes de separar com eficiência as espécies *Naegleria gruberi*, *Naegleria australiensis* e *Vermamoeba vermiformis*, todas no estágio de Trofozoíto, mesmo não possuindo características similares entre si.

Figura 37 – Representação de todo o *dataset* utilizado neste trabalho. Células no estágio: Cisto e Trofozoíto. Em a) o melhor resultado, em b) o pior resultado.



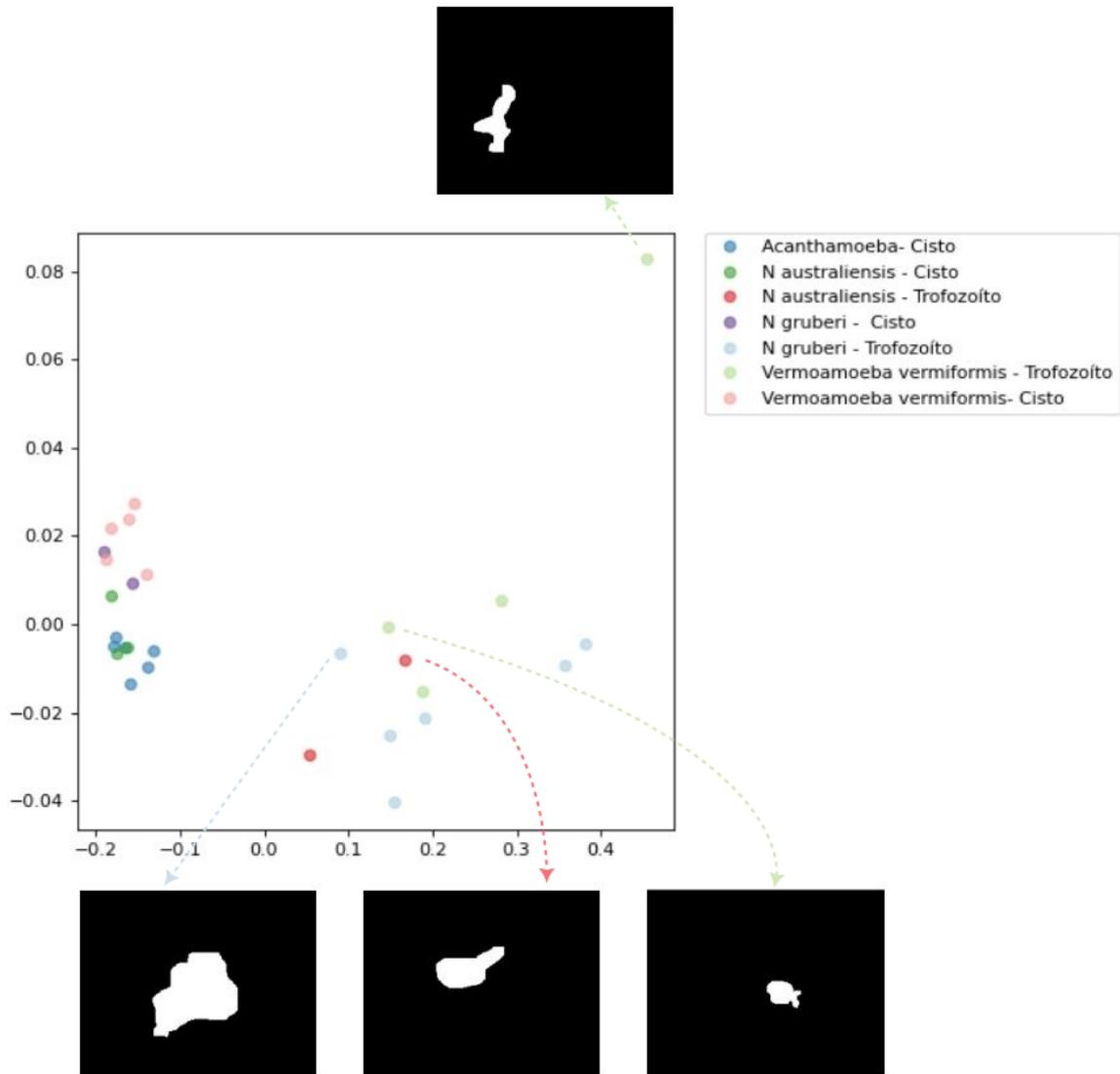
Fonte: Elaborado pela autora.

Figura 38 – Representação da espécies da Figura 37a.



Fonte: Elaborado pela autora.

Figura 39 – Representação da espécie da Figura 37b.



Fonte: Elaborado pela autora.

Métrica de consistência

A Tabela 2 apresenta uma análise comparativa dos resultados entre cistos e trofozoítos de diferentes espécies. Foi possível notar que os descritores Área e Feret produziram resultados mais eficazes quando as células das espécies apresentam tamanhos consideravelmente distintos, conforme ilustrado na Figura 25. Neste exemplo, a espécie *Vermamoeba vermiformis* no estágio de cisto exibe uma Região de Interesse (ROI) significativamente menor em comparação com a espécie *Acanthamoeba*, que também está no estágio de cisto e esse é um dos pilares para que a separação das espécies se torne mais assertiva. No entanto, observou-se que os descritores Circularidade e Solidity, quando ativos, não demonstraram eficácia na separação das espécies em ambos os estágios.

Após a análise de todas as espécies de amebas em seus estágios de cisto e trofozoíto, o resultado da métrica de silhueta foi registrado em 0,041. Esse valor reflete a situação na qual todos os descritores morfológicos foram considerados ativos, ou seja, receberam peso máximo de relevância no processo de análise. A ponderação máxima atribuída a todos os descritores sugere que a análise considerou igualmente todos os aspectos morfológicos das amebas, sem priorizar nenhum em particular. Esse resultado indica uma separação moderada entre os grupos das diferentes espécies de amebas, revelando uma diferenciação razoável com base nas características morfológicas analisadas. Vale notar que o resultado obtido após ajuste nos pesos dos descritores, ilustrado na Figura 37a, obteve uma melhora na silhueta ao se comparar com o *dataset* completo original, mostrando que é possível melhorar a silhueta com intervenção do usuário através da aplicação desenvolvida neste trabalho.

Tabela 2 – Análise comparativa de descritores entre espécies diferentes considerando as formas cisto e trofozoíto.

ESPÉCIES	MELHOR RESULTADO		PIOR RESULTADO	
	SILHUETA	DESCRITORES	SILHUETA	DESCRITORES
Cisto	0,434	Feret, Solidity	-0,262	Circularidade, Round
Trofozoíto	0,489	Área, Feret, MinFeret	-0,165	Circularidade, Skew, Solidity
Cisto + Trofozoíto	0,041	TODOS	—	—
Cisto + Trofozoíto	0,074	Área, Perímetro, Feret	-0,187	Circularidade, Skew, Solidity

Fonte: Elaborada pela autora.

6 CONCLUSÃO

6.1 Considerações Finais

A análise de espécies de amebas de vida livre (AVL) utilizando processamento digital de imagens e projeção multidimensional é uma abordagem promissora na análise morfológica, para auxiliar a diferenciação desses organismos microscópicos. O desenvolvimento de um sistema que utiliza descritores matemáticos e projeção multidimensional permite uma análise morfológica detalhada das espécies de AVL, contribuindo para o estudo e a compreensão desses protozoários.

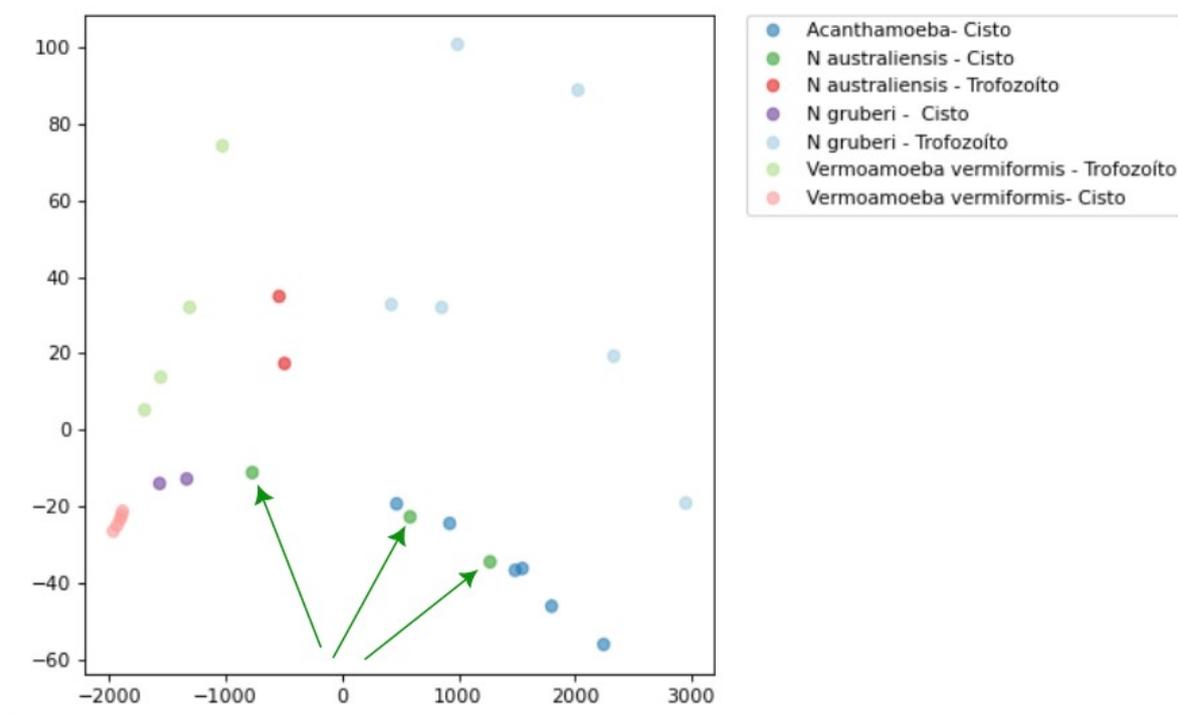
Ao longo deste projeto, foi demonstrado que é possível obter resultados efetivos na diferenciação das espécies de amebas com um número reduzido de descritores. Isso representa um avanço que pode auxiliar a elaboração de chaves de classificação, pois indicaria quais perguntas poderiam ser respondidas primeiro. Estas chaves são ferramentas importantes na taxonomia e na identificação de organismos, e são utilizadas para agrupar e distinguir espécies com base em características específicas por meio de perguntas.

Portanto, conclui-se que o sistema desenvolvido neste trabalho tem o potencial de impactar a análise e classificação de espécies de amebas de vida livre, fornecendo uma ferramenta auxiliar para pesquisadores e profissionais da área biológica. O uso de tecnologias avançadas, como processamento digital de imagens e projeção multidimensional, promete avançar significativamente nosso entendimento e capacidade de lidar com organismos microscópicos, contribuindo assim para a saúde pública e para o avanço científico.

6.2 Trabalhos Futuros

Para aprimorar a ferramenta, um trabalho futuro seria à medida que o usuário clicar nas células na projeção e aproximar os grupos de espécies semelhantes, ao mesmo tempo em que distancia os grupos distintos, essa ação resultará em um cálculo de otimização que ajusta os pesos e fornece os melhores descritores para a separação automatizada das classes. Isso é exemplificado na Figura 40, onde as setas verdes indicam o grupo de células da *Naegleria australiensis* que o usuário tentaria aproximar para torná-las mais próximas, determinando assim os melhores descritores para diferenciar as espécies. Para tal, pode-se adotar a estratégia de *Weighted Multidimensional Scaling (WMDS)* proposta por Self et al. (2018), no seu sistema Andromeda.

Figura 40 – Exibição dos dados analisados neste estudo, ressaltando uma possível progressão desta pesquisa em estudos subsequentes.



Fonte: Elaborada pela autora.

Referências

- ALFIERI, S. C. et al. Proteinase activities in total extracts and in medium conditioned by *acanthamoeba polyphaga* trophozoites. *Journal of Parasitology*, v. 86, n. 2, p. 220–227, 2000. Citado na página 22.
- ARAÚJO, W. O. de; COELHO, C. J. Análise de componentes principais (pca). *University Center of Anápolis, Annapolis*, 2009. Citado na página 28.
- ÁVILA, A. L.; SPÍNOLA, R. O. Introdução à engenharia de requisitos. *Engenharia de software Magazine*, p. 46–52, 2007. Citado na página 23.
- BANON, G. J. F.; BARRERA, J. *Bases da Morfologia Matemática para a análise de imagens binárias*. [S.l.]: UFPE-DI, 1994. Citado na página 10.
- BELLET, A.; HABRARD, A.; SEBBAN, M. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*, 2013. Citado na página 16.
- BELLINI, N. K. et al. Isolation of *naegleria* spp. from a brazilian water source. *Pathogens*, Multidisciplinary Digital Publishing Institute, v. 9, n. 2, p. 90, 2020. Citado 2 vezes nas páginas 2 e 22.
- CEDRIM, D. et al. Depth functions as a quality measure and for steering multidimensional projections. *Computers & Graphics*, Elsevier, v. 60, p. 93–106, 2016. Citado na página 17.
- FILHO, O. M.; NETO, H. V. *Processamento digital de imagens*. [S.l.]: Brasport, 1999. Citado na página 9.
- GINORIS, Y. et al. Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system. *Water research*, Elsevier, v. 41, n. 12, p. 2581–2589, 2007. Citado 7 vezes nas páginas , 19, 24, 32, 33, 55 e 56.
- GLOBO. *Ameba 'comedora de cérebro': caso de infecção na Flórida gera alerta nos EUA*. 2020. Disponível em: <<https://g1.globo.com/ciencia-e-saude/noticia/2020/07/06/ameba-comedora-de-cerebro-caso-de-infeccao-na-florida-gera-alerta-nos-eua.ghtml>>. Citado na página 2.
- GLOBO. *Ameba comedora de cérebros: 'Podemos ter casos no Brasil não diagnosticados', diz parasitologista*. 2023. Disponível em: <<https://g1.globo.com/ciencia-e-saude/noticia/2020/07/06/ameba-comedora-de-cerebro-caso-de-infeccao-na-florida-gera-alerta-nos-eua.ghtml>>. Citado na página 2.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Edgard Blucher, 2000. Citado 3 vezes nas páginas 9, 11 e 55.
- GONZALEZ, R. C.; WOODS, R. E. Image processing. *Digital image processing*, v. 2, p. 1, 2007. Citado na página 10.
- GONZALEZ, R. C.; WOODS, R. E. et al. Digital image processing. *Publishing house of electronics industry*, v. 141, n. 7, 2002. Citado 3 vezes nas páginas 7, 8 e 10.

- JEONG, D. H. et al. ipca: An interactive system for pca-based visual analytics. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2009. v. 28, n. 3, p. 767–774. Citado 2 vezes nas páginas 20 e 21.
- JR, R. H. *Segmentação de imagens por morfologia matemática*. Tese (Doutorado) — Instituto de Matemática e Estatística da Universidade de São Paulo, 07.03. 97., 1997. Citado na página 56.
- LABOR, S. *Microscópio Óptico Binocular - Partes, manuseio, limpeza e dica de compra*. 2016. Disponível em: <<https://shorturl.at/hprO3>>. Citado na página 4.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas e 15.
- LEVY, P. C. et al. Reconhecimento e segmentação do mycobacterium tuberculosis em imagens de microscopia de campo claro utilizando as características de cor e o algoritmo backpropagation. Universidade Federal do Amazonas, 2012. Citado 2 vezes nas páginas 4 e 5.
- MACHADO, B. B. et al. Exploração de espaços de características para imagens por meio de projeções multidimensionais. *Avanços em visão computacional*, p. 406, 2012. Citado na página 16.
- MACIVER, S. K.; PIÑERO, J. E.; LORENZO-MORALES, J. Is naegleria fowleri an emerging parasite? *Trends in parasitology*, Elsevier, v. 36, n. 1, p. 19–28, 2020. Citado na página 6.
- MANTILLA, E. R. *Avaliação da liberação mineral por análise de imagens digitais obtidas em microscópio óptico de luz refletida*. Dissertação (Mestrado) — Programa de Pós-graduação em Engenharia Metalúrgica, Materiais e de Minas, 2013. Citado na página 5.
- MATTA, W. N. Metodologia para detecção de máculas em micrografias utilizando morfologia matemática. 1998. Citado 3 vezes nas páginas 11, 12 e 13.
- MENEZES, L. Processamento de imagens digitais para visão robótica. *INTELLECTUS. Revista do Grupo Polis Educacional.*, p. 39, 2010. Citado na página 9.
- OLSON, J. M.; BREWER, C. A. An evaluation of color selections to accommodate map users with color-vision impairments. *Annals of the Association of American Geographers*, Wiley Online Library, v. 87, n. 1, p. 103–134, 1997. Citado na página 31.
- SANTOS, A. P. d. O. et al. Desenvolvimento de descritores de imagens para reconhecimento de padrões de plantas invasoras (folhas largas e folhas estreitas). Universidade Federal de São Carlos, 2009. Citado na página 14.
- SCHUSTER, F. L.; VISVESVARA, G. S. Free-living amoebae as opportunistic and non-opportunistic pathogens of humans and animals. *International journal for parasitology*, Elsevier, v. 34, n. 9, p. 1001–1027, 2004. Citado na página 1.
- SCORTECCI, J. F. *Estudo da via de incorporação de selenocisteínas: compreensão dos mecanismos de interações macromoleculares*. Tese (Doutorado) — Universidade de São Paulo, 2019. Citado na página 1.

- SCURI, A. E. Fundamentos da imagem digital. *Pontifícia Universidade Católica do Rio de Janeiro*, 1999. Citado na página 8.
- SELF, J. Z. et al. Observation-level and parametric interaction for high-dimensional data analysis. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, ACM New York, NY, USA, v. 8, n. 2, p. 1–36, 2018. Citado na página 47.
- SOHN, H.-J. et al. Efficient liquid media for encystation of pathogenic free-living amoebae. *The Korean journal of parasitology*, Korean Society for Parasitology, v. 55, n. 3, p. 233, 2017. Citado na página 22.
- SOUZA, T.; CORREIA, S. Estudo de técnicas de realce de imagens digitais e suas aplicações. *João Pessoa. Paraíba*, 2007. Citado na página 55.
- TEIXEIRA, L. H. et al. Ocorrência de amebas de vida-livre, dos gêneros *acanthamoeba* e *naegleria*, em pisos de ambientes internos, na universidade católica de santos, sp, brasil. *Universidade Católica de Santos*, 2008. Citado na página 6.
- TRABELSI, H. et al. Pathogenic free-living amoebae: epidemiology and clinical review. *Pathologie Biologie*, Elsevier, v. 60, n. 6, p. 399–405, 2012. Citado 2 vezes nas páginas 1 e 2.

Apêndices

APÊNDICE A – Instruções de uso do sistema

Funcionalidades básicas da interface

Com a função de auxiliar o usuário a utilizar o sistema foi desenvolvido algumas funcionalidades, como mostra na figura 41.

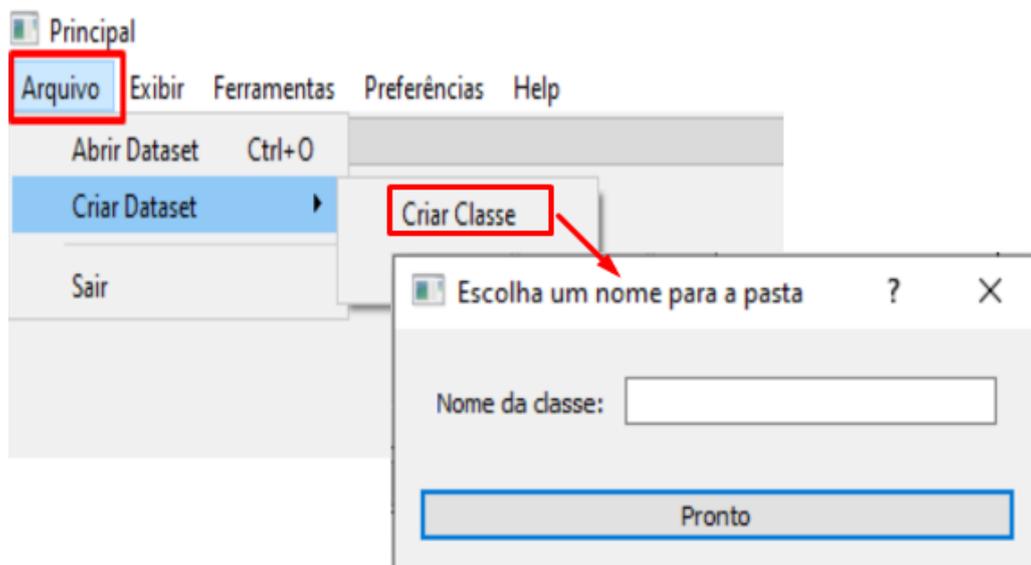
Figura 41 – Interface



Fonte: Elaborado pela autora.

Arquivo: Permite ao usuário selecionar obrigatoriamente o dataset a ser analisado ao clicar na opção “Abrir Dataset”. Caso o usuário opte por isso, após selecionar a pasta onde seu novo dataset será armazenado, ele pode criar um novo dataset selecionando “Criar Dataset”. Em seguida, é possível inserir os nomes das pastas (classes), clicando em “Criar Classe”, que serão utilizados para distinguir cada classe de espécies, conforme ilustrado na Figura 42.

Figura 42 – Interface

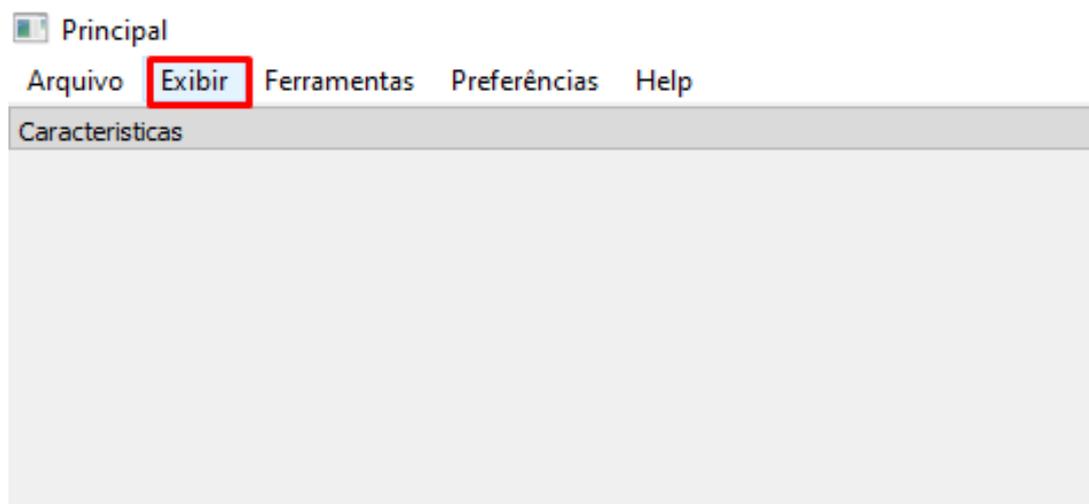


Fonte: Elaborado pela autora.

Exibir: Permite a visualização das imagens que estão sendo processadas, figura 43.

Ferramentas: Se a opção for criar o dataset, o passo subsequente é o “Pré-processamento”, que pode ser efetuado manualmente ou automaticamente. No procedimento manual, o usuário efetuará o recorte da região de interesse (corpo da célula) e, em seguida, gerará a imagem segmentada. Para a abordagem automática, uma vez realizado o recorte da

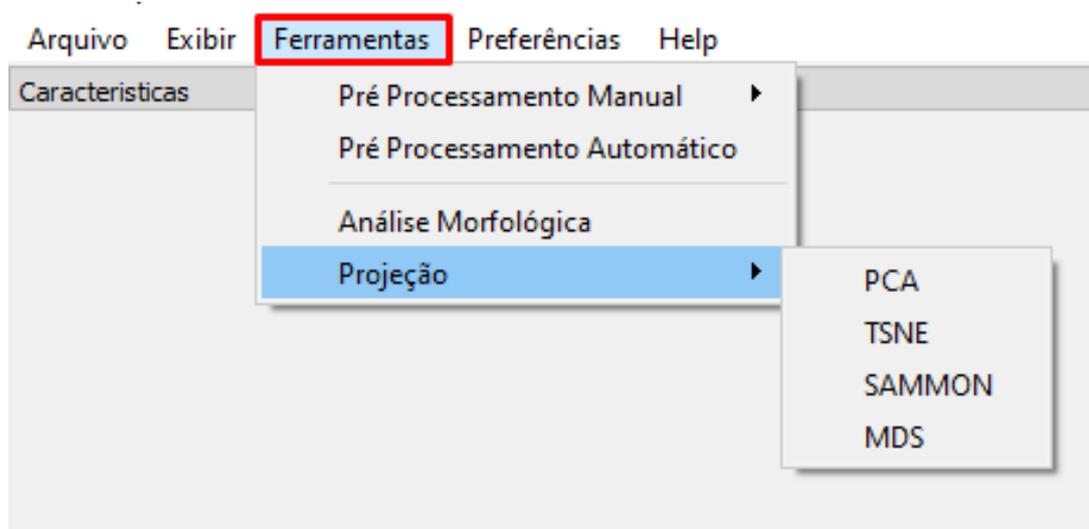
Figura 43 – Interface



Fonte: Elaborado pela autora.

região de interesse, ao pressionar o botão “Pré-Processamento Automático”, os filtros (Equalização, Mediana, Inversão, Binarização e Operadores) serão aplicados de forma sequencial, conforme demonstrado na Figura 52. Após a criação do dataset, para uma análise dos dados, basta selecionar “Análise Morfológica”, que apresentará os descritores na interface. Posteriormente, é viável optar por “Projeção” e selecionar um dos modelos disponíveis (PCA e TSNE), como evidenciado na Figura 44.

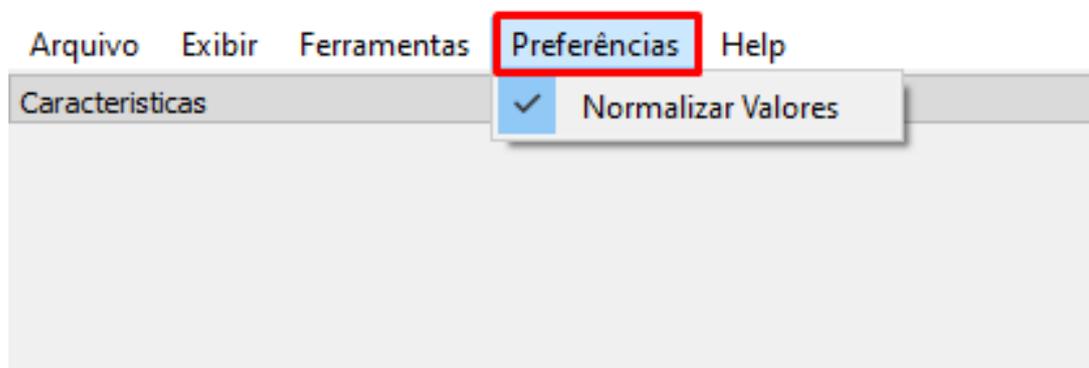
Figura 44 – Interface



Fonte: Elaborado pela autora.

Preferências: A decisão de analisar os dados normalizados ou não é deixada ao critério do usuário. Se optar por não normalizar, a opção “Normalizar Valores”, permanecerá desmarcada. No entanto, se a escolha for pela normalização, o sistema indicará isso com a opção selecionada, conforme ilustrado na Figura 45.

Figura 45 – Interface



Fonte: Elaborado pela autora.

Segmentação automática da imagem

Na fase da segmentação da imagem será uma estrutura baseado no artigo: *Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system* (GINORIS et al., 2007). Nesse projeto será utilizada a segmentação por similaridade. Onde acontecerá os processos morfológicos na imagem. Tudo isso na delimitação da região de interesse das células.

Para alcançar esse objetivo, será necessário aplicar uma série de filtros na imagem, um após o outro, para obter o processamento desejado. Por essa razão, neste trabalho, decidiu-se que tal procedimento será denominado como *Segmentação automática da imagem*.

São estes os filtros :

- Binarização:

A limiarização é uma técnica que consiste em converter imagens em tons de cinza para imagens binárias, por isso também é conhecida como binarização. A forma mais simples de limiarização consiste em converter os pixels cujo tom de cinza é maior ou igual a um valor de limiar em brancos e os demais em pretos.

É difícil escolher a melhor forma de se definir o limiar, pois dependerá das condições da imagem (contraste, iluminação, complexidade da cena) e também do resultado esperado com a limiarização (SOUZA; CORREIA, 2007).

- Inversão:

É a técnica que já obtida a imagem binarizada, ou seja, imagem preta e branca. Onde o pixel tiver o valor 0 (preto) se transformará no valor 255 (branco). Da mesma forma os valores dos pixels brancos (255), vão receber o valor 0 (preto). Assim acontece a inversão da imagem.

- Operadores Morfológicos:

A linguagem da morfologia matemática é a teoria dos conjuntos. Dessa maneira, a morfologia oferece uma abordagem unificada e poderosa para inúmeros problemas de processamento de imagens (GONZALEZ; WOODS, 2000). Uma imagem binária é composta por dois tipos de pixels: plano de fundo e o plano principal, que são representados normalmente usando preto e branco. A morfologia é baseada na teoria dos conjuntos e no

fato de que o conjunto de todos os pixels pretos constituem uma completa descrição de imagens binárias.

A ideia central da morfologia matemática é a descrição dos operadores entre imagens, ou geralmente entre reticulados completos, em termos de uma linguagem formal (JR, 1997). A morfologia é a aplicação de operadores de erosão e dilatação da imagem. Sendo que a dilatação deixa o objeto mais largo e a erosão deixa o objeto mais fino.

Para que o processamento automático seja feito é necessário seguir os passos mostrados das figuras 46 à 53.

Figura 46. Passo 1: O usuário terá que selecionar o dataset desejado para que o processamento manual seja realizado, onde as imagens estejam no seu estado original.

Figura 47 Passo 2: Dataset já selecionado, deverá seguir a rotina: Pré Processamento Manual -> Recortar Região de interesse.

Figura 48. Passo 3: Ao selecionar a rotina do passo anterior o programa exibirá uma tela onde será possível selecionar a imagem que deseja fazer o processamento.

Figura 49. Passo 4: Imagem selecionada, outra janela será exibida mostrando duas opções de recorte e a imagem que foi selecionada.

Figura 50. Passo 5: Para que o recorte seja feito o usuário terá que passar o cursor do mouse na região de interesse da célula que deseja o recorte, ao final basta dar dois cliques com o cursor para que o sistema “entenda” que o recorte chegou ao fim. Posteriormente deverá clicar no botão “Recorte 1” para finalizar o processo de recorte.

Figura 51. Passo 6: Com base no passo anterior, a figura mostra o resultado do recorte. Preservando o recorte da imagem original e aplicando em uma imagem criada com o “fundo” preto.

Figura 52. Passo 7: Para que o processamento automático seja executado basta o usuário clicar no botão “Pré Processamento Automático”. Onde aplicados os filtros baseados no artigo *Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system* (GINORIS et al., 2007).

Figura 53. Passo 8: Mostra o resultado do processamento automático da imagem.

Segmentação manual da imagem

Outra abordagem adotada neste trabalho para a segmentação das imagens é a ausência da aplicação dos filtros, conforme explicado anteriormente na fase de Segmentação automática. Este método será denominado : *Segmentação manual*. Para executá-lo, é essencial seguir os passos ilustrados nas figuras 55 a 60.

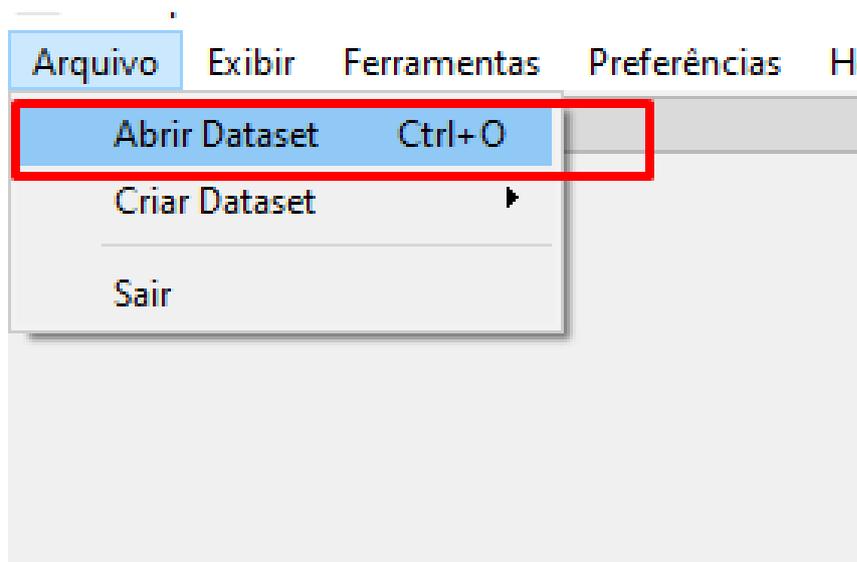
Figura 55 . Passo 1: O usuário terá que selecionar o dataset desejado para que o processamento manual seja realizado, onde as imagens estejam no seu estado original.

Figura 56. Passo 2: Dataset já selecionado, deverá seguir a rotina: Pré Processamento Manual -> Recortar Região de interesse.

Figura 57 . Passo 3: Ao selecionar a rotina do passo anterior o programa exibirá uma tela onde será possível selecionar a imagem que deseja fazer o processamento.

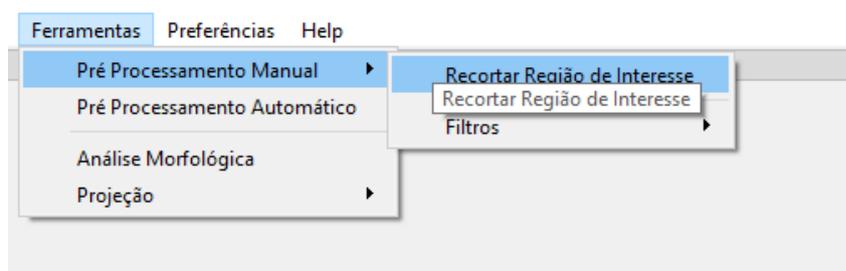
Figura 58 . Passo 4: Imagem selecionada, outra janela será exibida mostrando duas opções de recorte e a imagem que foi selecionada.

Figura 46 – Interface - Passo 1



Fonte: Elaborado pela autora.

Figura 47 – Interface - Passo 2



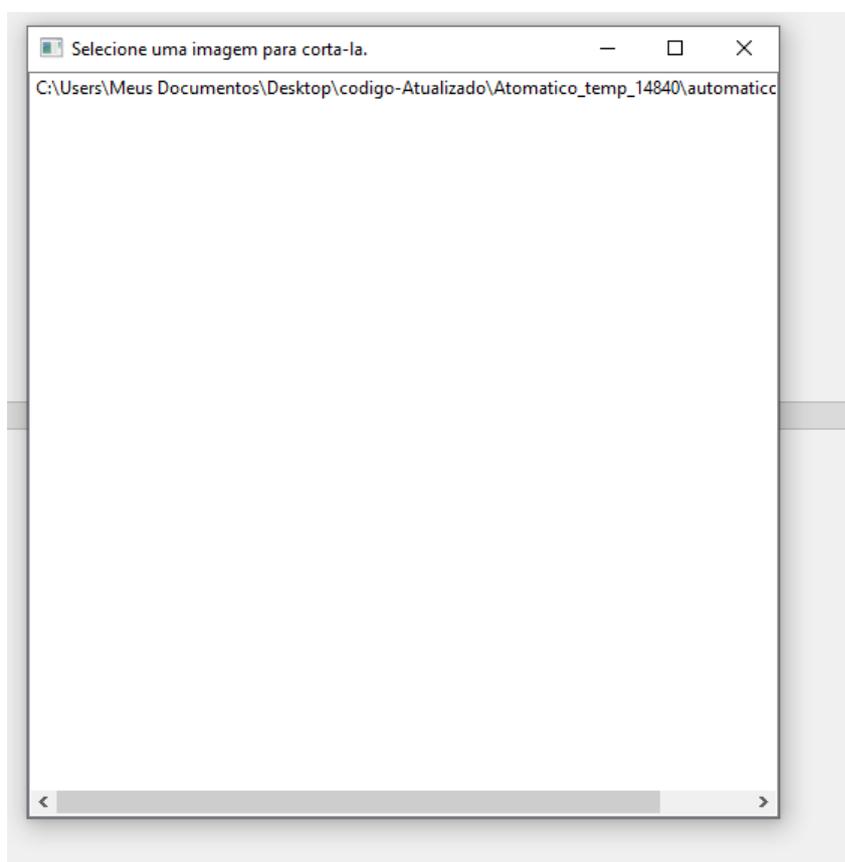
Fonte: Elaborado pela autora.

Figura 59. Passo 5: Para que o recorte seja feito o usuário terá que passar o cursor do mouse na região de interesse da célula que deseja o recorte, ao final basta dá dois cliques com o cursor para que o sistema “entenda” que o recorte chegou ao fim. Posteriormente deverá que clicar no botão “Recorte 2” para finalizar o processo.

Figura 60 . Passo 6: Mostra o resultado finalizado depois do clique no botão “Recorte 2”. Onde foi preservado o recorte da região de interesse, aplicando apenas pixels na cor branca nessa região e o “fundo” da imagem foi aplicado pixels da cor preta.

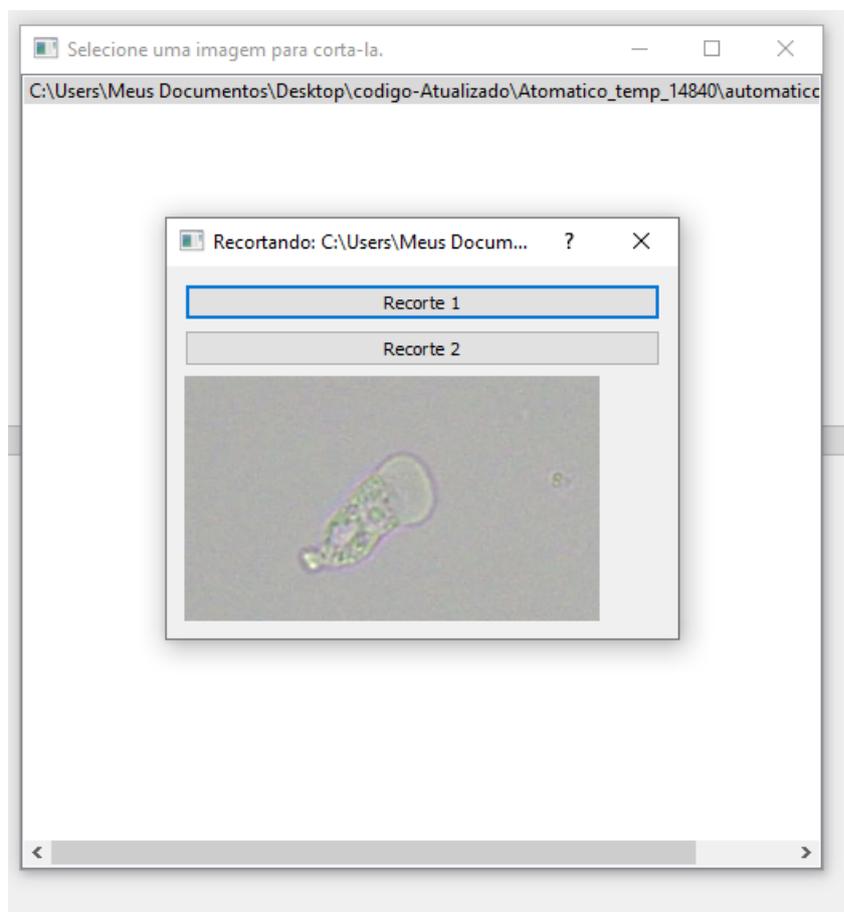
Caberá ao usuário decidir qual método utilizar: Segmentação manual ou Segmentação automática. A sugestão é escolher o método que ofereça os resultados mais satisfatórios em relação à segmentação da imagem.

Figura 48 – Interface - Passo 3



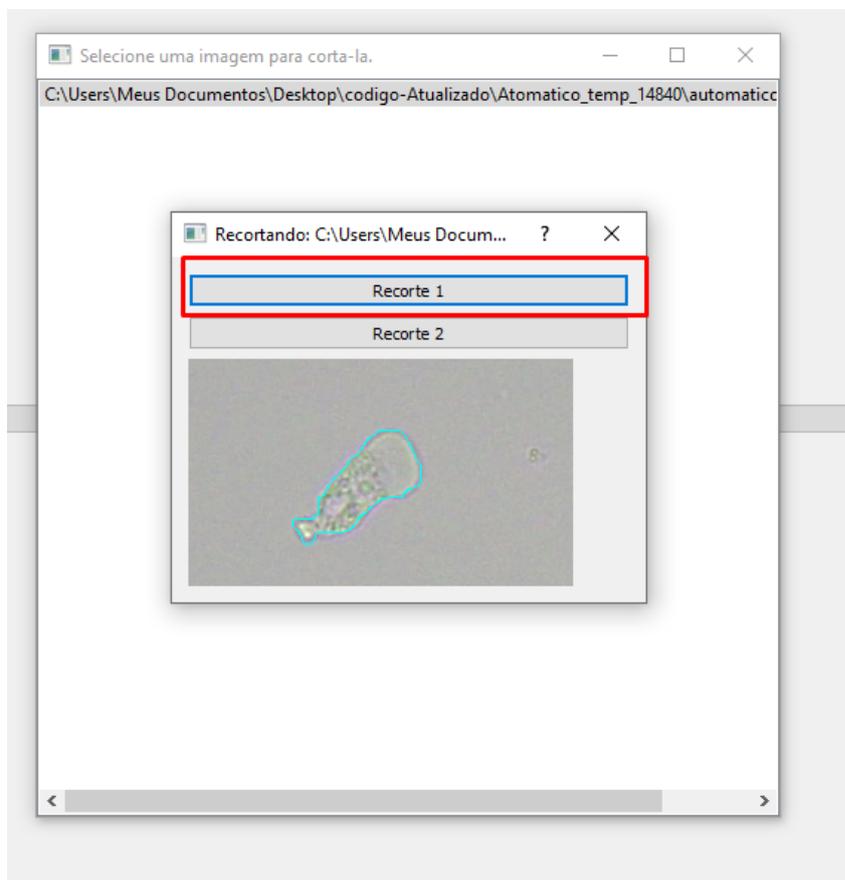
Fonte: Elaborado pela autora.

Figura 49 – Interface - Passo 4



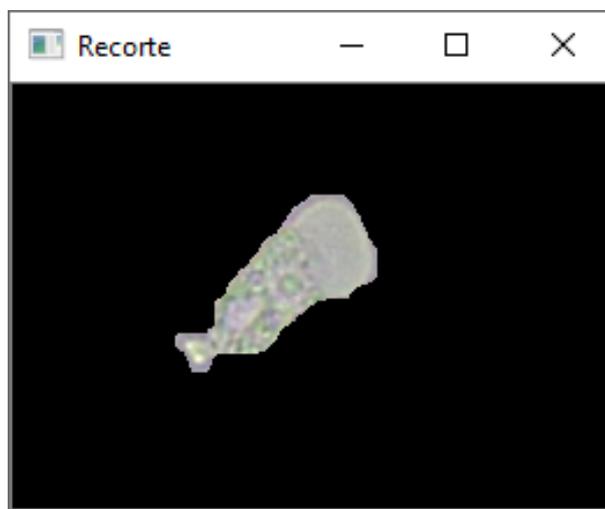
Fonte: Elaborado pela autora.

Figura 50 – Interface - Passo 5



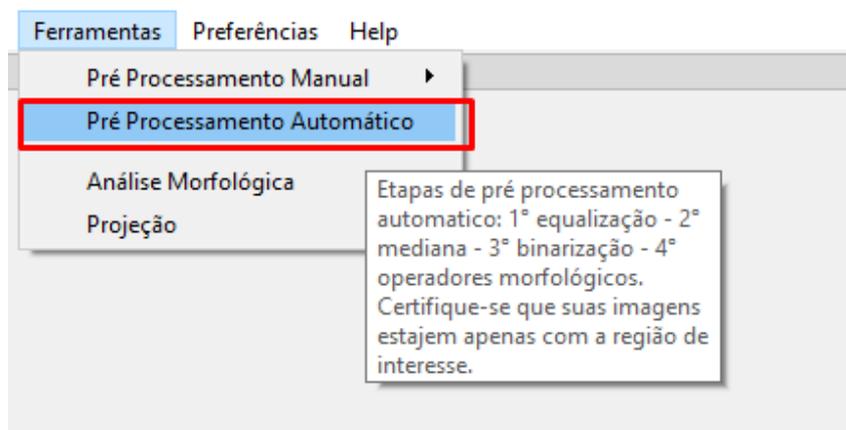
Fonte: Elaborado pela autora.

Figura 51 – Interface - Passo 6



Fonte: Elaborado pela autora.

Figura 52 – Interface - Passo 7



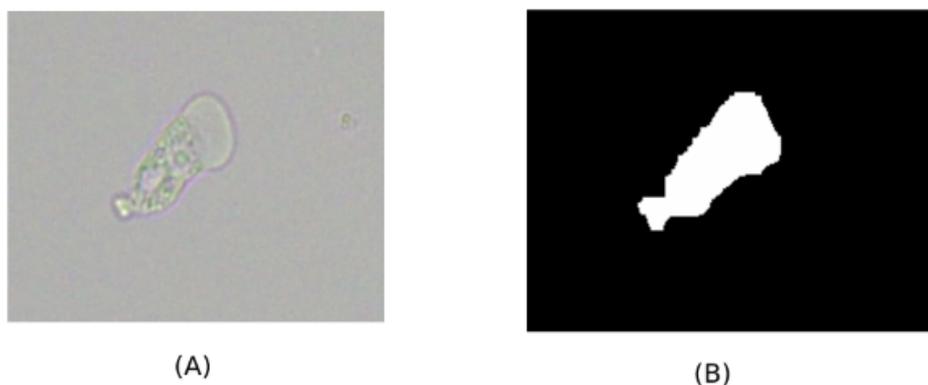
Fonte: Elaborado pela autora.

Figura 53 – Interface - Passo 8



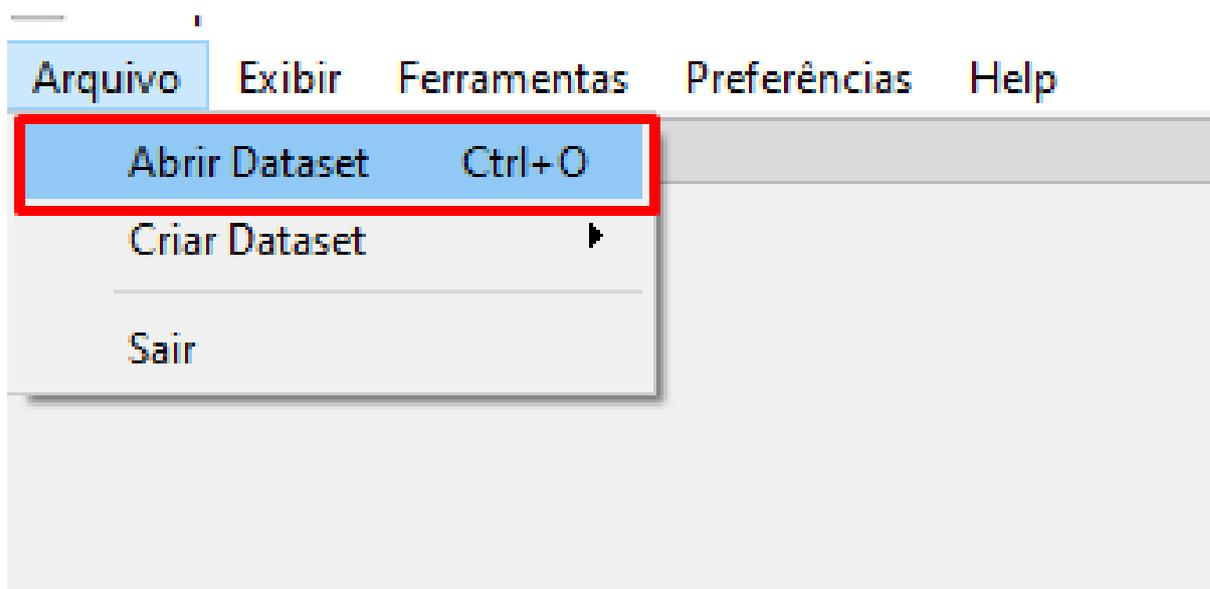
Fonte: Elaborado pela autora.

Figura 54 – Na figura (A) mostra a imagem original. A figura (B) mostra a figura passada pelo processamento automático da Figura 53.



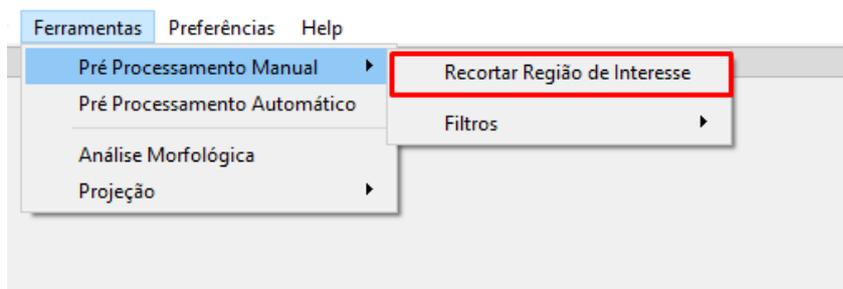
Fonte: Elaborado pela autora.

Figura 55 – Interface - Passo 1



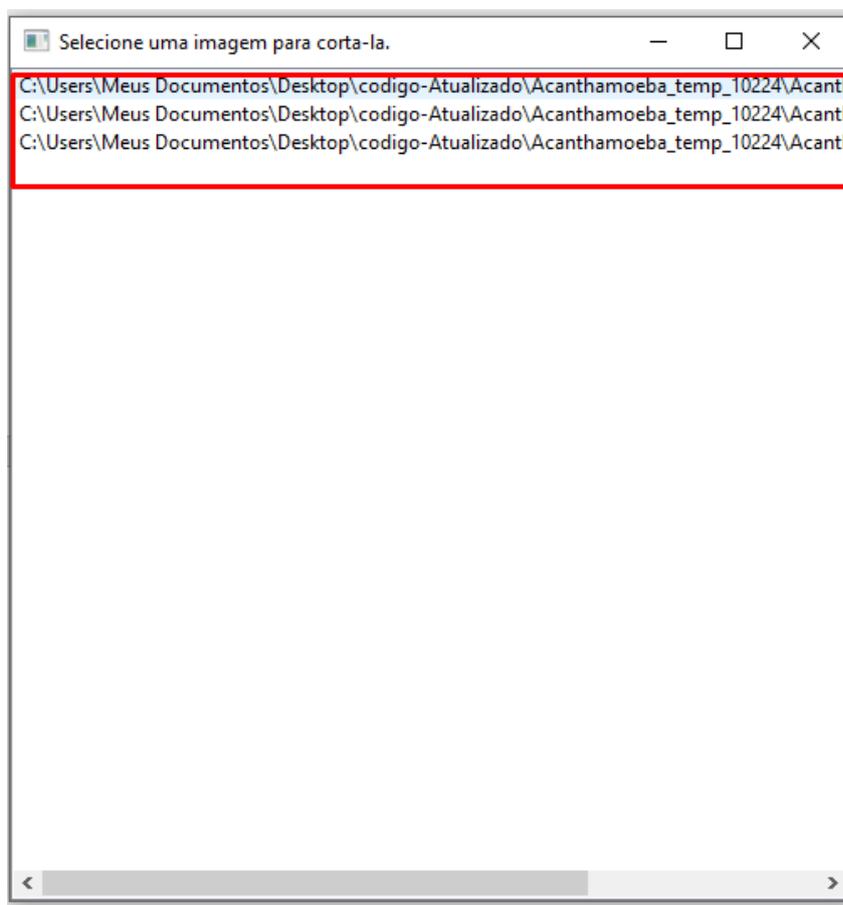
Fonte: Elaborado pela autora.

Figura 56 – Interface - Passo 2



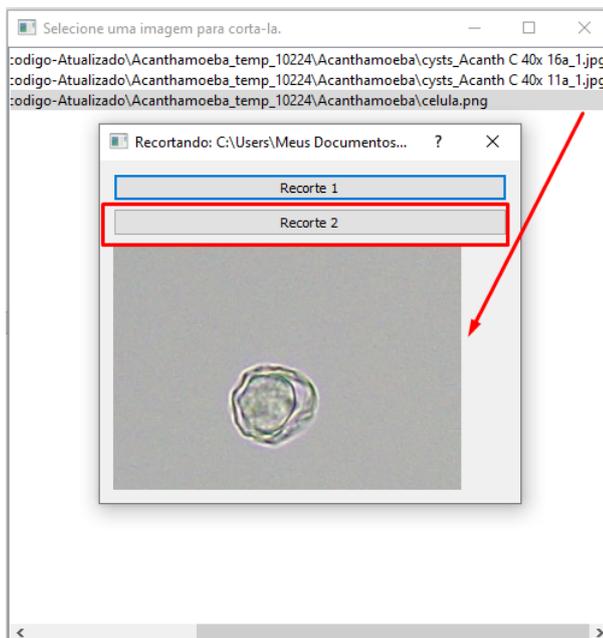
Fonte: Elaborado pela autora.

Figura 57 – Interface - Passo 3



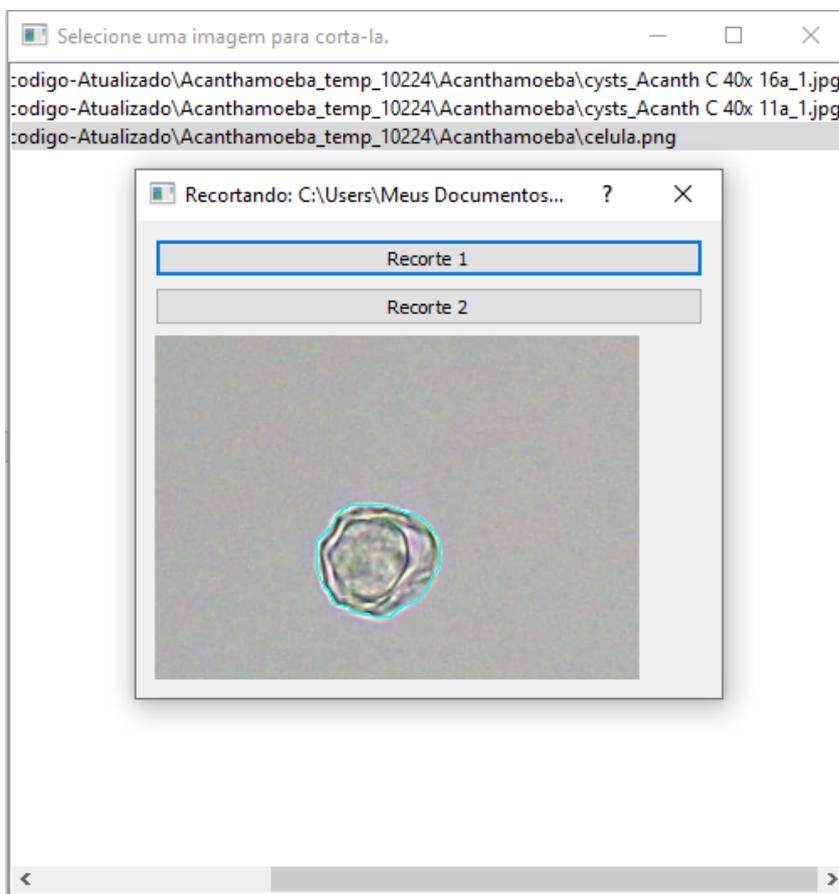
Fonte: Elaborado pela autora.

Figura 58 – Interface -Passo 4



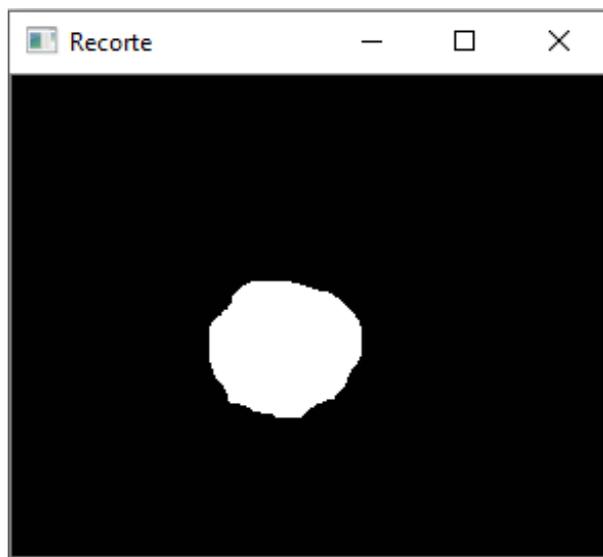
Fonte: Elaborado pela autora.

Figura 59 – Interface - Passo 5



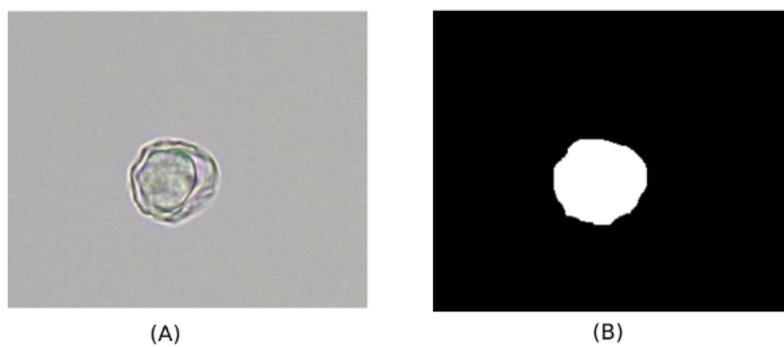
Fonte: Elaborado pela autora.

Figura 60 – Interface -Passo 6



Fonte: Elaborado pela autora.

Figura 61 – Na figura (A) mostra a imagem original. A figura (B) mostra a figura passada pelo processamento manual da Figura 60.



Fonte: Elaborado pela autora.